Comparing Machine Learning Models for Classifying Online Learning Satisfaction: A Supervised and Unsupervised Approach

Srisharvin A/L Meroiviran [1], Noor Hasliza Md Saad[2], Normaiza Binti Mohamad[1*], Farhad Nadi[1,3*]

[1]Faculty of Artificial Intelligence and Frontier Technologies, UNITAR International University, 47301 Petaling Jaya, Selangor, Malaysia

[2]School of Management, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia

***Corresponding Authors**

DOI: https://doi.org/10.47772/IJRISS.2026.100601210

Received: 24 June 2026; Accepted: 29 June 2026; Published: 15 July 2026

ABSTRACT

The rapid expansion of online education has made learner satisfaction a central measure of platform quality, yet many institutions still rely on end-of-course surveys that capture feedback too late to support timely intervention. This study examines whether supervised and unsupervised machine learning can predict and explain student satisfaction in online learning using real learner feedback. The publicly available Online Education System Review dataset from Kaggle, comprising 1,033 learner records, was used as the basis for the analysis. Four supervised classifiers, namely Random Forest, Naive Bayes, Logistic Regression, and K-Nearest Neighbours, were trained in WEKA to classify satisfaction into three levels, while Simple K-Means clustering was applied to uncover natural learner segments. Random Forest delivered the strongest and most balanced performance, with an accuracy of approximately 63.5 percent and a weighted ROC area of 0.757, followed closely by Naive Bayes and Logistic Regression, whereas K-Nearest Neighbours performed noticeably worse at about 51.6 percent. Across the supervised models, instructor support, interaction frequency, content quality, and platform usability emerged as the most influential predictors of satisfaction. Clustering revealed distinct learner groups that differed in engagement and demographic profile while sharing comparable satisfaction tendencies. The results indicate that ensemble classification combined with clustering provides a practical, interpretable, and reproducible approach for monitoring satisfaction and informing course design. The study contributes to learning analytics by showing how accessible tools can convert routine learner feedback into evidence for improving digital education, while acknowledging the limits of single-snapshot survey data.

Keywords: online learning satisfaction; machine learning; learning analytics; Random Forest; predictive modelling

INTRODUCTION

Digital platforms have moved from being a supplement to traditional instruction to becoming a primary channel for teaching and learning in both academic and corporate settings. This shift accelerated during recent global disruptions and has widened access for learners who were previously separated by geography, cost, or time (Tu et al., 2020). At the same time, the move online has exposed a persistent difficulty. Keeping learners engaged, motivated, and satisfied is harder when instruction is asynchronous and self-paced, and when the immediate social cues of a physical classroom are absent.

Satisfaction in online learning is best understood as a multidimensional construct that is shaped by the quality of instruction, the usability of the platform, the opportunities for interaction, and the degree of learner control. Several studies report that timely feedback, easy navigation, and the perceived relevance of the material are closely tied to how satisfied learners feel (Sulisworo et al., 2021). When these conditions are not met, learners

often disengage, and dissatisfaction has been linked to reduced persistence and higher dropout, a problem that remains visible in large-scale open courses.

A recurring limitation in practice concerns the way satisfaction is measured. Institutions have long depended on end-of-course surveys, which tend to yield static and delayed information. These instruments capture opinions only after the experience has ended, and they rarely surface the behavioural patterns that precede dissatisfaction. As a consequence, educators are frequently left reacting to problems rather than anticipating them, and the data that could guide evidence-based improvement is often too coarse to act upon.

Machine learning offers a route from reactive evaluation toward proactive support. Supervised algorithms can learn from learner feedback and behavioural indicators in order to classify satisfaction levels, while unsupervised methods can reveal natural groupings of learners that are not obvious from survey totals alone. Earlier work has shown that behavioural data combined with machine learning can predict learner performance and satisfaction with useful accuracy (Yuan et al., 2024), and that satisfaction itself functions as a strong mediator of continued platform use (Deng et al., 2023). These findings suggest that predictive modelling can generate insight that is both academically meaningful and practically actionable.

This paper investigates how predictive modelling can be used to analyse and anticipate online learning satisfaction using real learner feedback. Drawing on the publicly available Online Education System Review dataset, the study applies four supervised classifiers and one clustering algorithm to identify the factors most strongly associated with satisfaction and to segment learners by their feedback and engagement patterns. The aim is not only to predict satisfaction but to produce interpretable findings that course designers, educators, and educational technology developers can translate into concrete improvements.

