INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES
Vol 15, Issue 9, (2025) E-ISSN: 2222-6990

Prediction of Mobile Learning Video Conferencing Requirements for Higher Learning Institutions in China Using Machine Learning

Lei Lu[1,2], Nazean Jomhari[3*], Shang Pang[3], Mohd Nurul Hafiz Ibrahim[4]
[1]City Graduate School, City University Malaysia, Petaling Jaya, Malaysia, [2]Jiangxi College of Applied Technology, Ganzhou, China, [3]Department of Software Engineering, Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia, [4]City Graduate School, City University Malaysia, Petaling Jaya, Malaysia
Email: lulei1618@gmail.com, nazean@um.edu.my, 22106157@siswa.um.edu.my, mohd.nurul.hafiz@city.edu.my

DOI Link: http://dx.doi.org/10.6007/IJARBSS/v15-i9/26370

Published Date: 13 September 2025

Abstract
This study investigates the functional requirements of video conferencing platforms in Chinese higher education, where mobile learning has become increasingly essential. Based on data from 517 students and 211 teachers, we apply K-Means clustering and SHAP analysis to identify distinct user demand groups—low, medium, and high. Key features such as attention tracking, interactive tools, and feedback functions are found to significantly influence user preferences. Unlike traditional models focused on outcomes, this research emphasizes demand structure discovery through unsupervised learning. The findings offer actionable insights for platform developers and educational decision-makers, promoting the development of smarter and more personalized learning environments. This study also highlights the value of explainable AI in understanding user behavior and improving platform adaptability.
Keywords: Video Conferencing, Mobile Learning, Unsupervised Learning, Educational Technology, Functional Requirements

Introduction
Background
With the rapid development of technologies such as mobile Internet, smart terminals and cloud computing, mobile learning has gradually become the core carrier of the digital transformation of global higher educationGuo et al. (2025); Nadzinski et al. (2023). In China, especially since the outbreak of the COVID-19 pandemic, online teaching has become an important form of normal teaching in colleges and universities, further accelerating the

widespread application of the "mobile learning + video conferencing" model in colleges and universitiesHo et al. (2021). More and more college teachers and students use platforms such as "Tencent Conference", "DingTalk" and "Zhidao" to achieve remote teaching and learning, and use screen sharing, interactive whiteboards, real-time discussions and other functions for teaching interaction and knowledge transferLuo et al. (2023). This video conferencing-based teaching method provides technical support for improving teaching flexibility, resource coverage and educational equityLiang and Wang (2022).

However, video conferencing platforms have also exposed many problems in the actual application of higher education. For example, students of different grades, majors, genders and regions have significant differences in their user experience, and their expectations for platform functions are also highly diverseLuo et al. (2023). Some students tend to need more interactive functions, such as online quizzes, reflection prompts, real-time translation, etc.; while others focus on simple interfaces, network stability or self-learning tools.

Faced with such complex and diverse usage requirements, colleges and universities and platform developers urgently need to identify user demand characteristics in a more scientific and data-driven wayKhor (2022). Traditional average statistical analysis based on questionnaires is difficult to reveal the underlying structural laws, and subjective experience decisions are prone to resource mismatch. Therefore, how to build a systematic, objective and interpretable analysis mechanism to identify the implicit functional requirements of students and teachers in video conferencing platforms has become a key issue in current educational informatization researchMiranda et al. (2024).

To this end, this study starts with large-scale questionnaire data from students and teachers, and uses unsupervised learning methods, especially KMeans clustering algorithm and SHAP value interpretation technology, to analyze and visualize the key features that affect video conferencing needs, thereby realizing user demand group identification and data interpretationMiranda et al. (2024); Nadzinski et al. (2023). Different from traditional supervised learning models, this study does not aim at label prediction, but focuses on the exploratory identification of demand patterns, emphasizing the interpretability of demand structure and the diversity of user portraits. This method is not only applicable to scenarios without clear labels in reality, but also more in line with the actual characteristics of individual differentiation, ontological complexity and preference ambiguity in the field of education.

Through cluster analysis and interpretation of important characteristics of student and teacher demand characteristics, this study hopes to provide accurate and efficient optimization suggestions for university administrators, teaching decision makers and platform developers, and realize data-driven teaching resource configuration and platform function innovation based on real needsGuo et al. (2025). The ultimate goal is to promote the development of mobile learning platforms in a more intelligent, personalized and inclusive direction in the context of digital higher education, and help improve the overall level of education modernization.

Research Objectives

This study is dedicated to exploring how to use unsupervised learning methods to explore the potential demand structure of students for video conferencing platform functions in the context of increasingly popular mobile learning in higher education in China, so as to provide data support for the intelligent upgrade of teaching platforms and the personalization of educational services. The overall goals and specific sub-goals are as follows:

Overall goal
Build a user portrait identification and functional demand modeling framework based on unsupervised learning, discover students' key functional preferences and demand levels for video conferencing systems in mobile learning scenarios from structured questionnaire data, and achieve demand-oriented optimization of platform design and educational resource allocation.

Specific goals
(1)Identify the core dimensions of video conferencing needs
Based on the questions related to interface personalization, learning assistance, interactive functions, and teaching feedback in the student questionnaire (such as Q21-Q27), construct a variable system reflecting functional demand tendencies as input for cluster analysis.
(2)Identify demand groups using unsupervised learning algorithms
Use the KMeans clustering algorithm to divide student groups, explore the potential heterogeneity in video conferencing function preferences, and form a clear "user portrait".
(3)Combined with SHAP method to enhance the interpretability of clustering results
The key influencing factors of different clustering groups are analyzed through SHAP values to clarify which behavioral or attitudinal variables can best distinguish different types of needs.
(4)Analyze the individual difference characteristics of demand groups
Explore the distribution characteristics of individual variables such as gender, grade, and platform usage frequency among different clustering groups to reveal the background factors that affect the demand for video conferencing.
(5)Provide optimization suggestions for platform design and university decision-making
Based on the results of clustering and interpretation analysis, propose functional optimization directions and educational resource allocation strategies for different user groups to improve the adaptability and teaching support efficiency of video conferencing platforms.

Literature Review
Past Studies
Application of data mining and prediction technology in education
In recent years, with the advancement of artificial intelligence and big data technology, machine learning has been widely used in data analysis and prediction in the field of education. Khor (2022)proposed a data mining method based on student behavior data to predict academic achievement and behavior patterns. This method confirms the feasibility of machine learning technology in revealing student demand characteristics.

Nadzinski et al. (2023) conducted a systematic review of this field and believed that supervised learning technology (such as support vector machines, random forests) and clustering technology have inherent advantages in educational environments with high heterogeneity of student groups, which can significantly improve the accuracy of learning path modeling and resource recommendations.

Following this line of thought, Miranda et al. (2024)conducted a systematic review of data mining technology and believed that supervised learning technology (such as support vector machines, random forests) and clustering technology have inherent advantages in educational environments with high heterogeneity of student groups, which can significantly improve the accuracy of learning path modeling and resource recommendations. Miranda et al. (2024) proposed a model-agnostic interpretability method (Model-Agnostic Interpretability), and combined it with tools such as SHAP to make the prediction model more interpretable, so that teachers and education administrators can more clearly understand the causal relationship between learning behavior and learning outcomes.

Optimization of online teaching platform performance and resource utilization As an important infrastructure for online learning, the service quality of the video conferencing platform directly affects the teaching experience and teaching effect. In order to improve the availability, response speed and stability of the video conferencing platform, Liang and Wang (2022)constructed a QoS-based video service performance prediction model.

Lin et al. (2021)established a multivariate resource scheduling model for the problem of online teaching resource allocation, focusing on the optimization effect of dynamic resource allocation methods based on time and user characteristics on teaching effect.

In addition, Guo et al. (2025)studied how AI-driven mobile learning platforms organize the most ideal learning resource allocation. They proposed an AI-driven education platform framework that combines personalized learning suggestions with data-driven feedback processes to improve the platform's responsiveness and material adaptability.

Teaching Evaluation System and Personalized Platform Design

Platform personalization and intelligent teaching evaluation system have always been the main directions for optimizing digital teaching platforms. The interactive teaching evaluation system proposed by Li (2024) combines student feedback, platform function usage data, and teacher usage behavior to achieve dynamic monitoring and feedback of teaching effects.
Luo et al. (2023)studied the acceptance of video conferencing learning among Chinese college students and found that students' satisfaction with the platform interface design, ease of use, and platform interactive functions significantly affected their willingness to use, highlighting the role of platform personalization functions.

Teaching Mobile Network Technology and Infrastructure Modeling

Good network infrastructure is the basis for ensuring the smooth operation of video conferencing systems. Mei et al. (2020)proposed a mobile bandwidth prediction model based on long short-term memory (LSTM) neural network and Bayesian fusion, which provides strong technical support for the common network instability problem in mobile learning.

In addition, Ho et al. (2021)used machine learning methods to predict students' satisfaction with emergency remote teaching during the COVID-19 pandemic in their study, and determined that the platform technology status, network stability and teaching support level were the main variables of satisfaction, which once again confirmed the core role of infrastructure in video conferencing learning.

Research Gaps

The Evolution of the Learning Needs Prediction Research Paradigm

Although traditional learning prediction models mostly use supervised learning methods and focus on "result variables" such as student grades and completion rates Khor (2022); (Tahiru et al., 2023), researchers have gradually realized in recent years that analysis based on subjective behavioral preferences and functional needs is equally critical. Starting from the "functional needs" of video conferencing tools, this study attempts to explore the differentiated needs of student groups using unsupervised learning methods, which is a new research paradigm that is different from the traditional "prediction results".

Scarce Application of Unsupervised Modeling in Educational Data Mining

A large number of literature (Miranda et al., 2024); Nadzinski et al. (2023) focuses on using supervised learning models for personalized recommendations and path modeling, but there is still a lack of systematic exploration of demand clustering and preference discovery in unlabeled scenarios. Cluster analysis, as a typical representative of unsupervised learning, can effectively mine hidden structures between heterogeneous groups, and is particularly suitable for demand mining tasks in education that lack explicit output variables.

Gaps in Structured Modeling Methods for Educational Questionnaire Data

Currently, mainstream educational modeling data mostly comes from operational data such as learning management system (LMS) logs and click behavior trajectories, while the machine learning application of multiple-choice questions and Likert scale data in structured subjective questionnaires is still in the exploratory stage. This study uses questionnaire data from 517 Chinese college students, and through clustering modeling of questions 21-23 and 27, it realizes the mining of "student video learning function demand pattern", which has methodological innovation significance in data type and modeling method.

Shift of research objectives: from predicting variables to "explaining demand structure"

Although this study did not use clear labels for supervised learning modeling, the "clustering labels" obtained by unsupervised learning methods can indirectly reflect the strength and weakness of functional requirements. Combined with feature importance (such as SHAP explanation) and difference visualization analysis, it can provide user segmentation logic for the subsequent construction of personalized learning platforms. This kind of "demand explanation-oriented" research broadens the horizons of traditional "score prediction" research.

The combination of model interpretability and educational practice needs to be strengthened

Traditional clustering research mostly stays at the level of mathematical modeling and classification accuracy, lacking in-depth interpretation of educational significance. With the help of model-independent interpretation methods such as SHAP, this study can associate students' performance in different dimensions (skill mastery, function usage perception, video learning expectations, etc.) with their demand groups, thereby providing more actionable decision support for platform function design and tiered teaching resource delivery.

