

Machine Learning-Enabled Quality Analytics for Higher Education: A Multi-Dimensional Classification and Decision-Support Framework

Tripti Singh 

Research Scholar,
Department of Computer Science and Engineering,
Dr.K.N.Modi University,
Newari, Rajasthan, India
 triptisinghcsc@gmail.com
<https://orcid.org/0009-0002-6049-9459>



Publication History: Manuscript Reference No: IRJCS/RS/Vol.13/Issue09/SPCS10081

Research Article | Open Access | Double-Blind Peer-Reviewed| Article ID: IRJCS/RS/Vol.13/Issue 09/CS6.SPCS10081

Received: 12, August, Revised: 03, September 2026, Accepted: 14, September 2026, Published Online: 30, September 2026.

<https://www.irjcs.com/volumes/Vol13/iss-09/01.SPCS10081.pdf>

Article Citation: Tripti(2026), Machine Learning-Enabled Quality Analytics for Higher Education: A Multi-Dimensional Classification and Decision-Support Framework, IRJCS: International Research Journal of Computer Science, Volume 13, Issue 09 of 2026 pages 642-651 **Doi:** <https://doi.org/10.26562/irjcs.2026.v1309.01> **BibTeX Key:** Tripti@Machine2026.

IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, <https://doi.org/10.26562/irjcs.2026.v1309.01> The journal's official abbreviation is IRJCS.

ORCID: <https://orcid.org/0009-0004-9398-7488> About the License: Copyright©2026 copyright by the authors. This article is an open access and license under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Higher education institutions generate heterogeneous information through curriculum processes, teaching-learning activities, learning resources, student support, research and extension, internal quality assurance, and learner engagement. Converting these heterogeneous signals into actionable quality intelligence is a computationally challenging classification problem because educational quality is multidimensional, noisy, partially observed, and context-dependent. This paper proposes and experimentally evaluates a machine-learning quality analytics pipeline to classify higher-education quality states. We constructed a computational dataset of 500 synthetic student-level observations using seven normalized quality dimensions. Three supervised learning algorithms Logistic Regression, Decision Tree and Random Forest were evaluated using a stratified 70:30 hold-out protocol and five-fold cross-validation. Performance was assessed through accuracy, precision, recall, F1-score and ROC-AUC. Random Forest achieved 69.33% hold-out accuracy and an F1-score of 79.46%, while Logistic Regression achieved 69.33% accuracy and 79.28% F1. We also used five-fold cross-validation to assess performance stability. Feature-importance analysis identified the relative contribution of the quality dimensions, and an ablation experiment examined the effect of removing the most influential predictor. The study also develops a probabilistic feedback-validation formulation and an outcome-oriented Faculty Effectiveness Index architecture for integration into a quality-management pipeline. The results demonstrate the feasibility of treating institutional quality management as a computational decision-support problem rather than as a collection of isolated descriptive indicators. Because the numerical dataset is synthetic, the reported performance values are methodological demonstrations rather than estimates of any real institutional population. The framework provides a technically explicit basis for subsequent validation with ethically collected institutional data.

Keywords: higher education; machine learning; educational data mining; quality analytics; classification; Random Forest; Logistic Regression; Decision Tree; feedback validation; quality assurance; decision support

1. INTRODUCTION

Quality assurance in higher education increasingly operates within a data-rich environment. Institutions collect information on student progression, assessment, attendance, feedback, curriculum delivery, learning resources, research activity, support services, and internal quality assurance. The computational problem is to transform these heterogeneous observations into reliable evidence that can support timely academic decisions. Conventional descriptive statistics remain important, but they do not always expose nonlinear interactions, probabilistic uncertainty, hidden patterns, or combinations of indicators associated with different quality states. Machine learning offers a complementary analytical layer. In a supervised classification setting, a system learns a mapping from a feature vector X to an outcome Y . For educational quality analytics, X can represent curriculum, teaching-learning, infrastructure, support, research, internal quality assurance and engagement indicators, while Y can represent a defined quality state. The usefulness of such a system depends not only on predictive performance but also on measurement validity, interpretability, error costs, fairness, data governance and institutional context. Recent learning-analytics literature demonstrates the expanding use of computational methods for educational decision support. Systematic reviews have examined learning analytics in distance education and the design of learning-analytics dashboards, while broader reviews describe AI applications across higher education.

These studies also indicate that analytics should connect to learning and institutional action rather than be treated simply as visualization. The computational framework developed here treats quality management as a pipeline consisting of data acquisition, preprocessing, feature construction, probabilistic validation, model training, model evaluation, interpretability analysis, decision support and feedback into institutional improvement. The approach is deliberately multi-dimensional. Rather than relying on a single satisfaction score to represent quality, it combines several domains that describe the educational process. The main contributions are: (i) a formal representation of higher-education quality as a multi-feature classification problem; (ii) an experimental comparison of three supervised classifiers; (iii) five-fold cross-validation in addition to a hold-out evaluation; (iv) feature importance and ablation analyses; (v) a probabilistic feedback-validation formulation; and (vi) a practical architecture for integrating predictive analytics into continuous quality management.