The contribution of the study is both methodological and practical. Methodologically, it treats satisfaction as a primary target for prediction and pairs supervised classification with unsupervised segmentation, so that the analysis explains what drives satisfaction and also reveals how different learners experience the same platform. Practically, it relies on accessible, widely used tools and modest hardware, which makes the workflow straightforward to replicate. By moving from opinion-driven, end-of-course evaluation toward a quantitative and repeatable model, the study aims to support continuous, evidence-based improvement of online learning rather than one-off assessment.

LITERATURE REVIEW

From Traditional to Digital Learning

The delivery of education has shifted steadily from structured, in-person formats toward flexible and on-demand digital provision. This change grants learners greater autonomy, but it also transfers more responsibility for motivation and pace onto the learner. Where the traditional classroom sustains attention through direct supervision and routine, online environments can produce irregular participation and, in some cases, withdrawal when motivation and satisfaction are not adequately supported (Addas et al., 2024). The central concern, therefore, is not access to technology but the systematic measurement and improvement of satisfaction across diverse platforms.

Engagement and Feedback

Engagement is widely treated as a measurable signal of quality in online learning, and institutions commonly track proxies such as login frequency, time on task, and completion of materials. Cinar et al. (2024) note that such indicators provide an objective picture of activity, yet they capture little of the emotional or cognitive involvement that sustains learning over time. Feedback quality appears to be especially important. Prompt, specific feedback is associated with stronger self-regulation and higher satisfaction (Covrig et al., 2023), whereas delayed or non-actionable feedback tends to demoralise learners even when surface activity remains high (Desai

et al., 2024). These observations imply that engagement analytics should consider the quality of interaction, not only its volume.

Retention, Completion, and Performance

Satisfaction is closely connected to retention and completion. Learners who perceive their online experience as clear, responsive, and relevant are more likely to persist, while unmet expectations are linked to disengagement and dropout. Although game-like elements such as badges and progress indicators have been proposed as one means of sustaining motivation (Hassan et al., 2025), the broader evidence suggests that satisfaction depends on a combination of factors, including content clarity, instructor responsiveness, peer interaction, and platform usability, rather than on any single feature. Academic performance is also bound up with satisfaction, since learners who feel supported tend to study more consistently and submit work on time (Joseph, 2024). In digital settings, usability and timely feedback mediate this relationship directly, because the ability to identify and correct mistakes in real time strengthens both confidence and progress (Lampropoulos & Sidiropoulos, 2024). Adaptive, data-driven platforms that provide progressive feedback have been associated with stronger learning, particularly in technical subjects (Naseer et al., 2025).

Artificial Intelligence in Learning Analytics and the Research Gap

Across these themes, artificial intelligence and machine learning are increasingly used to convert raw interaction data into predictive insight. Models such as logistic regression, decision trees, and ensemble methods can estimate the risk of disengagement and the likelihood of dissatisfaction, enabling earlier intervention (Yeganeh et al., 2025). What remains comparatively underexplored is the treatment of satisfaction as a primary outcome, predicted from real feedback data at a scale that supports institutional decision-making. Much existing work either centres on general engagement and performance or relies on small-scale or self-report studies whose findings are difficult to generalise. The present study addresses this gap by modelling satisfaction directly with both supervised and unsupervised techniques on an open, reasonably sized dataset, and by reporting interpretable predictors that practitioners can use.

MATERIALS AND METHODS

Research Design

The study adopts a quantitative and predictive design. Rather than relying on descriptive survey summaries, it uses machine learning to estimate satisfaction levels from demographic and behavioural variables and to surface relationships that conventional statistics may not reveal. The design combines supervised learning (Logistic Regression, Random Forest, Naive Bayes, and K-Nearest Neighbours) with unsupervised learning (Simple K-Means clustering), so that the analysis both predicts satisfaction for individual learners and identifies meaningful subgroups within the population. Strong preprocessing, k-fold cross-validation, and supporting visualisation were built into the design to reduce bias, limit overfitting, and keep the workflow transparent and reproducible.

Dataset and Sampling

The analysis draws on the Online Education System Review dataset, an openly available collection on Kaggle (Sujaradha, 2021). The dataset records structured feedback from 1,033 learners across a range of demographic and educational backgrounds, with variables covering course ratings, instructor quality, device usage, study habits, and overall satisfaction. The study uses secondary data analysis, working directly with this existing dataset rather than collecting new primary responses. This approach is efficient, draws on genuine learner feedback, and avoids the ethical exposure associated with new data collection.