Theoretical Framework

Theoretical Basis and Behavioral Model Support

This study aims to explore the heterogeneity of the needs of college students and teachers in the video conferencing environment, identify their actual usage expectations and potential functional demands for the platform, and predict needs from multi-dimensional behavior and background characteristics. In order to more scientifically construct a video conferencing demand analysis framework, the study integrates the technology acceptance theory (TAM), the unified technology acceptance model (UTAUT), the educational informationization path model (EDIPT) and the machine learning interpretability theory to establish a multi-level demand prediction foundation.

Technology Acceptance Model (TAM)

TAM was proposed by Davis (1989) and is the most representative theoretical framework in the field of educational technology to explain user adoption behavior. It emphasizes that "perceived ease of use (PEOU)" and "perceived usefulness (PU)" are key factors affecting user behavioral intention (BI). Many studies have confirmed that this model is suitable for explaining the motivation of college users to adopt digital learning tools Bailey et al. (2022). In the questionnaire design of this study, multiple dimensional questions (such as frequency of use, satisfaction, platform interaction evaluation, etc.) can be mapped to TAM key indicators to construct potential demand types and assist in the explanation of the behavioral mechanism of unsupervised models.

Unified Technology Acceptance and Use Model (UTAUT)

The UTAUT model proposed by Venkatesh（2003）expands four major influencing factors based on TAM: performance expectations, effort expectations, social influence, and facilitating conditions. This theory is particularly suitable for complex systems such as collaborative teaching between teachers and students and uneven distribution of technical resources in educational scenarios. This study maps the obstacles, influence of others, support and guarantees faced by students or teachers in platform use into the UTAUT framework to identify the social infrastructure of demand differences.

Educational Demand Identification and EDIPT Framework

The five-stage model of educational information technology development (EDIPT) emphasizes the evolutionary path of teaching digitization from "tool adoption" to "demand feedback" to "intelligent intervention", and proposes an iterative mechanism of "learning demand drive-platform function optimization-learning path adaptation". In this study, platform demand is not only manifested as functional interest, but also as deep-level learning assistance demands and participation in the teaching process. This multi-dimensional demand structure is an important target for mining through clustering models and feature analysis methods (such as SHAP).

Adaptability of Unsupervised Learning and Demand Modeling

Under the premise of no explicit annotation of variables, unsupervised learning (such as K-Means clustering) can reveal the natural distribution structure and potential patterns of individuals under multiple behavioral variables, and is often used to identify heterogeneity in complex systems. Through the joint modeling of multiple feedbacks from students or teachers, this study, based on the path logic of "behavioral characteristics → clustering grouping →

reasoning demand intensity", shifts the traditional "result prediction" to "demand classification and behavioral insight", and constructs an intelligent analysis paradigm suitable for educational decision-making.

Research Methodology

Comparison and Innovation of Research Methods

Review of Existing Research Methods

In current research on digitalization of higher education, mainstream methods are mostly focused on supervised learning paths of questionnaire analysis + regression/classification modeling. For example, Guo (2025) adopted a hybrid method to build a functional adaptation model based on questionnaires and platform behavior logs, and used deep learning and platform design optimization for verification. However, most of these studies use "learning outcomes" (such as grades and satisfaction) as target variables, lacking in-depth discussion of "functional demand preferences".

Methodological Innovation of this Study

This study combines structured questionnaires with machine learning algorithms to propose an unsupervised learning framework with cluster analysis as the core and SHAP interpretation as the auxiliary, which has the following three innovative features:
Transformation of target definition: For the first time, "video conference learning function demand score" is used as the core variable, no longer limited to traditional outcome variables, focusing on education platform optimization and user support strategies.

Algorithm path reconstruction: From traditional supervised modeling to unsupervised clustering (K-Means), combined with SHAP value to evaluate key influencing factors, and build a "data-driven + explanation-friendly" model logic.

Dual-subject analysis perspective: adopt a dual questionnaire design (517 students + 211 teachers), and model and analyze them separately to realize the heterogeneity mining and comparison of teacher and student needs, and provide multi-angle support for function development.

Research Methods

Data Collection Method

The research team designed and distributed two structured questionnaires, one for college students and one for teachers. The questionnaire design integrates the TAM model and UTAUT theory, combined with the specific application scenarios of video conferencing teaching, covering multiple dimensions such as platform usage habits, functional preferences, technical support, psychological perception and future expectations.

The student questionnaire focuses on learning behavior, platform experience and functional evaluation, constructs demand variables and explores influencing preferences;
The teacher questionnaire focuses on teaching conditions, teaching behavior changes and platform feedback, focusing on identifying potential support needs from the perspective of education supply.

The questionnaire mostly uses the Likert five-point scale to ensure the feasibility of data structuring and quantitative modeling.

The data was collected online, including the university academic affairs system, class groups, teacher communities, etc., and finally 517 valid student questionnaires and 211 teacher questionnaires were obtained, which are representative and diverse samples.

Sample Setting and Feature Variable Selection

To ensure that cluster analysis can accurately reflect the core demand preferences and behavior patterns of users in the video conferencing teaching environment, this study carefully selected several questions from the numerous items in the questionnaire as cluster feature variables, respectively targeting the role differences and decision-making preferences of students and teachers, and designed as follows:

(I) Student cluster feature selection logic

The student-side selected questions include:

Q21: Platform and tool proficiency

Reflects the students' familiarity with the operation of video conferencing platforms, learning software and social tools, which is the premise for their smooth use of online learning environments.

Q22: Learning function preference and usage experience

Covering the use of functions such as quizzes, notes, and translation, directly corresponding to platform function support and learning interaction.

Q23: Interactive function evaluation

Including the preference for interactive tools such as bullet screens, virtual backgrounds, and expressions, it is a key indicator to measure students' response to immersive and gamified teaching.

Q27: Participation and classroom behavior

Including data such as whether to speak, show, and turn on the camera, it is a direct reflection of learning initiative and online participation.

The above items construct student learning portraits from four dimensions: technical adaptability, functional demand behavior, situational interactive feedback, and learning participation, which can better support unsupervised clustering modeling based on "video conference learning needs".

(II) Teacher clustering feature selection logic

The teacher-side selected items include:

Q20: Teaching behavior and task practice type

The items cover online teaching, teaching resource development, homework feedback and collaborative project evaluation, etc., which fully reflect the implementation path and degree of dependence of teachers on video teaching tasks.

Q25: Acceptance of teaching interactive elements

Involving functional preferences such as animation, virtual images, and interactive icons, reflecting the subjective acceptance level of teachers to enhance teaching immersion and interactivity.

Q27: Feedback on classroom interactive functions

Evaluating the "discussion group function" is a key indicator for teachers to observe student participation and their own guidance effects.

Q28: Data support function demands

Including learning data analysis, personalized recommendations, online corrections, etc., are the functional expectations of teachers for the platform to provide "intelligent teaching support", which has extremely high practical guidance value.
These items cover multiple dimensions from teaching operations, interactive experience to data support, which helps to comprehensively characterize the functional dependence and preference differences of teachers on video conferencing platforms, and are an ideal feature set for building a clustering model on the supply side of the platform.

Data Analysis Method

In order to explore the heterogeneity of user groups' needs in video conferencing teaching, this study constructed the following analysis process:
(I) KMeans clustering modeling
Standardize the selected Likert score variables;
Use the KMeans algorithm to perform unsupervised clustering of samples;
Calculate the "overall demand intensity" of each category based on the mean of the cluster center and map it to "Low/Medium/High Demand";
Use principal component analysis (PCA) for cluster visualization to verify the sample distribution boundary and structural clarity.
(II) SHAP explanation mechanism
Considering that unsupervised clustering itself lacks causal reasoning ability, this study further uses XGBoost to build a "cluster label predictor" and combines SHAP (Shapley Additive Explanation) values to evaluate the impact of each variable on the prediction results:
Each type of demand corresponds to a set of SHAP values;
Extract the top ten SHAP features and evaluate their impact on different demand groups;
Draw a multi-color bar chart to compare the explanatory strength of each feature in high, medium and low demand groups.
This process makes the "driving variable" behind each cluster visual, interpretable and operational, enhancing the practical value of the results for educators and platform developers.

Analysis Tools and Experimental Environment

This study uses the Python programming language, and the main dependent libraries include:
pandas and matplotlib: for data preprocessing and visualization;
scikit-learn: for clustering and standardization;
xgboost and shap: for model construction and interpretation analysis.
The entire analysis process is performed locally, and pycharm is used to output charts and track results, ensuring that the results are transparent and reproducible.

Research Ethics

In order to protect the rights and interests of participants and data security involved in the research process, this study takes the following safeguards at the ethical level:
Informed consent: There is a statement at the beginning of the questionnaire to explain the purpose of the research, data usage, anonymity and voluntary participation mechanism. Participants click to enter to indicate that they are informed and agree;
Privacy protection: The questionnaire does not collect any personal identification information such as name, phone number, email address, etc.; all data are stored by number, and researchers cannot trace the identity of participants;

Data security: The original data is uniformly stored in an encrypted folder, and the researcher himself has access rights. It cannot be disseminated or reused without permission;

Ethical approval: The research design and questionnaire content have been reviewed by the school-level ethics, and the survey will be carried out after obtaining formal approval;
The results are open and transparent: No personally identifiable data is displayed in the research report, and all analysis results are presented and discussed in terms of group characteristics.

Summary
This chapter systematically describes the research methods, data sources and analysis process used in this study. By comparing the hybrid methods and supervised learning models commonly used in existing studies, the three major innovations of this research method in terms of data structure, prediction target definition and model interpretation mechanism are clarified. The study not only obtained high-quality samples from Chinese college students (517) and teachers (211) based on two structured questionnaires, but also constructed a set of characteristic variables focusing on platform function preferences and classroom behaviors by reasonably screening items to ensure that cluster analysis has solid theoretical and practical support.

In terms of analysis methods, this study uses the KMeans unsupervised learning algorithm to group student and teacher data separately to explore potential video conferencing teaching demand patterns. At the same time, in order to make up for the lack of interpretability of the clustering model, the SHAP (SHapley Additive Explanations) mechanism is introduced to visualize the core variables that affect demand grouping, thereby improving the transparency and practicality of the model in educational decision-making.

Analysis and Modeling
EDA for Student Data
To ensure that the input features of the clustering model are representative and have predictive value, this study conducted a systematic exploratory data analysis on the four key questions (Q21, Q22, Q23, and Q27) selected before modeling, covering students' skill foundation, behavioral performance, functional preferences, and expansion needs for video conferencing platforms.

Question Q21: Proficiency in online learning skills
Question Q21 evaluates students' mastery of online learning tools from three aspects: "teacher platform", "learning software" and "social application". The bar chart results show that most students' scores on the three sub-items are concentrated between 3 and 5 points. In particular, in the dimension of "social application", the number of students who scored the highest score (5 points) is the largest, indicating that the proficiency of this type of tool is the strongest.

This shows that the interviewed students generally have a good foundation in digital skills, especially in interactive applications, which provides the operational premise for them to use various functions (such as chat, interaction, group collaboration, etc.) in the video conferencing platform.

