International Journal of Education and Social Development
ISSN: 3078-9931 (Print), ISSN: 3078-2287 (Online) | Vol. 3, No. 2, 2025

Feature Selection and Model Application of Machine Learning in Educational Quality Assessment

Xiaolan Li

School of Cardiff University, Wales, UK

Abstract: In today's society, education, as a key force driving personal growth and social progress, is highly valued for its quality. Educational quality assessment is an important means to ensure and improve the quality of education. It not only helps educational institutions understand their own teaching effectiveness and identify problems in the teaching process, but also provides scientific basis for educational decision-making, promotes the rational allocation of educational resources, and ultimately achieves the goals of educational equity and quality improvement.

Keywords: Machine learning; Educational quality assessment; Feature Selection.

1. Introduction

The traditional methods of evaluating educational quality mainly rely on manual experience and simple data statistics, such as exam scores, teacher evaluations, etc. Although these methods can to some extent reflect certain aspects of educational quality, they have many limitations. On the one hand, evaluation criteria are often relatively single, overly focusing on students' academic performance and neglecting their development in comprehensive quality, innovation ability, practical ability, etc., making it difficult to comprehensively and objectively reflect students' true level and the actual effectiveness of education. On the other hand, the manual evaluation process is highly subjective and easily influenced by factors such as the evaluator's personal preferences and experience, leading to doubts about the accuracy and reliability of the evaluation results. In addition, traditional evaluation methods are inefficient, the data collection and analysis process is cumbersome, and it is difficult to provide timely feedback on evaluation results, which cannot meet the needs of rapid development and dynamic adjustment of education. [1]

With the rapid development of information technology, machine learning, as an important branch of artificial intelligence, has brought new opportunities and solutions for education quality assessment. Machine learning is a multidisciplinary field that enables computers to automatically learn patterns and patterns from large amounts of data, thereby achieving prediction and classification of unknown data. In the field of education, machine learning can deeply mine and analyze massive amounts of educational data, which not only includes students' academic performance, but also encompasses various information such as their learning behavior, habits, interests, and social interactions. By comprehensively analyzing these multi-source heterogeneous data, machine learning can construct more comprehensive and accurate education quality assessment models, providing richer and deeper perspectives for education quality assessment. Applying machine learning technology to educational quality assessment can not only overcome the limitations of traditional assessment methods, improve the accuracy, comprehensiveness, and efficiency of assessment, but also bring new research ideas and methods to the field of education, promote innovative development of educational assessment, and have important theoretical and practical significance. [2]

2. Overview of Machine Learning and Educational Quality Assessment

According to whether there is annotation information in the learning process, machine learning can be mainly divided into three categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is the most common type of machine learning, with its training dataset containing input features and corresponding annotation information (also known as labels). The model learns from these annotated data and establishes a mapping relationship between input features and annotations, thereby predicting new unknown data. For example, in image classification tasks, each image in the training dataset is labeled with its corresponding category (such as cat, dog, car, etc.). After learning the relationship between the features and categories of these images, the model can classify and predict new unlabeled images. Common supervised learning algorithms include linear regression, logistic regression, decision tree, support vector machine, naive Bayes, etc. [3]

The training dataset for unsupervised learning only has input features and no annotation information. The goal of the model is to discover the intrinsic structure, patterns, and patterns of the data, such as its distribution, clustering, principal components, etc., from these unlabeled data. For example, in customer segmentation tasks, unsupervised learning is performed on customer consumption behavior, preferences, and other data to group customers with similar characteristics into one category, thereby achieving customer segmentation and providing a basis for precision marketing.

The education quality evaluation index system is a multidimensional and multi-level complex system aimed at comprehensively, objectively, and scientifically measuring the quality and effectiveness of educational activities. This system covers multiple key aspects such as teaching process quality, teaching effectiveness quality, school management and support system, and each aspect includes a series of specific evaluation indicators that are interrelated and influence each other, forming an organic whole. [4]