The sample is reasonably balanced and supports subgroup comparison. As summarised in Table 1, the data include 614 male and 419 female learners. The age distribution is concentrated among younger learners, with the largest groups aged 20, 19, and 18, alongside a smaller tail extending to age 40. Learners also vary in

education level, from school leavers and diploma holders to undergraduate and postgraduate students, which allows the models to learn patterns across a range of learner conditions rather than a single profile.

Characteristic	Category	Number of learners
Gender	Male	614
Gender	Female	419
Age (largest groups)	20 years	249
Age (largest groups)	19 years	226
Age (largest groups)	18 years	220
Total instances	All learners	1,033

Table 1. Selected demographic characteristics of the Online Education System Review dataset.

Tools and Environment

WEKA version 3.9.6 served as the primary analytical engine for preprocessing, attribute selection, classification, and clustering, since its graphical interface makes the workflow accessible to practitioners without a programming background. Python, with the Pandas, Matplotlib, and Seaborn libraries, was used selectively to generate additional visual outputs such as a correlation heatmap and a device-usage chart. Microsoft Excel supported initial inspection and validation of the raw comma-separated files. All work was carried out on a standard laptop (Intel Core i5-10300H, 8 GB RAM, Windows 11 64-bit), which demonstrates that the approach does not require specialised or costly computing resources.

Data Preparation and Preprocessing

The raw dataset was imported into the WEKA Explorer, where its structure and attribute distributions were inspected. Missing values were addressed using the ReplaceMissingValues filter, which substitutes the attribute mean for numeric fields and the mode for nominal fields, so that the algorithms received complete and consistent input. Duplicate records were located through sorting and filtering and then removed, which prevented repeated instances from being over-weighted during training. Categorical variables, such as gender and education level, were encoded into numeric form, and numeric attributes were standardised where appropriate so that differences in scale would not distort the models. A final visual inspection of the cleaned data confirmed that the structure was intact and ready for modelling.

Feature Selection, Training, and Validation

Attribute evaluators available in WEKA, including Information Gain and correlation-based ranking, were used to identify the variables that contributed most to predicting satisfaction. The four supervised classifiers were then trained on the prepared data, with algorithm-specific parameters tuned across repeated runs, for example the number of trees in Random Forest, set to 100, and the number of neighbours in K-Nearest Neighbours, set to one in the reported configuration. Model performance was assessed using ten-fold stratified cross-validation, in which the data are divided into ten balanced folds and each fold serves once as the test set. This procedure provides a more realistic estimate of how the models would perform on new learner data and reduces the risk of overfitting to a single train-test split.

Models were evaluated with standard classification metrics. Accuracy measures the overall proportion of correct predictions, while precision and recall describe how well each satisfaction class is handled. The F-measure combines precision and recall into a single value, and the confusion matrix reports the distribution of true and false predictions across classes. The ROC area summarises the ability of a model to separate classes, with values

above 0.70 generally regarded as reasonable and values above 0.80 as strong in educational data mining. Simple K-Means clustering was then applied to the cleaned features to group learners by natural similarity, adding an exploratory layer that complements the supervised models.

Ethical Considerations

Because the study uses an openly published and already anonymised dataset, it carries minimal ethical risk. The data contain no personally identifiable information, results are reported only in aggregate form, and no attempt was made to re-identify any participant. Attention was also given to fairness during modelling, including the use of cross-validation and a balanced treatment of demographic groups, so that predictions would not systematically over-represent or under-represent any subgroup. All preprocessing and modelling steps were documented to support transparency and replication.

RESULTS AND DISCUSSION

Overview and Model Comparison

All four classifiers were trained to predict satisfaction across three levels (Average, Bad, and Good) using ten-fold stratified cross-validation on 1,033 instances. Three of the models produced broadly similar accuracy in the low-sixties, while the fourth lagged behind. Random Forest classified about 63.50 percent of instances correctly, Naive Bayes 63.31 percent, and Logistic Regression 62.63 percent, whereas K-Nearest Neighbours reached only 51.60 percent. Table 2 summarises the comparison across accuracy, the Kappa statistic, and the weighted F-measure, ROC area, and PRC area.

Model	Accuracy (%)	Kappa	F-measure (wtd.)	ROC area (wtd.)	PRC area (wtd.)
Random Forest	63.50	0.3695	0.624	0.757	0.654
Naive Bayes	63.31	0.3795	0.628	0.757	0.637
Logistic Regression	62.63	0.3476	0.615	0.745	0.621
K-Nearest Neighbours	51.60	0.1904	0.511	0.593	0.454

Table 2. Comparison of supervised learning models based on ten-fold cross-validation in WEKA.