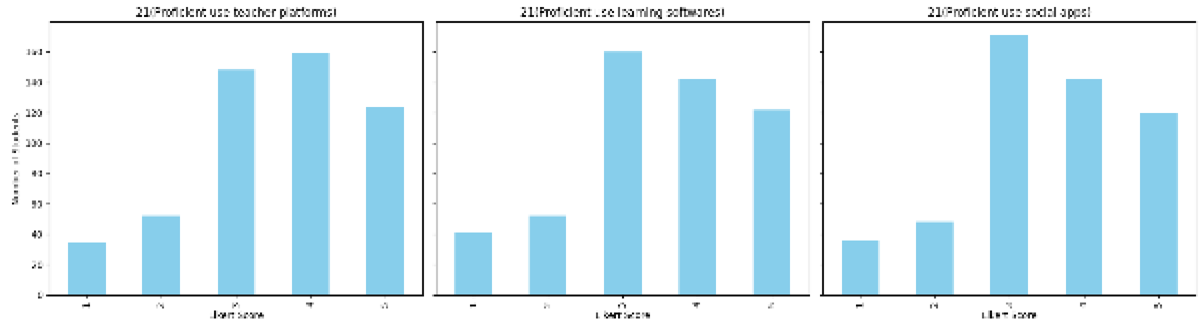

Figure 4.1 Question Q21: Proficiency in online learning skills

Question Q22: Online learning behavior and adaptation feelings

Question Q22 is designed with eight dimensions including "task completion", "classroom concentration", "interaction with teachers", and "group cooperation" to reflect students' behavior patterns and emotional adaptation on the video conferencing platform. The results of the line graph analysis show that the scores of all variables are generally concentrated in 3-4 points, reflecting that most students have a neutral or relatively positive attitude towards the online teaching model.

It is worth noting that the proportion of high scores in items such as "group collaboration effectiveness" and "improvement of learning effects" is relatively low, suggesting that the platform's social functions and support effects still have room for improvement in some students. This difference in behavioral adaptability provides an important behavioral basis for subsequent demand clustering analysis.

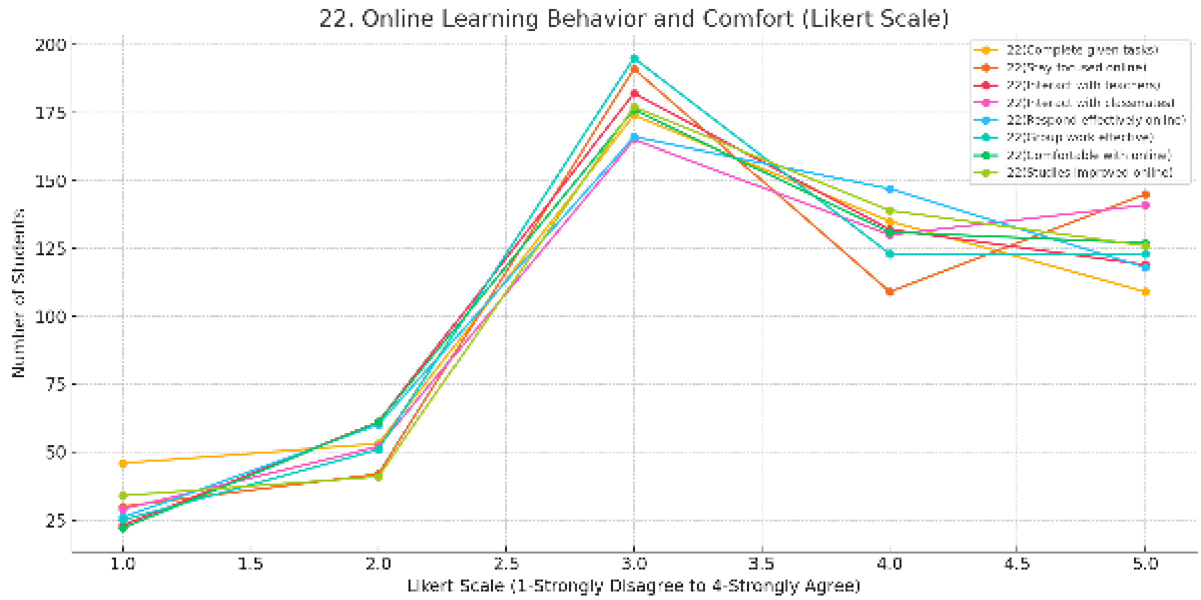

Figure 4.2 Question Q22: Online learning behavior and adaptation feelings

Question Q23: Functional Preference and Interaction Expectation

Question Q23 examines students' interest in ten types of interactive functions (such as attention tracking, self-assessment tools, screen sharing, virtual gifts, online quizzes, etc.). As can be seen from the line graph, the scores of most functions are mainly distributed between 3 and 4 points, among which "self-assessment tools", "interactive quizzes", "attention tracking" and other functions have received more high scores, indicating that they are more popular.

In addition, some design and personalized functions (such as "virtual gifts" and "avatar customization") have relatively low scores, indicating that there are individual differences in their applicability. The results of this question can reveal students' specific preferences for platform functionality and have a direct contribution to demand structure modeling.

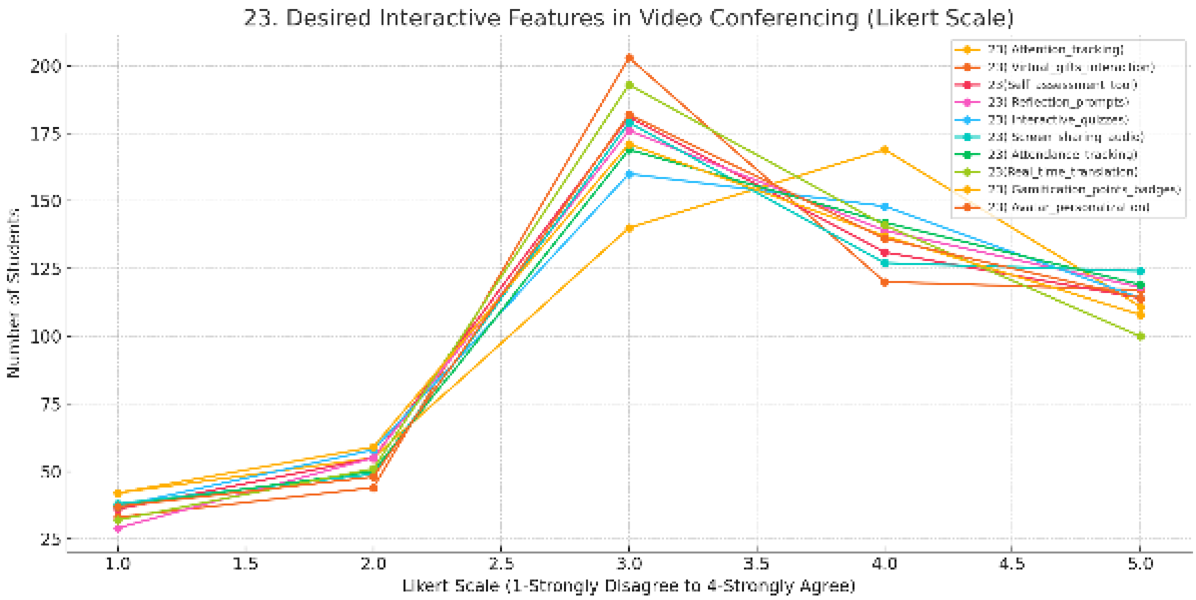

Figure 4.3 Question Q23: Functional Preference and Interaction Expectation

Question Q27: Additional Functional Requirements in Video Conferencing

Question Q27 is a multiple-choice question, which shows students' actual selection behavior for additional functions such as "knowledge point reminder", "real-time translation", and "smart notes". The statistical chart results show that "knowledge point reminder" has the highest selection frequency (437 people), followed by "real-time translation" (350 people) and "smart notes" (272 people), while only 4 people chose the "other" option, showing that the functional preferences are concentrated and clear.

Compared with the scoring question, this question can better reflect students' real usage expectations through direct behavioral selection, which helps the clustering model to be more targeted and practical in explaining functional requirements.

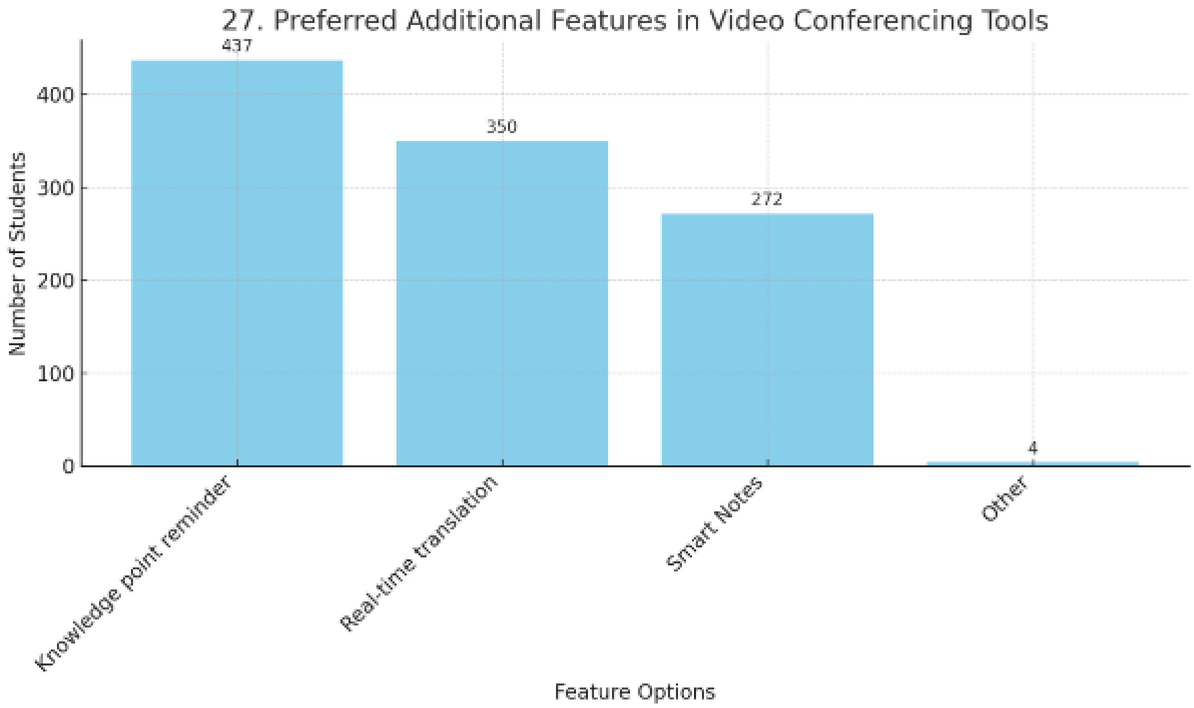

Figure 4.4 Question Q27: Additional Functional Requirements in Video Conferencing

EDA for teacher data

In order to more comprehensively understand the usage habits, technical support needs and functional preferences of college teachers in video conferencing teaching, this paper conducted exploratory data analysis (EDA) on questions 20, 25, 27 and 28 in the teacher questionnaire that are closely related to the prediction target. This step not only provides theoretical support for subsequent variable selection and cluster modeling, but also helps to reveal the heterogeneity of teacher-side needs and key areas of focus.