The quality of the teaching process is an important component of educational quality assessment, which focuses

on the implementation process and methods of educational and teaching activities. Teaching staff is one of the key factors in the quality of the teaching process, including indicators such as teachers' education, professional titles, teaching experience, and professional competence. Highly educated, highly skilled, and experienced teachers are often able to provide students with more professional and in-depth knowledge transmission and guidance. The rationality of curriculum design is also crucial, as it should meet educational goals and students' developmental needs, cover a wide range of subject areas, and emphasize the combination of theory and practice. The diversity and effectiveness of teaching methods also affect the quality of the teaching process. Teachers should flexibly use various teaching methods such as lecture, discussion, case teaching, project-based teaching, etc. based on the teaching content and student characteristics, to stimulate students' interest and initiative in learning. In addition, the richness and utilization efficiency of teaching resources are also important manifestations of the quality of the teaching process, including textbooks, library materials, laboratory equipment, multimedia teaching facilities, etc. Adequate and advanced teaching resources can provide strong support for teaching activities and promote the improvement of teaching quality. A certain school has invested in the construction of modern laboratories equipped with advanced experimental equipment, providing students with good experimental conditions and enabling them to better master experimental skills and scientific research methods in experimental courses. [5]

The school management and support system is an important guarantee for the quality of education, providing necessary conditions and support for the smooth implementation of teaching activities. The degree of improvement of school management system directly affects the operational efficiency and educational quality of the school. A scientifically reasonable teaching quality monitoring system can promptly identify problems in the teaching process and take corresponding measures for improvement. The efficiency and service awareness of administrative management are also crucial, as they can ensure the orderly progress of various tasks in schools. The administrative department of the school actively provides services for teachers and students, timely solves the problems they encounter in work and study, and creates a good environment for the smooth development of teaching activities. Logistics support is the foundation for the normal operation of schools, including campus safety, food hygiene, facility maintenance, and other aspects. At the same time, the school pays attention to food hygiene management and strictly supervises the cafeteria, providing students with safe and healthy food. In addition, the school regularly maintains and updates teaching facilities and equipment to ensure the normal progress of teaching activities. [6]

3. Feature Selection of Machine Learning in Educational Quality Assessment

Feature selection plays a crucial role in the evaluation of educational quality, as it is a key step in constructing efficient and accurate evaluation models. Data in the field of education typically has high-dimensional characteristics, containing a large number of potential features such as students' academic performance, study time, study location, learning devices, social relationships, etc. Some of these features are highly correlated with education quality assessment and can provide key information for evaluation; And another part may have weak correlation with the evaluation target, even containing noise, which not only fails to improve the evaluation effect, but also interferes with the learning and judgment of the model. For example, the brand of learning equipment used by students during the learning process may not be directly related to their learning effectiveness. If it is included in the evaluation model, it will not only increase the complexity of the model, but also introduce unnecessary interference factors, which may affect the accuracy of the model. Through reasonable feature selection, the most representative and influential features can be selected from numerous original features, eliminating redundant, irrelevant, or noisy features, thereby improving the performance of the model. This is mainly reflected in the following aspects:

Feature selection can improve the accuracy and generalization ability of the model. When a model is trained with a large number of irrelevant or redundant features, overfitting is prone to occur, where the model performs well on training data but performs poorly on test data or practical applications. By selecting features closely related to the objectives of educational quality assessment, the model can focus more on learning key patterns and patterns in the data, reduce noise interference, and improve its predictive and generalization abilities for unknown data. Taking predicting students' exam scores as an example, if the model only considers key features such as students' learning time and attitude, and ignores other information unrelated to exam scores, such as students' interests and hobbies, the model can more accurately capture factors that affect exam scores, thereby improving the accuracy of predictions.

Feature selection helps enhance the interpretability of the model. In the evaluation of educational quality, the interpretability of models is very important. Education decision-makers and teachers need to understand the decision-making basis of models in order to take targeted measures to improve teaching. Excessive features can make the model complex and difficult to interpret its output results. Through feature selection, the retained features are often key factors that have a significant impact on the evaluation of educational quality, making the decision-making process of the model clearer and more intuitive. For example, when evaluating the teaching quality of teachers, if the model selects key features such as students' classroom participation, homework completion, and exam scores, the relationship between these features and teaching quality is easy to understand. Teachers can analyze the problems in their teaching based on these features and take corresponding improvement measures.