The three leading models performed at a comparable level, which suggests that the dataset contains consistent patterns that different algorithms can detect. Naive Bayes recorded the highest Kappa and weighted F-measure by a small margin and trained quickly, which makes it attractive where computing resources are limited. Logistic Regression offered transparent, interpretable coefficients. Random Forest combined competitive accuracy with the strongest weighted PRC area and a useful by-product, namely a ranking of feature importance, and was therefore selected as the most practical model overall. K-Nearest Neighbours, configured with a single neighbour, struggled with the overlapping and subjective nature of survey responses and is best read as a baseline that highlights the advantage of the other methods. Figure 1 illustrates the accuracy gap, and Figure 2 compares the per-class ROC area across the four models. The following subsections report each model in turn.

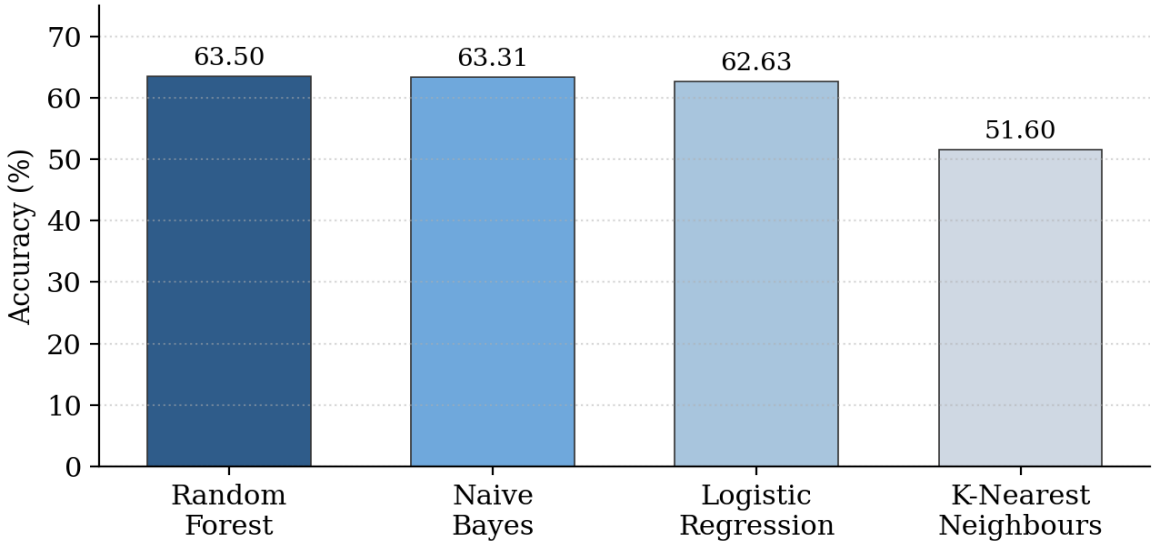

Figure 1. Classification accuracy of the four supervised models (ten-fold cross-validation).