Q20: Distribution of needs for support areas in online teaching

Figure 4.5 shows the areas where teachers need extra help most in online teaching. The data shows that "Develop teaching resources" and "Improve teaching ability" are the two most frequently mentioned items, with 118 and 97 people choosing them respectively, accounting for 56% and 46% of the total sample. In addition, "Pursue personal research interests" was also mentioned by 84 people. This reflects that teachers still face the dual challenges of course content preparation and professional growth in the process of digital teaching transformation. This to a certain extent verifies the importance of Facilitating Conditions in the promotion of educational technology, and also suggests that platform developers should consider embedding teaching resource support modules in the tools to assist teachers in lesson preparation and resource sharing.

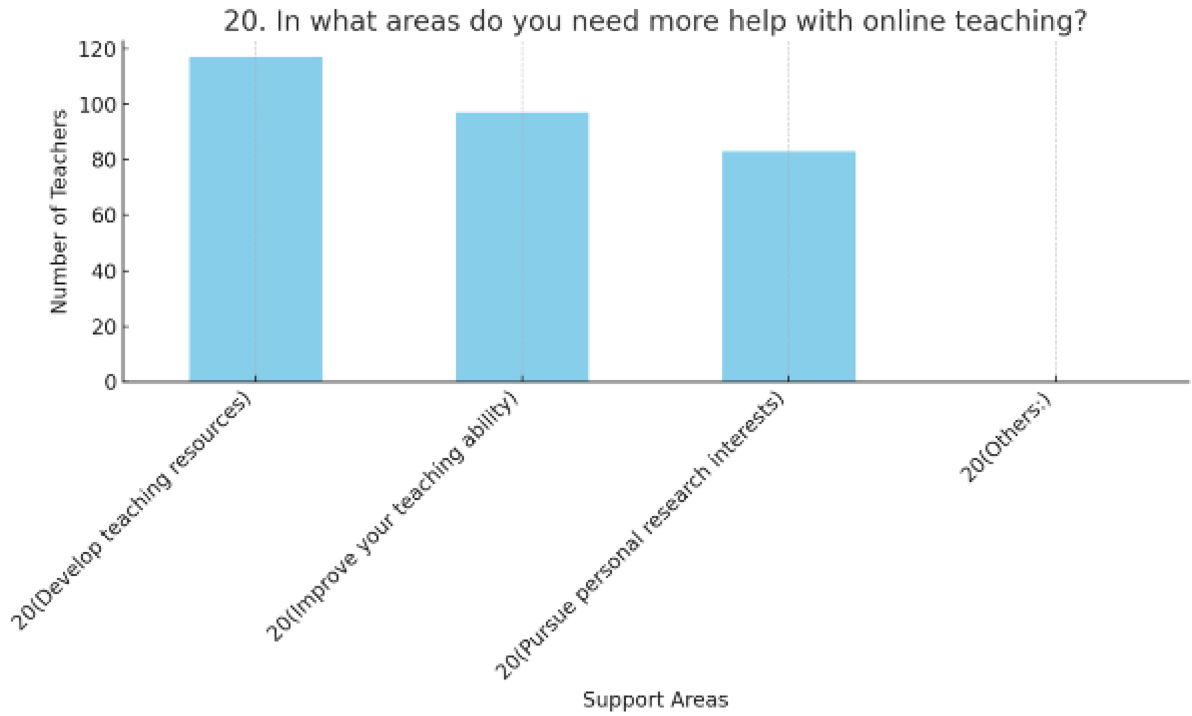

Figure 4.5 Q20: Distribution of needs for support areas in online teaching

Q25: Teachers' preference for visual design elements of video interfaces

Figure 4.6 shows the evaluation results of teachers on three types of visual design elements (animation effects, avatars, and interactive icons). "Avatar design" was favored by the most teachers, with a total of 176 people approving it, significantly higher than "animation effects" (98 people) and "interactive icons" (140 people). This result shows that compared with complex dynamic elements, teachers prefer clear, concrete, and personalized expressions.

This is consistent with the user's demand for interface stability and clarity in the teaching platform, and further echoes the dual influence of "perceived ease of use" (PEOU) and "perceived usefulness" (PU) on teacher behavior in the Technology Acceptance Model (TAM).

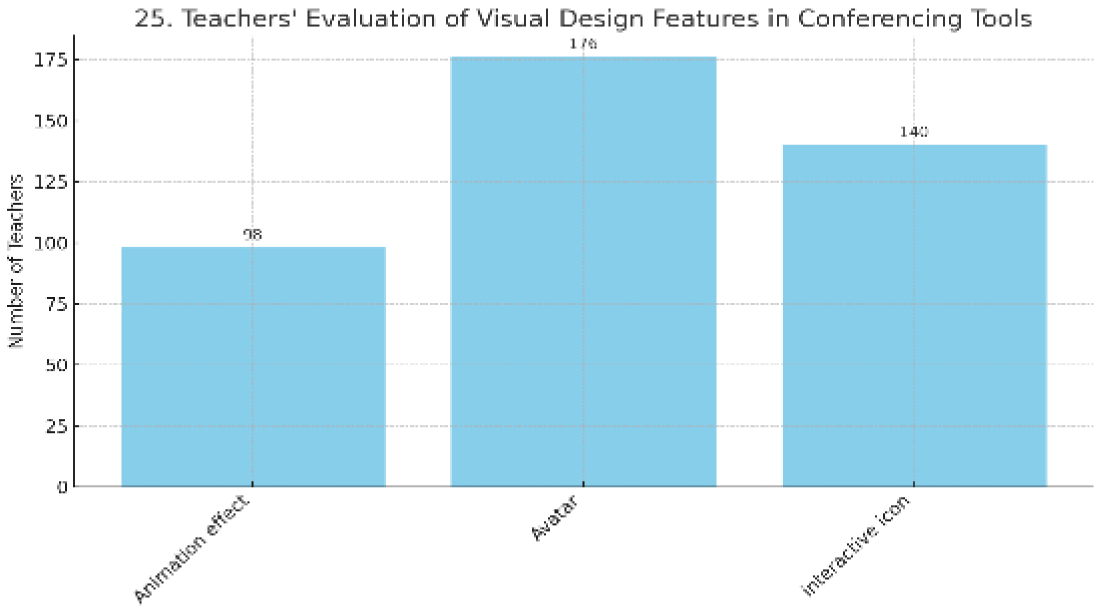

Figure 4.6 Q25: Teachers' preference for visual design elements of video interfaces

Q27: Teachers' attitude towards current video conferencing functions (such as group discussion)

Figure 4.7 is a Likert rating item, reflecting teachers' satisfaction with the "group discussion" function in current video conferencing tools. The data shows that most teachers are concentrated in the middle to low rating area, scoring 2 points (75 people) or 3 points (56 people), while only 26 people strongly agree (4 points). This shows that the current group discussion function has not yet fully met the needs of teachers in teaching scenarios.

This finding reveals the potential room for improvement in the interactive design of platform functions, and also shows that teachers have an "expectation gap" in the use of interactive functions, which is an important reference for subsequent product optimization.

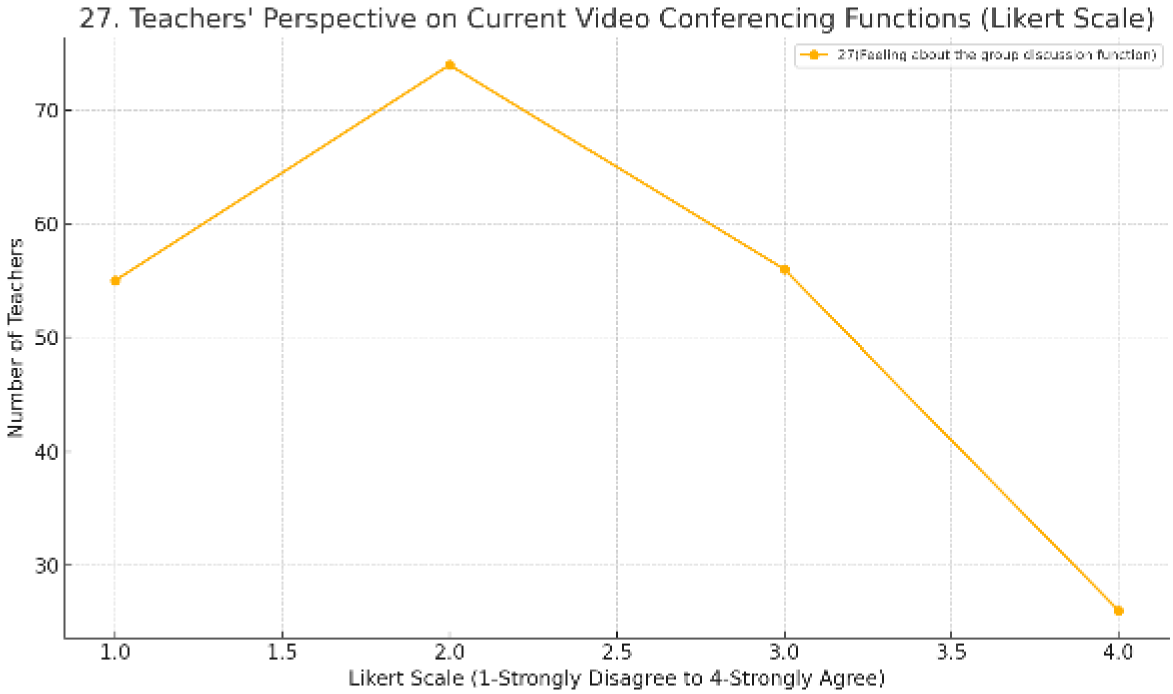

Figure 4.7 Q27: Teachers' attitude towards current video conferencing functions

Q28: Teachers' preferred video conference interactive functions

As shown in Figure 4.8, the three new functions most preferred by teachers are: "Personalized learning recommendation" (185 people), "Online homework correction" (136 people), and "Learning data analysis" (105 people). This shows that teachers have high expectations for personalized feedback mechanisms and teaching evaluation tools based on learning data.

From the perspective of the UTAUT model, this result confirms the driving force of "performance expectation", that is, teachers are more inclined to use platform functions that can improve teaching effectiveness.