4. Challenges and Countermeasures of Machine Learning Applications

Data quality is the cornerstone of the application of machine learning in educational quality assessment. However, in practical applications, educational data often faces many quality issues, which seriously affect the performance of machine learning models and the accuracy of evaluation results. Data loss is one of the common data quality issues. In educational data, there may be data loss in areas such as student grades, learning behavior records, and teacher teaching evaluations. For example, due to system

malfunctions or human negligence, some students' course homework grades for a certain semester were not recorded, or some teachers' teaching reflection reports were not submitted in a timely manner, resulting in missing relevant data. Data loss can lead to incomplete information, making it difficult for machine learning models to fully understand the situation of the evaluated object, thereby affecting the training and prediction performance of the model. If the key data of students' homework grades is missing when predicting their overall grades, the model may not accurately capture their learning process and effort level, resulting in biased prediction results.

Noise data is also an important factor affecting data quality. Noise data refers to errors, anomalies, or interference information present in the data, which may be caused by data collection equipment failures, data entry errors, or interference during data transmission. In student performance data, there may be errors in grade input, such as mistakenly recording 85 points as 58 points; In student learning time records, there may be outliers, such as recording learning time as negative or significantly outside the normal range. These noisy data can interfere with the learning process of machine learning models, causing them to learn incorrect patterns and patterns, and reducing the accuracy and reliability of the models. If the model is trained based on incorrect grade data, it may draw incorrect conclusions, such as thinking that a student's academic performance is poor when in fact their actual grades are excellent.

The issue of inconsistent data cannot be ignored. Education data comes from a wide range of sources, including school academic management systems, online learning platforms, student information management systems, etc. The data formats, definitions, and standards of different data sources may vary, which can easily lead to inconsistent data. In different systems, there may be differences in the recording of students' grade, major, and other information, or inconsistent definitions of course names and numbers. Inconsistent data can cause confusion in machine learning models when processing data, affecting the accuracy of model training and evaluation results. When analyzing students' professional learning situation, if different data sources have inconsistent records of students' majors, the model may not be able to accurately identify the students' majors, thus making it difficult to conduct effective analysis and evaluation.

5. Conclusion

This study delves into the feature selection and model application of machine learning in educational quality assessment, and has achieved a series of significant theoretical and practical results. In terms of feature selection, a comprehensive analysis was conducted on multidimensional educational data features such as students, teachers, and teaching resources, clarifying the important impact of each feature on educational quality evaluation. The characteristics of students' grades, study habits, and study time are key indicators reflecting their learning status and effectiveness. The teaching experience, teaching methods, and educational background of teachers directly affect the quality of teaching. The characteristics of teaching resources such as teaching equipment, textbook quality, and library resources provide material basis and knowledge support for educational activities. On this basis, the system studied various feature selection methods such as filtering, wrapping, and embedding, and successfully applied them to educational data processing. The filtering method is based on the statistical characteristics of features, and through methods such as correlation analysis and chi square test, quickly selects features that are highly correlated with the educational quality assessment objectives, laying the foundation for subsequent analysis. The wrapper approach is performance oriented towards specific machine learning models, such as recursive feature elimination and genetic algorithm-based feature selection, which can fully consider the interactions between features and find the feature combination that maximizes model performance.

References

[1] Reunanen J .Model Selection and Assessment Using Cross-indexing[J].IEEE, 2007.DOI:10.1109/IJCNN.2007.4371365.

[2] Assegie T A , Murugan S , Govindarajan R ,et al.Improving the Performance of Machine Learning with Sequential Feature Selection and Grid Search[J].Przegląd Elektrotechniczny, 2024(7).

[3] Niu H , Guan B , Pang E .Application of SE Module Based YOLO v8 Network and Machine Learning in Laser Notch Quality Evaluation[C]//International Conference on the Frontiers of Robotics and Software Engineering.Springer, Singapore, 2025.

[4] Lima H S , Oliveira G F V D , Ferreira R D S ,et al.Machine learning-based soil quality assessment for enhancing environmental monitoring in iron ore mining-impacted ecosystems[J].Journal of Environmental Management, 2024, 356(000):11.

[5] Hemdanou A L , Sefian M L , Achtoun Y ,et al.Comparative analysis of feature selection and extraction methods for student performance prediction across different machine learning models[J].Computers and Education: Artificial Intelligence, 2024, 7(000).

[6] Dostál, Ji?í,Serafín, ?estmír, Havelka M ,et al.Assessment of Quality of Material Educational Tools for Technical Education[J].Procedia - Social and Behavioral Sciences, 2012, 69(Complete):709-718.