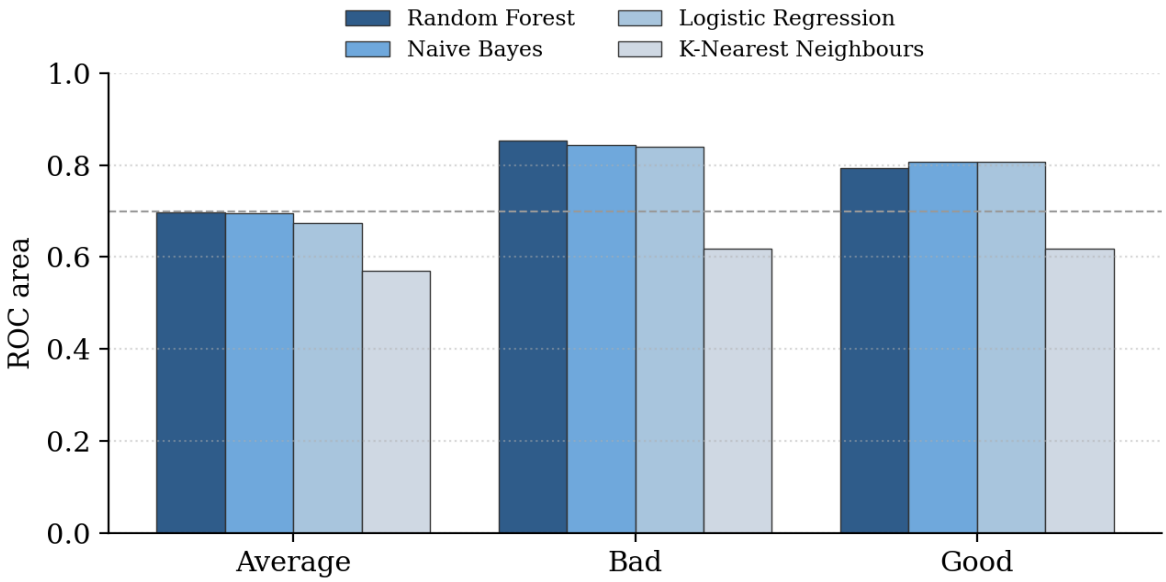

Figure 2. ROC area by satisfaction class for each model; the dashed line marks the 0.70 reference.

Random Forest

The Random Forest model was run with 100 trees under ten-fold cross-validation. Its overall results are summarised in Table 3. The model classified 655 of 1,033 instances correctly, a Kappa of 0.3695 indicates fair agreement beyond chance, and the error measures sit at a moderate level for survey data of this kind.

Metric	Value
Correctly classified	655 (63.50%)
Incorrectly classified	377 (36.50%)
Kappa statistic	0.3695
Mean absolute error	0.3279

Metric	Value
Root mean squared error	0.4012
Relative absolute error	80.31%
Root relative squared error	88.81%
Total instances	1,033

Table 3. Stratified cross-validation summary for the Random Forest model.

Class-level behaviour is reported in Table 4. The model recognised the Average class most readily, with a true positive rate of 0.782, but was more conservative on the Good class, where recall fell to 0.398 despite reasonable precision. Discrimination was strongest for the Bad class, which reached a ROC area of 0.854. This pattern indicates that the model is comparatively reliable at distinguishing clearly dissatisfied learners, which is valuable because these are precisely the learners who may benefit most from early support.

Metric	Average	Bad	Good	Weighted avg.
TP rate	0.782	0.552	0.398	0.635
FP rate	0.488	0.096	0.078	0.297
Precision	0.638	0.636	0.621	0.634
Recall	0.782	0.552	0.398	0.635
F-measure	0.703	0.591	0.485	0.624
MCC	0.306	0.480	0.379	0.364
ROC area	0.697	0.854	0.793	0.757
PRC area	0.680	0.655	0.596	0.654

Table 4. Detailed accuracy by class for the Random Forest model.

The confusion matrix in Table 5 makes the same behaviour concrete. Most errors involve learners whose true label is Bad or Good being assigned to the Average category, which is consistent with the tendency of survey responses to cluster around the middle of a satisfaction scale.

Actual / Predicted	Average	Bad	Good
Average	423	63	55
Bad	102	133	6
Good	138	13	100

Table 5. Confusion matrix for the Random Forest model.

Naive Bayes

Naive Bayes produced almost the same headline accuracy as Random Forest, classifying 654 instances correctly, and it returned the highest Kappa of the four models at 0.3795 (Table 6). Its main advantage is speed and simplicity, which suits categorical and mixed survey data.

Metric	Value
Correctly classified	654 (63.31%)
Incorrectly classified	379 (36.69%)
Kappa statistic	0.3795
Mean absolute error	0.2761
Root mean squared error	0.4234
Relative absolute error	67.63%
Root relative squared error	93.71%
Total instances	1,033

Table 6. Stratified cross-validation summary for the Naive Bayes model.

At the class level, Naive Bayes was strongest at separating dissatisfied learners, with a ROC area of 0.843 for the Bad class and 0.807 for the Good class, giving a weighted ROC area of 0.757 that matches Random Forest. Precision was 0.646 for Average, 0.573 for Bad, and 0.661 for Good, while recall was highest for the Average class at 0.738 and lower for the Good class at 0.466. The confusion matrix in Table 7 again shows that the most common errors are Bad and Good cases drawn toward the Average category.

Actual / Predicted	Average	Bad	Good
Average	399	88	54
Bad	97	138	6
Good	122	12	117

Table 7. Confusion matrix for the Naive Bayes model.

Logistic Regression

Logistic Regression classified 647 instances correctly for an accuracy of 62.63 percent, with a Kappa of 0.3476 (Table 8). Its appeal lies in interpretability, since the model exposes which attributes carry the most weight in the prediction, which helps institutions act on the results without a specialised technical team.