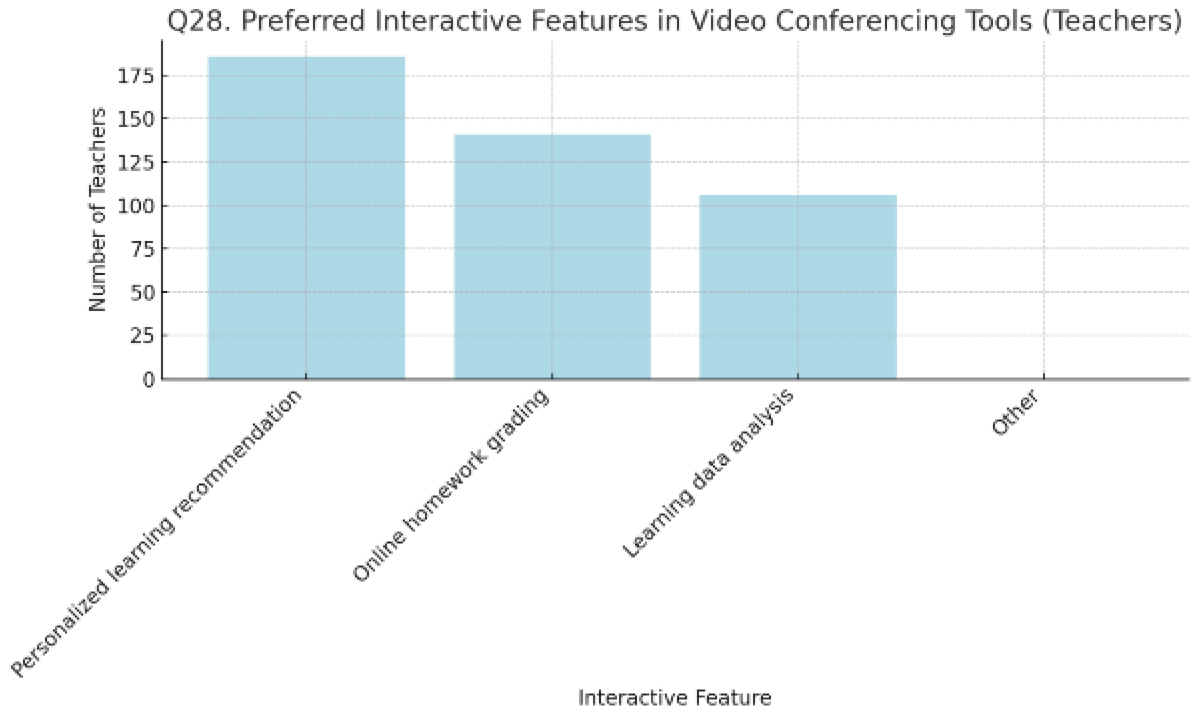

Figure 4.8 Q28: Teachers' preferred video conference interactive functions

Student Data Feature Selection and Variable Description

In order to improve the pertinence and explanatory power of the modeling, this study selected four questions in the questionnaire that are closely related to learning needs:

Q21: Proficiency in the use of online learning tools/platforms, covering the use of teacher platforms, learning software and social applications, reflecting students' technical familiarity;

Q22: Online learning behavior and comfort, including concentration, interactivity, group collaboration, coping ability, etc.;

Q23: Expected interactive functions in video conferencing, such as voice sharing, attention tracking, interactive quizzes, etc.;

Q27: Preference for additional functions in the video conferencing system, such as knowledge point reminders, real-time translation and smart notes, etc.

These characteristics comprehensively reflect students' digital literacy, learning behavior, functional preferences and potential expectations, which are highly consistent with the modeling goal of the study on "learning video conferencing needs".

K-Means modeling(Student Data)

In order to identify the potential learning demand patterns in the student population, the K-Means algorithm is used to perform cluster analysis on the above characteristic variables. In order to determine the optimal number of clusters, this study uses the Elbow Method to calculate the SSE (Sum of Squared Errors) under different K values. The results are shown in Figure 4.9:

There is an obvious "elbow" at K=3, indicating that dividing students into three categories is a more reasonable choice, which can take into account clustering accuracy and avoid overfitting.

Subsequently, the model is constructed with K=3, and two-dimensional visualization is performed through Principal Component Analysis (PCA). The results are shown in Figure 4.10: Figure 4.10 shows that the three types of students are clearly distributed in the principal component space, and the cluster boundaries are relatively clear, indicating that the selected features have good clustering discrimination.

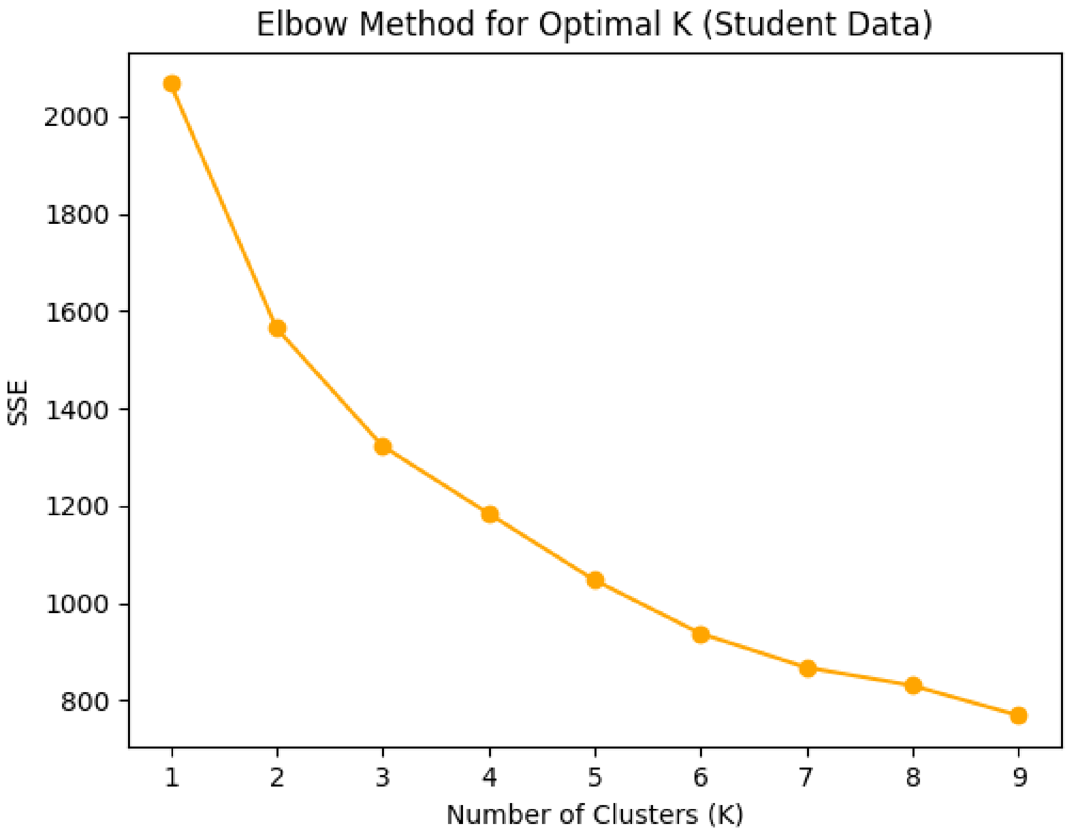

Figure 4.9 Elbow method for student data

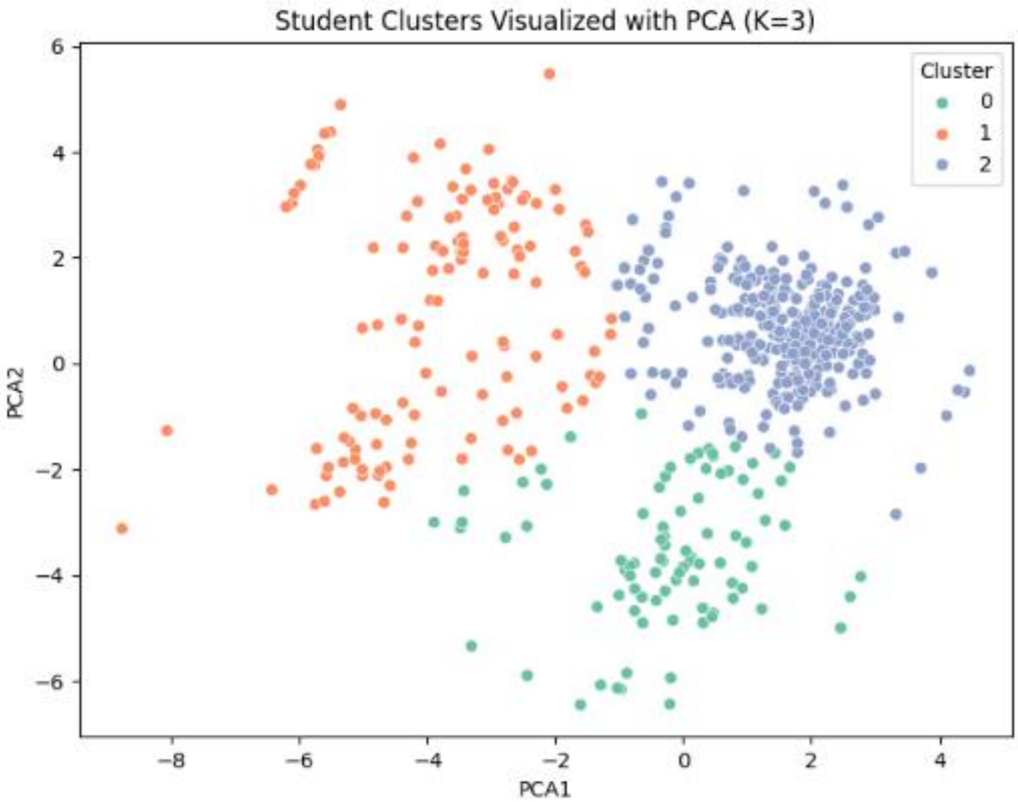

Figure 4.10 PCA for student data

Interpretation of Clustering Results(Student Data)

In order to give the clustering results interpretable semantic labels (Low / Medium / High Demand), this study calculated the mean of the variables in each cluster center. According to the score ranking, Cluster 2 showed the highest overall functional demand, and Cluster 1 showed the weakest demand. The three groups were labeled as "High Demand", "Medium Demand", and "Low Demand" respectively. This mapping relationship is used for semantic interpretation and color identification in subsequent heat maps and SHAP maps.

```
Cluster average demand score (from low to high):
 1   -0.572980
0   -0.232993
2    0.303693
```

Figure 4.11 student data cluster average demand score

The heat map (see Figure 4.12) shows the standardized scores of different demand groups on the Q21, Q22, Q23 and Q27 series variables, which clearly shows the significant differences in the three types of students in terms of functional acceptance, willingness to participate and emotional response.

High Demand student group (Cluster 2):

The scores in most functional dimensions are significantly higher than those of other groups,

especially in:
Q21: Proficient in using teaching platforms, learning software, and social tools;
Q22: More frequent use of interactive functions (such as raising hands, sharing screens, and expression tools);
Q23: More active participation in classroom interactions (such as turning on the camera, showing homework, and speaking);
Q27: More likely to show positive emotions (such as happiness and fulfillment) in post-class feedback.
This group shows a high degree of technical adaptability and learning initiative, and is an important target user for platform function innovation and application promotion.
Medium Demand student group (Cluster 0):
The scores of most dimensions are above average;
The performance in interactive use and teaching feedback dimensions is acceptable, but the participation behavior is not continuous enough;
The learning emotion dimension reflects a certain volatility.
This shows that this group has a certain foundation for platform use, but has not yet formed a high-frequency and deep use habit, and needs to improve participation stickiness through function optimization and teaching design.
Low Demand student group (Cluster 1):
The lowest score in almost all dimensions, especially in Q22 and Q23 (interactive functions and classroom participation);
The frequency of use of video conferencing functions is low, the participation is weak, and the learning emotion is relatively negative.
It is speculated that such students may have problems such as technical barriers, learning burnout or lack of motivation, and are the focus of educational intervention.

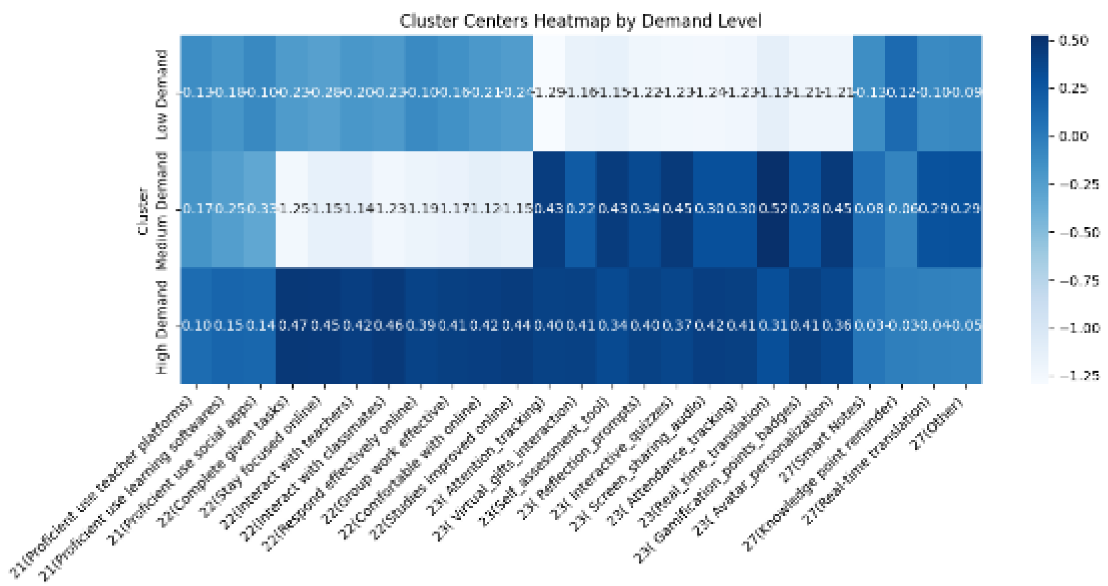

Figure 4.12 Heatmap for student data

SHAP-Based Feature Importance Analysis(student data)
Based on the cluster analysis in the previous section, this study further uses the XGBoost classification model and the SHAP interpretation mechanism to identify the key factors that affect the classification of students' video conferencing functional requirements

(high/medium/low). By interpreting the model output feature by feature, the importance of different variables to the demand prediction results is revealed, providing a basis for educators to optimize teaching design and platform functions.

Model Training and Interpretation Method

The clustering results (High/Medium/Low Demand) are used as the target variable, and the input features are the Likert scale scores under the four categories of questions Q21, Q22, Q23 and Q27. This study uses the XGBoost multi-classification model for modeling, and uses the SHAP value decomposition model output to calculate the average impact intensity of each feature on the prediction results of the three groups.

Figure 4.13 shows the average contribution value (mean(|SHAP value|)) of the top 10 influential features in the SHAP analysis in the three types of student groups. The horizontal axis represents the importance of the feature to the model prediction output, and the colors distinguish the three types of demand groups.

Identification and Interpretation of Key Influencing Factors

The results of the SHAP analysis show that the following variables have significant predictive power in distinguishing the intensity of student demand:

Interactivity and collaborative behavior (Q22 and Q23)

22(Interact with classmates) and 22(Interact with teachers) are the top-ranked factors, indicating that social interaction needs are the core variables affecting the demand level;

23(Screen sharing_audio) and 23(Interactive_quizzes) show that high-demand groups are more inclined to use the multimedia and interactive functions of the video conferencing platform;

The social functions and interactive support of the teaching platform are the key to stimulate students' high usage needs.

Attention maintenance and concentration

23(Attention_tracking) and 22(Stay focused online) significantly distinguish high-demand and low-demand groups;

Whether the platform has mechanisms to support attention management (such as prompts, reminders, and eye tracking) has a particularly significant impact on high-frequency users.

Learning experience and feedback perception

22(Comfortable with online) and 22(Complete given tasks) reflect the impact of platform experience quality on demand stratification from the perspectives of perceived convenience and task completion, respectively;

Middle and high-demand groups generally have a more positive experience of using the platform and are more efficient in completing tasks.

Visual/virtual customization

23(Avatar_personalization) reflects the appeal of gamification and personalization elements to some students (especially high-demand groups); although it is not a mainstream factor, it has a potential impact on specific subgroups and should be considered in function expansion.

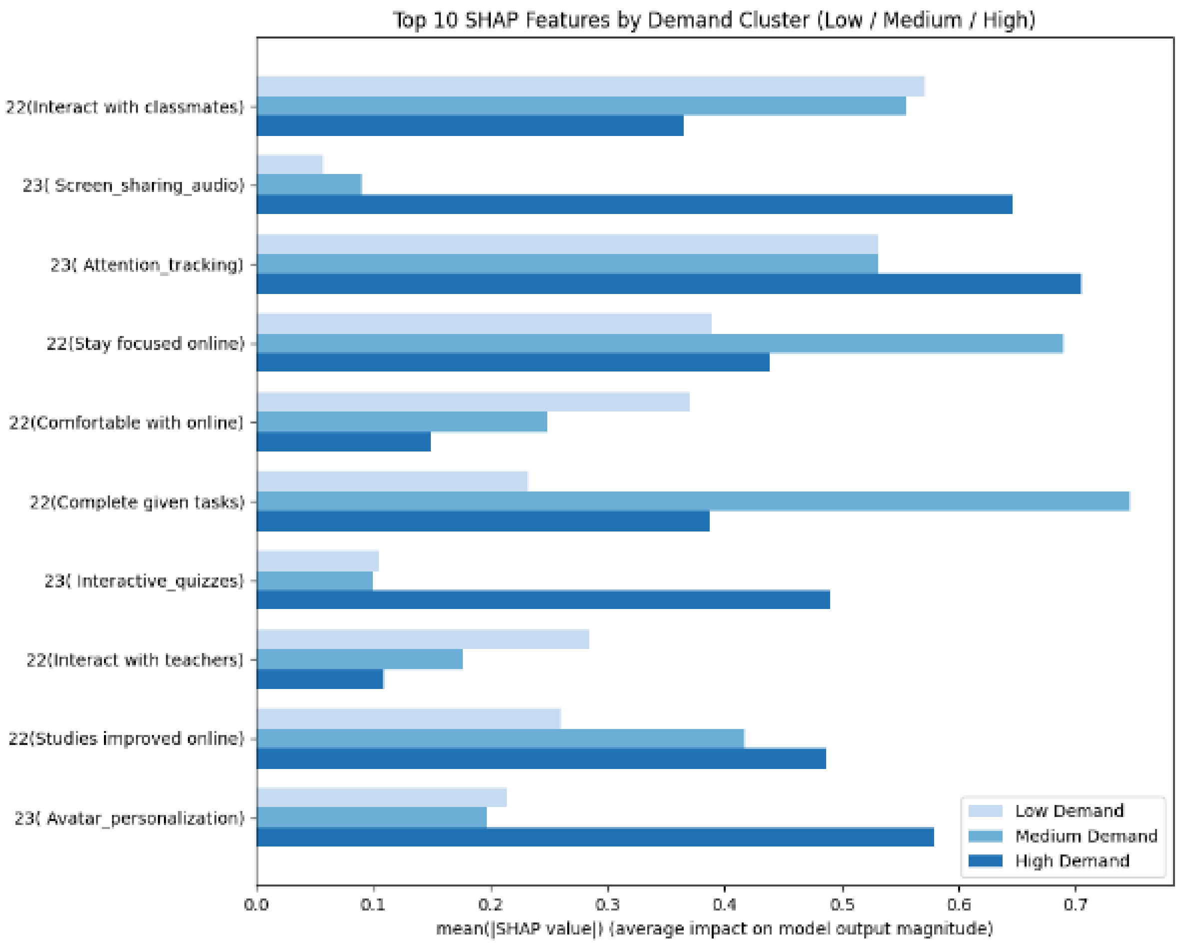

Figure 4.13 SHAP for student data

Teacher Data Feature Selection and Variable Description

In order to more accurately characterize teachers' support needs and functional preferences in video conference teaching, this study selected the following four clustering features from the questionnaire:

Q20 Teaching activity related variables: including multiple dimensions such as teaching resource development, homework design, personalized recommendation, collaborative evaluation, etc., which comprehensively reflect teachers' actual use and expectations of the platform's auxiliary teaching functions;
Q25 Teaching interface preference: covering visual and interactive elements such as animation effects, virtual images, and interactive icons, reflecting teachers' attitudes towards platform interface optimization;
Q27 Group discussion function evaluation: measuring teachers' functional satisfaction and usage feedback on the "group discussion" module in the platform, reflecting the demand for teaching interaction;
Q28 Learning data function requirements: such as learning analysis, automatic correction and personalized recommendation, are key variables to measure whether teachers have data-driven teaching awareness.
These characteristics comprehensively show the core concerns of teachers in the digital teaching process, covering both the use preferences of functional tools and their initiative

and technology acceptance in teaching innovation.

K-Means Modeling(Teacher Data)

In order to further identify the heterogeneous characteristics of the functional requirements of the teacher group in video conference teaching, this study also uses the K-Means clustering algorithm to perform unsupervised modeling on the above four types of variables. To determine the optimal number of clusters, the Elbow Method is first used to evaluate the sum of squared errors (SSE) corresponding to different K values, as shown in Figure 4.14:

Figure 4.11 shows that the SSE decrease rate slows down significantly near K=3, and an "elbow inflection point" appears, indicating that the three types of clustering schemes can strike a balance between complexity and fit, and are suitable for subsequent analysis.

Based on this, this study sets K=3 for modeling, and further uses Principal Component Analysis (PCA) to visualize the clustering results in two-dimensional space, as shown in Figure 4.15:
Figure 4.15 shows the distribution of teacher samples in the PCA dimensionality reduction space. The three types of cluster points show a certain degree of separation, especially in the direction of the first principal component. The difference is more obvious, which verifies the ability of variables to distinguish functional preferences.

This process not only improves the model's recognition accuracy of the heterogeneity of teacher needs, but also provides a solid foundation for the subsequent interpretation of feature contributions in combination with the SHAP method.