Metric	Value
Correctly classified	647 (62.63%)
Incorrectly classified	386 (37.37%)
Kappa statistic	0.3476
Mean absolute error	0.3156
Root mean squared error	0.4078
Relative absolute error	77.30%
Root relative squared error	90.27%
Total instances	1,033

Table 8. Stratified cross-validation summary for the Logistic Regression model.

The model discriminated the Bad class well, with a ROC area of 0.839, and the Good class with 0.808, for a weighted ROC area of 0.745. Precision reached 0.669 for the Good class and 0.605 for the Bad class, while recall was highest for the Average class at 0.786. As Table 9 shows, the error pattern mirrors the other models, with middle-of-scale predictions absorbing most of the misclassifications.

Actual / Predicted	Average	Bad	Good
Average	425	64	52
Bad	129	107	5
Good	130	6	115

Table 9. Confusion matrix for the Logistic Regression model.

K-Nearest Neighbours

K-Nearest Neighbours, implemented as IBk with a single neighbour, classified 533 instances correctly for an accuracy of 51.60 percent and a Kappa of 0.1904 (Table 10). The high root relative squared error reflects the difficulty the method had with overlapping responses, where nearby points often carried different labels.

Metric	Value
Correctly classified	533 (51.60%)
Incorrectly classified	500 (48.40%)
Kappa statistic	0.1904
Mean absolute error	0.3231
Root mean squared error	0.5671
Relative absolute error	79.13%
Root relative squared error	125.54%
Total instances	1,033

Table 10. Stratified cross-validation summary for the K-Nearest Neighbours model.

Class-level scores were modest, with ROC areas between 0.570 and 0.619 and precision of 0.443 for the Bad class against a low recall of 0.390. The confusion matrix in Table 11 confirms weaker separation than the other models. K-Nearest Neighbours is therefore most useful here as a comparison point that reinforces the value of ensemble and probabilistic methods on this kind of data.

Actual / Predicted	Average	Bad	Good
Average	342	53	106
Bad	122	94	25
Good	129	25	97

Table 11. Confusion matrix for the K-Nearest Neighbours model.

Learner Segments from Clustering

Simple K-Means clustering divided the learners into five segments. Rather than predicting a label, the clusters describe how groups of learners differ in their characteristics. Table 12 presents selected centroid values for the most interpretable attributes. Across all clusters, learners predominantly attended classes on laptops, reported a middle economic status, and recorded an Average satisfaction level on the output class, so these constant attributes are omitted from the table. The clusters differ mainly in age, study time, the frequency of clearing doubts with faculty online, and self-reported performance. The segment with the highest tendency to clear doubts with faculty also reported the highest performance, which is consistent with the view that interaction with instructors is associated with stronger outcomes.

Feature	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Age (years)	19.80	19.61	19.31	19.66	20.12
Study time (hours)	6.95	7.03	6.78	7.11	6.88
Clearing doubts (online)	2.30	2.41	2.65	2.89	3.19
Performance (online)	6.70	6.30	6.27	6.64	6.97
Satisfaction level	Average	Average	Average	Average	Average
Instances	250	164	126	194	203

Table 12. Selected cluster centroids from Simple K-Means clustering (constant attributes omitted).

Supporting Analysis in Python

Two additional outputs were produced in Python to corroborate the WEKA results. A correlation heatmap of the numeric features indicated that interaction-related variables move together. The strongest relationship was between online interaction and clearing doubts with faculty, at roughly 0.72, suggesting that learners who interact actively are also more likely to seek clarification from instructors. Online interaction also showed a moderate association of about 0.56 with performance. These patterns reinforce the feature-importance signal from Random Forest, namely that engagement and instructor interaction are meaningful correlates of satisfaction and success.

A device-usage chart showed that most learners accessed online classes on a laptop, at approximately 65.1 percent, followed by mobile devices at 32.3 percent and desktops at 2.6 percent, as shown in Figure 3. Device type is relevant because material that is not adapted to smaller screens may reduce comfort and, by extension, satisfaction for the substantial minority of learners on mobile devices. This descriptive finding offers a straightforward and actionable target for responsive course design.

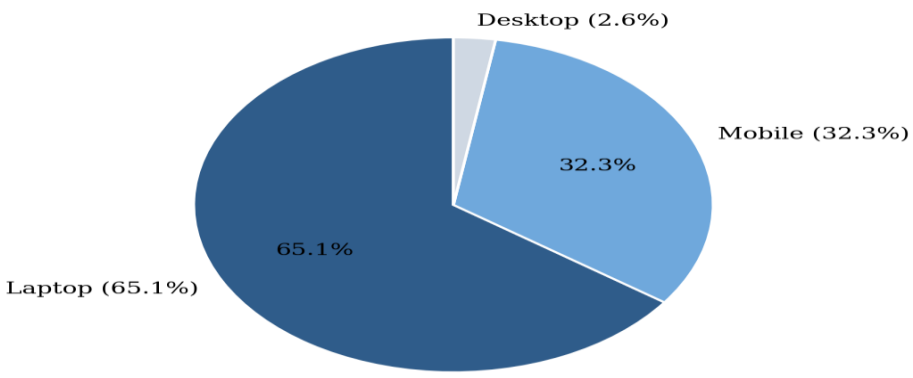

Figure 3. Distribution of devices used by learners to attend online classes.

DISCUSSION

Taken together, the supervised and unsupervised results offer a fuller picture than either approach would on its own. Classification estimates satisfaction and ranks its drivers, while clustering exposes how combinations of characteristics distinguish one group of learners from another. The convergence of evidence across WEKA models and Python visualisations increases confidence that the patterns reflect genuine structure in the data rather than artefacts of a single method. The practical implication is that institutions can identify learners at risk of low satisfaction earlier and can direct improvements toward the factors that matter most, particularly instructor support, interaction, content clarity, and usability. The headline accuracy in the low-sixties is reasonable for early-stage educational data, in which subjective and overlapping responses place a natural ceiling on separability, and it suggests clear scope for improvement through richer features and more advanced algorithms.

Practical Implications

The findings translate into several concrete courses of action for institutions and educators. Because the models can flag learners who are likely to report low satisfaction, they support early intervention rather than after-the-fact evaluation. An instructor who knows part way through a course which learners share the characteristics associated with dissatisfaction can reach out sooner, adjust pacing, or offer additional support before disengagement becomes entrenched. The value of the approach lies less in a precise score for any single learner than in directing limited teaching attention toward those who are most likely to benefit from it.

The analysis also points toward responsive course and platform design. The factors most consistently associated with satisfaction, namely instructor support, frequency of interaction, clarity of content, and ease of use, are all elements that designers can influence directly. Where the clustering reveals segments that differ in engagement and study patterns, the same course can be adapted to serve them better, for example by varying the amount of guided interaction or the structure of feedback opportunities. The device usage pattern observed in the data, in which laptops dominate but a substantial share of access occurs on mobile devices, reinforces the case for designing materials that remain usable across screen sizes rather than assuming a single mode of access.

Finally, the low cost of the approach is itself a practical contribution. Because the workflow relies on accessible, widely used tools and modest hardware, it is within reach of institutions and individual educators who do not have dedicated data science teams or specialised infrastructure. This lowers the barrier to adopting predictive analytics and makes it more feasible to embed continuous, evidence-based review into ordinary teaching practice rather than reserving it for well-resourced programmes.

CONCLUSION

This study set out to predict and explain online learning satisfaction using machine learning applied to real learner feedback. Working with the Online Education System Review dataset, it trained four supervised classifiers and one clustering algorithm, evaluated them with cross-validation, and triangulated the findings with Python visualisations. Random Forest emerged as the most practical model, pairing competitive accuracy of about 63.5 percent with interpretable feature importance and reliable identification of clearly dissatisfied learners. Naive Bayes and Logistic Regression performed at a similar level and offer speed and transparency respectively, while K-Nearest Neighbours served mainly as a baseline. Clustering added depth by revealing learner segments that differ in engagement and demographic profile.

The central contribution is methodological as much as empirical. The study shows that routine, openly available learner feedback can be transformed into evidence for action using accessible tools and modest hardware, which lowers the barrier for institutions and individual educators to adopt predictive analytics. The findings consistently point to instructor support, interaction frequency, content quality, and platform usability as the levers most associated with satisfaction, and they suggest that early, targeted intervention with at-risk learners is both feasible and worthwhile.