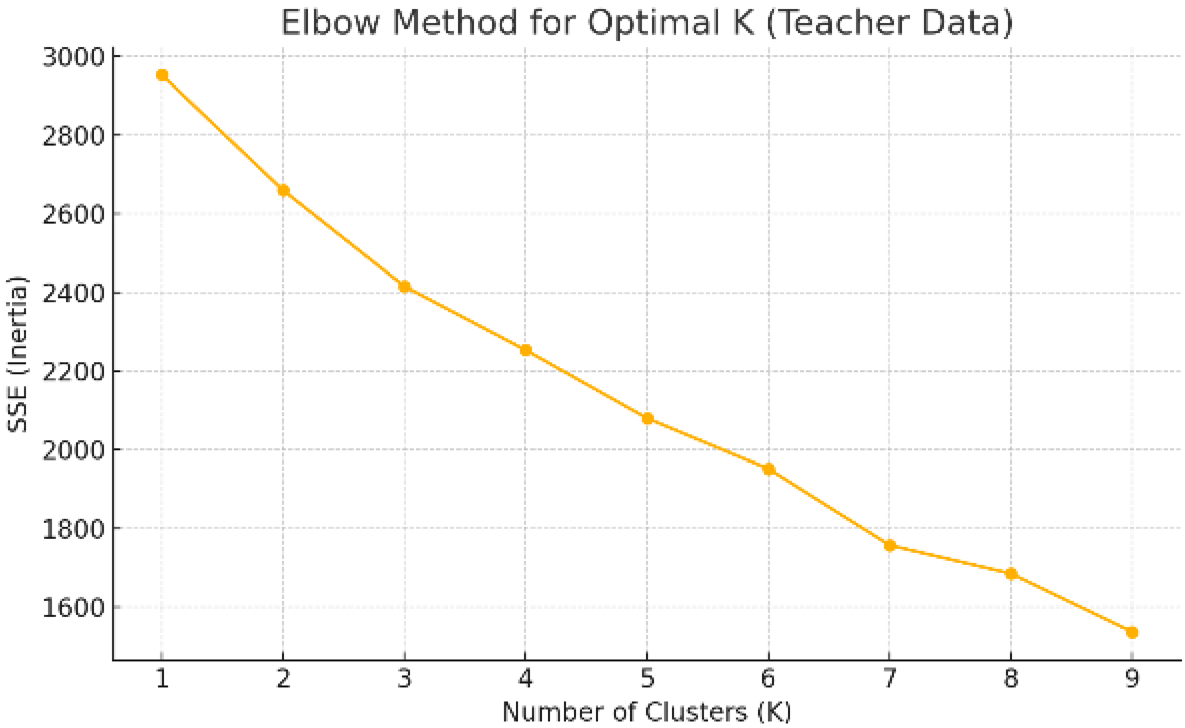

Figure 4.14 Elbow method for teacher data

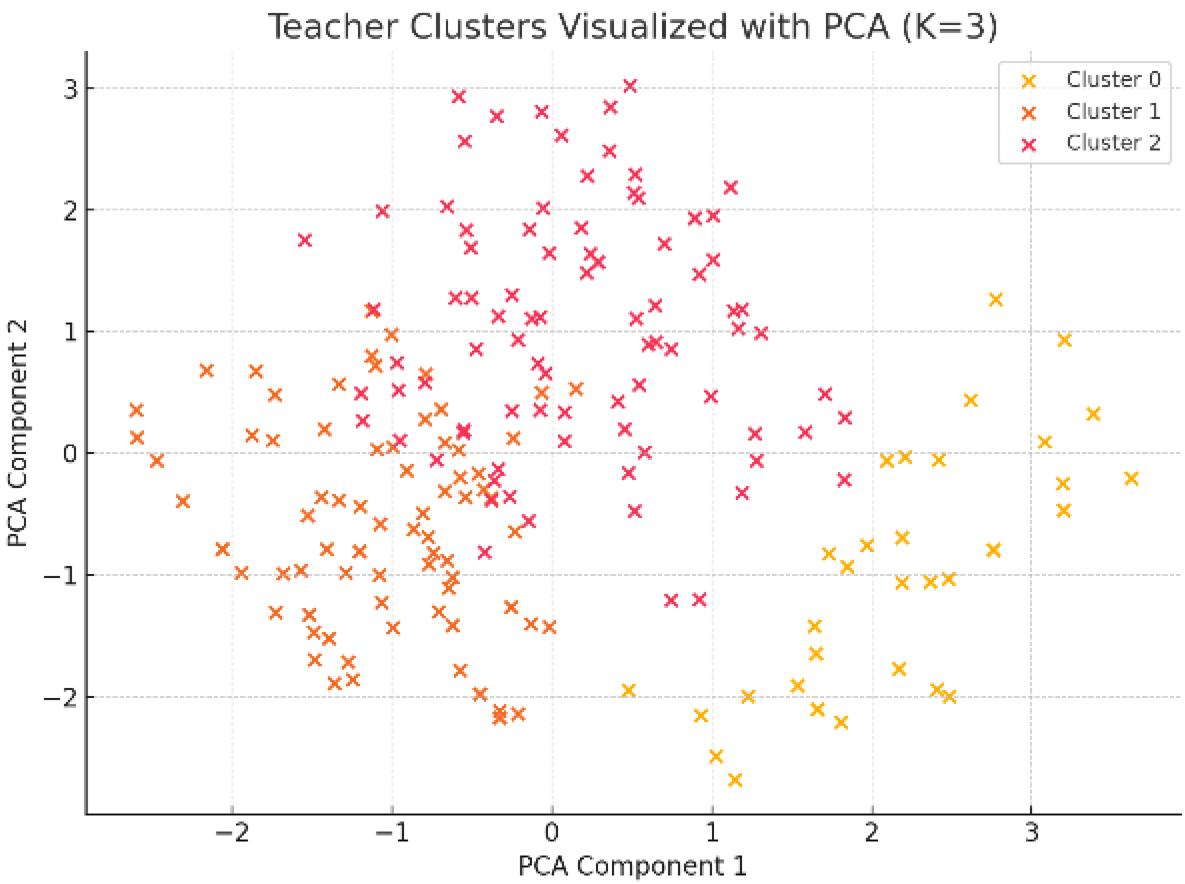

Figure 4.15 PCA for teacher data

Interpretation of Clustering Results(Teacher Data)

In order to identify the different functional demand types of teachers in video conference teaching, the study selected key items (Q20, Q25, Q27, Q28) in the teacher questionnaire to construct feature vectors, and divided the teacher group into three categories through K-Means clustering. The mean scores of the clustering results are as follows: According to the mean scores, they are sorted from low to high, and are defined as Low Demand (Cluster 2), Medium Demand (Cluster 1), and High Demand (Cluster 0).

```
Cluster
2    0.608163
1    0.615955
0    0.629870
```

Figure 4.16 teacher data cluster average demand score

The heat map (see Figure 4.17) shows the Q28 series (learning data analysis, personalized recommendation, online homework correction) is the core variable to distinguish the three types of teachers' demand levels. Especially in "personalized recommendation" and "learning data analysis", the high-demand group is significantly higher than other groups, indicating that they have stronger expectations for intelligent teaching auxiliary functions.

Question Q25 (virtual image elements) also maintains a high score in the high-demand group, showing that some teachers have higher requirements for classroom interactivity and attractiveness.

Question Q27 (group discussion function) has a lower score in the low-demand group, indicating that this type of teacher may prefer the traditional teaching mode and lack interest in or use habits of social interaction modules.

In contrast, the differences in "teaching resource development" and "homework evaluation" in the Q20 series (teaching behavior and platform support) are relatively flat among the three groups, indicating that they are common needs in teaching, but have limited role in distinguishing demand stratification.

Overall, the high-demand teacher group shows a significant preference for platform personalization, data-driven support, and interactive functions, while low-demand teachers are more concerned with the basic support functions of teaching. This type of demand portrait helps the education platform implement precise tiered services on the teacher side, such as pushing teaching data visualization modules and personalized recommendation functions to high-demand teachers, and providing teaching resource templates and operational training support to low-demand teachers.

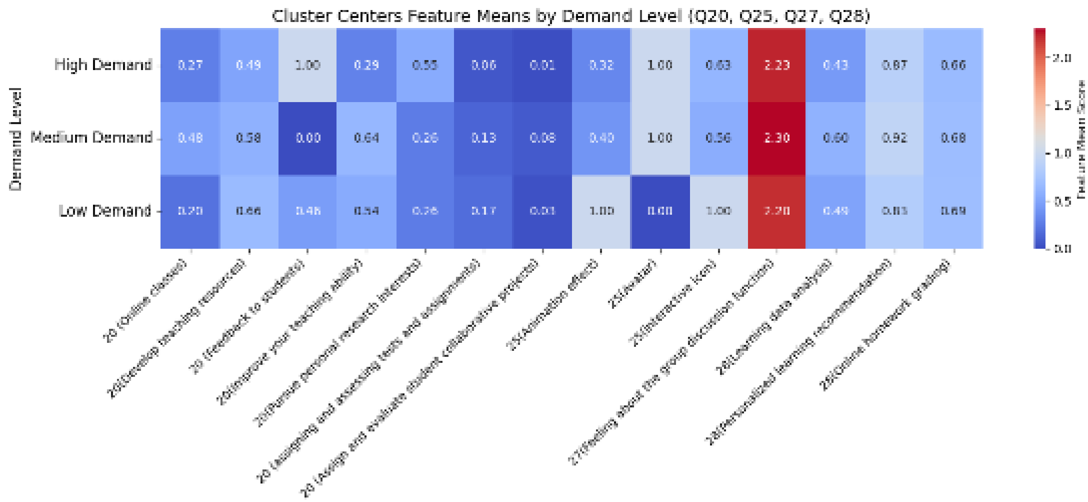

Figure 4.17 Heatmap for teacher data

SHAP-Based Feature Importance Analysis(teacher data)
To further explain the formation mechanism of functional preferences of different teacher demand categories, the study introduced SHAP value to perform explanatory analysis on the K-Means clustering model, and plotted the 10 most representative important variables in each group (see Figure 4.18). The horizontal axis is the importance of each feature in the model prediction on different groups (average absolute SHAP value), and the vertical axis is the specific feature name.

Figure 4.18 reveals the following important conclusions:
“20 (Feedback to students)” is the feature with the highest SHAP value in all teacher groups, especially in the low demand and medium demand groups. This shows that teachers’ attention to the teaching feedback mechanism is one of the main factors for the model to

distinguish different demand groups.

“25 (Avatar)” and “25 (Animation effect)” have greater influence in the medium and high demand groups, respectively, suggesting that interactivity and visual elements are gradually recognized by teachers in teaching, especially technology-sensitive groups.

Some teachers showed strong explanatory weights for research-related items (such as "20 (Pursue personal research interests)"), especially in the high-demand group, indicating that these teachers are more inclined to achieve the dual goals of self-research and teaching expansion through the platform.

Relatively speaking, the SHAP values of "28 (Personalized learning recommendation)" and "25 (interactive icon)" in each group are relatively low. Although these functions scored high in the questionnaire, they have limited group discrimination in the model, which may indicate that they are "consensus preferences" and are not enough to distinguish different types of teachers.

Overall, the SHAP analysis not only quantifies the influence of each functional variable in the model, but also intuitively reveals the key differences in functional perception and priority ranking among different levels of teacher needs. The study suggests that in product design, the feedback mechanism, virtual interaction and teacher research support modules can be optimized first to respond to the key expectations of the high-demand group, while retaining the basic teaching support functions to serve low-demand users, forming a hierarchical and refined functional layout strategy.