Several limitations qualify these results. The analysis relied on the variables available in a single dataset, which did not include richer behavioural traces such as time spent on individual modules or discussion activity. It also rested on a single snapshot of learner feedback, whereas satisfaction is likely to change over the course of a programme. Accuracy in the low-sixties, while acceptable for an early-stage study, leaves clear room for improvement. Future work could incorporate longitudinal and learning-management-system data, test more advanced algorithms such as gradient boosting or neural networks, validate the models on larger and more diverse populations, and integrate predictions into a live dashboard for instructors. Throughout, the responsible and transparent use of artificial intelligence, with attention to privacy and fairness, should remain central so that predictive tools support rather than replace human judgement in education.

Declarations

Ethical Approval

This research used a publicly available and already anonymised secondary dataset that contains no personally identifiable information. No new data were collected from human participants, and all results are reported in aggregate form, in line with institutional requirements for the responsible handling of educational data.

Conflict of Interest

The authors declare that there is no conflict of interest associated with this work.

Data Availability

The dataset analysed in this study, the Online Education System Review, is publicly available on Kaggle (Sujaradha, 2021) at https://www.kaggle.com/datasets/sujaradha/online-education-system-review.

REFERENCES

1. Addas, A., Naseer, F., Tahir, M., & Muhammad, N. K. (2024). Enhancing higher-education governance through telepresence robots and gamification: Strategies for sustainable practices in the AI-driven digital era. Education Sciences, 14(12), 1324. https://doi.org/10.3390/educsci14121324
2. Cinar, O. E., Rafferty, K., Cutting, D., & Wang, H. (2024). AI-powered VR for enhanced learning compared to traditional methods. Electronics, 13(23), 4787. https://doi.org/10.3390/electronics13234787
3. Covrig, M., Goia, S. I., Igret, R. S., Marinas, C. V., Miron, A. D., & Roman, M. (2023). Students' engagement and motivation in gamified learning. Amfiteatru Economic, 25(Special Issue 17), 1003–1023. https://doi.org/10.24818/EA/2023/S17/1003
4. Deng, P., Chen, B., & Wang, L. (2023). Predicting students' continued intention to use e-learning platform for college English study: The mediating effect of e-satisfaction and habit. Frontiers in Psychology, 14, Article 1182980. https://doi.org/10.3389/fpsyg.2023.1182980
5. Desai, S. P., Munshi, M. M., Hanji, S. V., & Pabba, C. (2024). Impact of the learning environmental factors on group engagement: A deep learning approach. South Asian Journal of Management, 31(6), 150–173. https://doi.org/10.62206/sajm.31.6.2024.150-173
6. Hassan, B., Muhammad, O. R., Siddiqi, Y., Muhammad, F. W., & Rabiya, A. S. (2025). CONNECT: An AI-powered solution for student authentication and engagement in cross-cultural digital learning environments. Computers, 14(3), 77. https://doi.org/10.3390/computers14030077
7. Joseph, A. R. (2024). The impact of personalized learning on intrinsic motivation, academic proficiency, and behavioral issues [Doctoral dissertation. ProQuest Dissertations and Theses (Order, (31146874)).
8. Lampropoulos, G., & Sidiropoulos, A. (2024). Impact of gamification on students' learning outcomes and academic performance: A longitudinal study comparing online, traditional, and gamified learning. Education Sciences, 14(4), 367. https://doi.org/10.3390/educsci14040367

9. Naseer, F., Muhammad, N. K., Addas, A., Awais, Q., & Ayub, N. (2025). Game mechanics and artificial intelligence personalization: A framework for adaptive learning systems. Education Sciences, 15(3), 301. https://doi.org/10.3390/educsci15030301
10. Sujaradha, G. (2021). Online education system review [Data set. Kaggle. https://www.kaggle.com/datasets/sujaradha/online-education-system-review
11. Sulisworo, D., Wulandari, Y., Effendi, M. S., & Alias, M. (2021). Exploring the online learning response to predict students' satisfaction. Journal of Physics: Conference Series, 012117. https://doi.org/10.1088/1742-6596/1783/1/012117
12. Tu, Y., Chen, W., & Brinton, C. G. (2020). A deep learning approach to behavior-based learner modeling. https://doi.org/10.48550/arXiv.2001.08328
13. Yeganeh, L. N., Nicole, S. F., Chen, Y., Simpson, A., & Hatami, M. (2025). The future of education: A multi-layered metaverse classroom model for immersive and inclusive learning. Future Internet, 17(2), 63. https://doi.org/10.3390/fi17020063
14. Yuan, J., Qiu, X., Wu, J., Guo, J., Li, W., & Wang, Y. (2024). Integrating behavior analysis with machine learning to predict online learning performance: A scientometric review and empirical study. https://doi.org/10.48550/arXiv.2406.11847