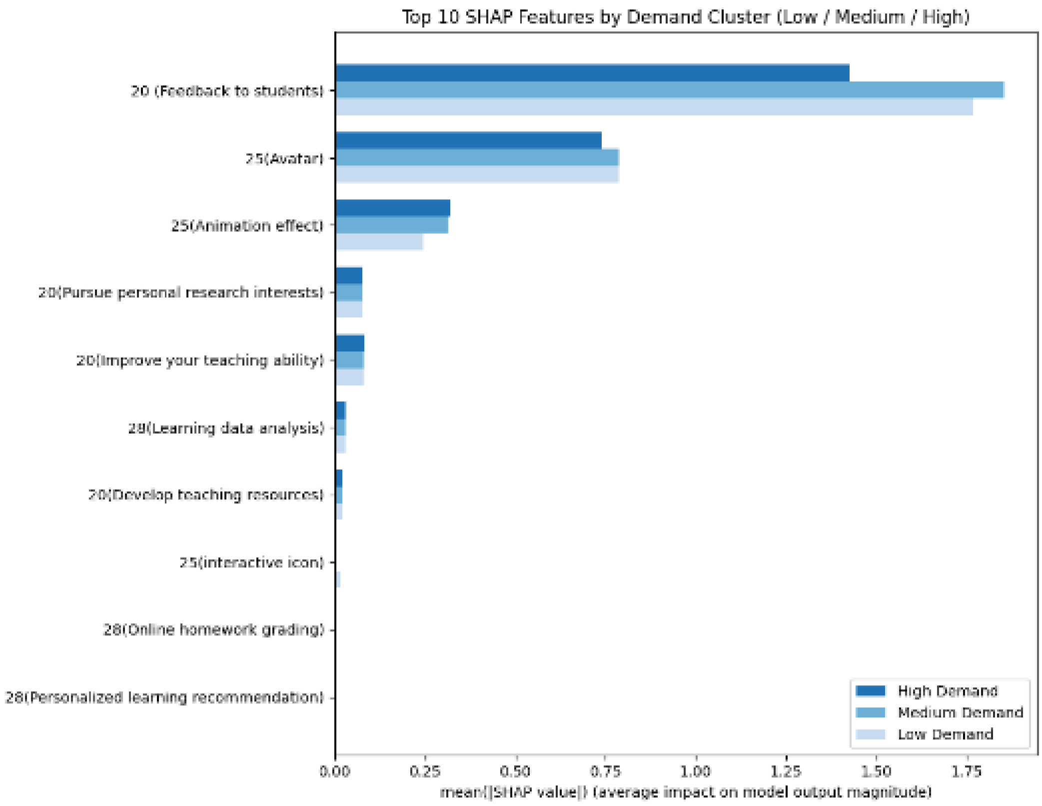

Figure 4.18 SHAP for teacher data

Conclusion and Implications

Research Conclusions

This study focuses on the video conferencing teaching scenario in the context of higher education in China, aiming to explore and identify the potential functional demand patterns of college students and teachers in distance learning through unsupervised learning methods. Compared with the traditional method that relies on subjective description or single indicator measurement, this study introduces the K-Means clustering algorithm and SHAP (SHapley Additive Explanations) explanation mechanism to construct a systematic and explanatory demand identification model, which enhances the practicality and explanatory power of the research.

The data comes from a total of 517 student questionnaires and 211 teacher questionnaires, covering multiple dimensional variables such as online learning behavior, platform usage habits, functional preferences and teaching feedback. In terms of feature selection, the study selected the items that best reflect the "functional usage needs" for students and teachers respectively, and constructed a feature matrix based on this for unsupervised clustering analysis. The optimal number of clusters was determined to be K=3 through the elbow method, and users were successfully divided into low-demand groups, medium-demand groups and high-demand groups, realizing user segmentation and demand map construction. Combined with the results of SHAP analysis, the study further analyzed the key influencing factors behind different clusters from the perspective of feature contribution, revealing the structural laws behind the differences in demand:

Among the student groups, the high-demand group showed significant dependence on "attention tracking", "task completion reminders", and "interactive collaboration with peers", indicating that they prefer structured management and social-driven online learning models; while the low-demand group paid more attention to situational factors such as "interface comfort" and "personalization of virtual images", reflecting their relatively weak motivation to participate in platform use.

Among the teacher groups, SHAP analysis showed that "giving students feedback" was the primary factor affecting demand clustering. In addition, "personalization of teaching platforms (such as virtual avatars, animation effects)" and "support for teaching resource configuration" also occupied a core position in the high-demand group, reflecting the strong demand of teachers for feedback efficiency, teaching situation optimization and intelligent resource distribution.

Expanded Implications for Educational Practice and Platform Construction

This study uses SHAP analysis to reveal the functional preferences and pain points of different groups in video conferencing teaching, and provides the following implications for multiple educational participants:

Implications for Platform Developers

Prioritize the development of core functional modules for interaction and monitoring

Students with high demand value functions such as "attention tracking", "task feedback" and "voice sharing" the most. The platform should give priority to the development of such functions at the bottom of the system to support real-time behavior perception and teacher-student interaction, thereby improving students' concentration and teachers' control.

Introduce lightweight entertainment design to stimulate the motivation of low-demand users
Low-demand student groups prefer personalized functions such as "virtual image" and "interface beautification". The platform can introduce gamification components (such as achievement badges, exclusive avatars) or visual theme customization to attract students with weak learning participation and improve their platform stickiness.

Integrate the teaching resource feedback system to improve the teacher's experience
The feature with the highest SHAP value on the teacher side is "giving students feedback". The platform should optimize modules such as "batch feedback", "voice comment entry", and "wrong question push" to help teachers improve their work efficiency and personalized guidance capabilities.

Create a visually friendly teaching interface to reduce fatigue
The high-demand group of teachers prefers visual tools such as animations and virtual images. The platform design needs to take into account both function and aesthetics to build an immersive teaching environment, thereby alleviating the pressure of online teaching and improving students' attention.

Build an "intelligent teaching analysis" module as an advanced function option
Although the average demand for functions such as "learning data analysis" and "personalized recommendation" is not high, it shows potential among high-demand teachers. It is recommended that the platform develop an optional intelligent analysis module for teachers with willingness and ability to use, and meet user needs in a layered manner.

Implications for education administrators
Promote hierarchical training of teachers' digital capabilities based on SHAP results
Different teachers have significant differences in their needs for feedback tools, animations, and data functions. Education management departments can carry out precise training based on demand clustering, set up technical operations, interactive strategies, and data literacy modules for different levels, and achieve personalized improvement of teaching capabilities.
Use clustering results to formulate differentiated teaching resource allocation strategies

For teachers with weak demand, more user-friendly operating platforms and standardized functional templates can be provided first; for teachers with high demand, advanced resources such as AI teaching assistant tools and personalized learning analysis can be configured to improve the overall resource utilization efficiency.

Promote cross-platform interoperability standards to achieve unified teaching experience
This study reveals that students pay attention to interaction and feedback consistency. Administrators should promote data interoperability and functional docking between different video conferencing platforms to reduce the cognitive burden and technical barriers caused by frequent tool switching between teachers and students.

Implications for Course Designers

Use functional demand clustering as a reference for the design of course participation mode
Different student groups have different preferences for interaction, reminders and feedback mechanisms. Course designers can integrate differentiated teaching strategies into the course based on clustering labels, such as group tasks, timed answering, and text and picture synchronization, so as to improve students' sense of participation and learning achievement in the platform.

Enhance the linkage mechanism between teachers and platforms in teaching feedback
Teachers hope that the "feedback" function will be more efficient and convenient. Course design can encourage the platform to automatically generate student performance reports after class, assist teachers in timely individual tutoring, recommend review tasks, and achieve "data-driven feedback".

Implications for Policymakers

Incorporate the stratification of teacher and student needs into the perspective of intelligent education policy design
This study provides a path to identify differentiated needs for educational technology based on unsupervised learning, which can be used as a basis for user portraits in the process of intelligent education policy formulation, and support the more accurate implementation of measures such as classified subsidies, function popularization, and capacity building.
Encourage universities to introduce explainable AI tools to improve teaching transparency
The explainability of SHAP analysis can be used in university teaching management and teaching evaluation, such as as a platform upgrade evaluation standard and a functional satisfaction tracking indicator, which will help policymakers promote the implementation of "explainable AI" at the institutional level.

Theoretical and Contextual Contributions

This research makes significant theoretical and contextual contributions to the fields of educational technology and learning analytics. Theoretically, it moves beyond the established paradigms of the Technology Acceptance Model (TAM) and UATAUT, which primarily explain whether users adopt technology. By employing unsupervised machine learning, this study shifts the focus from adoption to differentiated needs, uncovering the latent structure of user preferences that are not captured by traditional survey means or outcome-based supervised models. It demonstrates how explainable AI (XAI) techniques like SHAP can bridge the gap between black-box predictive models and actionable educational insights, thereby enriching the theoretical understanding of user behavior as a multi-faceted, segmentable construct rather than a monolithic outcome.

Contextually, this study provides a crucial, data-driven framework for addressing the specific challenges of digital transformation in Chinese higher education. The rapid, large-scale shift to mobile learning necessitated by national circumstances created a diverse and often underserved user base. This research offers a granular map of this landscape, identifying distinct user archetypes—from the high-demand, interaction-seeking student to the feedback-focused, data-oriented teacher. For Chinese platform developers (e.g., Tencent Conference, DingTalk), these findings provide a clear blueprint for prioritizing feature development to enhance market competitiveness and user satisfaction. For administrators and policymakers in China, the study offers a methodology to move beyond one-size-fits-all solutions, enabling more efficient and equitable resource allocation and targeted teacher training programs that align with the national agenda for educational modernization and equity. Therefore, this work serves as both a methodological guide and a practical handbook for building more intelligent, personalized, and effective mobile learning ecosystems within the unique context of modern China.

References

Bailey, D. R., Almusharraf, N., & Almusharraf, A. (2022). Video conferencing in the e-learning context: explaining learning outcome with the technology acceptance model. *Educ Inf Technol (Dordr)*, *27*(6), 7679-7698. https://doi.org/10.1007/s10639-022-10949-1

Guo, S., Halim, H. B. A., & Saad, M. R. B. M. (2025). Leveraging AI-enabled mobile learning platforms to enhance the effectiveness of English teaching in universities. *Scientific Reports*, *15*(1), 15873. https://doi.org/10.1038/s41598-025-00801-0

Ho, I. M. K., Cheong, K. Y., & Weldon, A. (2021). Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. *PLOS ONE*, *16*(4), e0249423. https://doi.org/10.1371/journal.pone.0249423

Khor, E. T. (2022). A data mining approach using machine learning algorithms for early detection of low-performing students. *International Journal of Information and Learning Technology*, *39*(2), 122-132. https://doi.org/10.1108/ijilt-09-2021-0144

Li, D. M. (2024). An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm. *Computers in Human Behavior*, *157*, Article 108211. https://doi.org/10.1016/j.chb.2024.108211

Liang, H., & Wang, S. (2022). An Online Service Performance Prediction Learning Method. *International Journal of Grid and High Performance Computing*, *14*(1). https://doi.org/10.4018/ijghpc.301577

Luo, R., Wang, J., & Wang, Y. (2023). Undergraduate students' perceptions of using videoconferencing for EFL learning: Evidence from Tencent Meeting application. *Heliyon*, *9*(12), e22993. https://doi.org/10.1016/j.heliyon.2023.e22993

Mei, L. F., Hu, R. C., Cao, H. W., Liu, Y., Han, Z. F., Li, F., & Li, J. (2020). Realtime mobile bandwidth prediction using LSTM neural network and Bayesian fusion. *Computer Networks*, *182*, Article 107515. https://doi.org/10.1016/j.comnet.2020.107515

Miranda, E., Aryuni, M., Rahmawati, M. I., Hiererra, S. E., & Dian Sano, A. V. (2024). Machine learning's model-agnostic interpretability on the prediction of students' academic performance in video-conference-assisted online learning during the covid-19 pandemic. *Computers and Education: Artificial Intelligence*, *7*, 100312. https://doi.org/https://doi.org/10.1016/j.caeai.2024.100312

Nadzinski, G., Gerazov, B., Zlatinov, S., Kartalov, T., Dimitrovska, M. M., Gjoreski, H., Chavdarov, R., Kokolanski, Z., Atanasov, I., Horstmann, J., Sterle, U., & Gams, M. (2023). Data Science and Machine Learning Teaching Practices with Focus on Vocational Education and Training. *Informatics in Education*, *22*(4), 671-690. https://doi.org/10.15388/infedu.2023.28

Tahiru, F., Parbanath, S., & Agbesi, S. (2023). Machine Learning-based Predictive Systems in Higher Education: A Bibliometric Analysis. *Journal of Scientometric Research*, *12*(2), 436-447. https://doi.org/10.5530/jscires.12.2.040