2. RELATED WORK

2.1 Educational data mining

Educational data mining applies statistical and computational methods to educational datasets to discover patterns, predict outcomes and support interventions. Researchers have used classification, clustering, association analysis, and sequential modelling to study student performance, engagement, and progression. A quality-management application differs from a conventional student-performance prediction task because its target is not necessarily a single examination result. The target may instead represent a composite institutional process state.

2.2 Learning analytics in higher education

Learning analytics has expanded from descriptive dashboards toward predictive and prescriptive decision support. Systematic evidence shows substantial research activity around analytics in online and distance education, including student engagement and dashboard design. The computational challenge is to preserve indicators' educational meaning while making them machine-readable. A technically accurate model can still be educationally inappropriate if its target variable is poorly defined.

2.3 Artificial intelligence and institutional quality

AI applications in higher education include automated assessment, recommendation and student support, learning analytics, institutional analytics and quality monitoring. Recent systematic reviews identify opportunities to improve educational quality but also emphasize ethical, organizational, and infrastructural constraints. UNESCO's guidance on AI in education emphasizes human-centered use, privacy, validation and institutional preparedness.

2.4 Research gap

Four technical gaps are addressed here. First, many quality discussions remain descriptive rather than predictive. Second, different quality dimensions are often treated independently despite their interaction. Third, model evaluation is sometimes reported through a single accuracy value without robustness analysis. Fourth, computational outputs are not always connected to explicit feedback-control architecture. This study addresses these gaps through multi-feature classification, cross-validation, error analysis, feature importance and a feedback-oriented quality-management pipeline.

3. PROBLEM FORMULATION

Let each learner record be represented by a vector $X_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$, where $p = 7$ quality dimensions, and the target $Y_i \in \{0,1\}$ denotes a lower-quality synthetic state (0) or a higher-quality synthetic state (1) for the learning problem $f(X_i) \rightarrow Y_i$. Utilising probabilistic classification, the Naive Bayes formulation assumes conditional independence of predictors given the class via $P(Y = c|X) \propto P(Y = c) \prod_j P(X_j|Y = c)$ —a computationally attractive yet potentially unrealistic baseline for correlated educational indicators like curriculum, teaching quality, and engagement. To build a robust quality-management index, a weighted effectiveness index is formulated as $FEI = \frac{\sum W_i F_i}{\sum W_i}$ using feedback scores F_i and probability-based validity weights W_i to discount inconsistent observations. These mechanics are further evaluated alongside supervised models such as Logistic Regression ($P(Y = 1|X) = \frac{1}{1+e^{-z}}$ where $z = \beta_0 + \sum \beta_j X_j$), Decision Trees (using split criteria like Gini impurity), and Random Forests (ensemble aggregations), with model performance comprehensively assessed via accuracy, precision, recall, F1-score, and ROC-AUC to prevent class-specific error concealment.

4. DATA AND EXPERIMENTAL METHODOLOGY

4.1 Experimental dataset

The experiment uses 500 computationally generated observations. Each observation contains seven continuous predictors on a 0–100 scale. The dimensions are Curriculum, Teaching-Learning, Infrastructure, Student Support, Research/Extension, IQA/Innovation and Engagement. A fixed random seed of 42 ensures reproducibility.

Table 1 Feature specification

Feature	Type	Range	Analytical role
Curriculum	Continuous	0–100	Predictor
Teaching-Learning	Continuous	0–100	Predictor
Infrastructure	Continuous	0–100	Predictor
Student Support	Continuous	0–100	Predictor
Research/Extension	Continuous	0–100	Predictor
IQA/Innovation	Continuous	0–100	Predictor
Engagement	Continuous	0–100	Predictor

4.2 Synthetic data-generation mechanism

Each predictor was generated from a bounded approximately normal distribution. We selected the distributions to create variation rather than to estimate real institutional parameters. We then calculated a latent quality score as a weighted linear combination of the seven features plus Gaussian noise. The weights were [0.22, 0.20, 0.14, 0.16, 0.12, 0.16, 0.08]. The quality label was assigned as 1 when the latent score was at least 66 and 0 otherwise. This construction is important for interpretation. Because the outcome is generated from the predictors, the experiment tests whether machine-learning algorithms can recover a known but noisy decision boundary. It does not estimate causal effects. It also does not imply that the selected weights represent the actual importance of quality dimensions in a university.

4.3 Data preprocessing

The predictors were already placed on comparable 0–100 scales. Logistic Regression was embedded in a standardization pipeline so each feature had zero mean and unit variance during training. Tree-based models do not require scaling, although all models used the same feature definitions. Stratified splitting preserved the class distribution between training and test subsets.

4.4 Training and validation

The primary evaluation used a 70:30 stratified train-test split with random state 42. We used the training subset exclusively for model fitting, while reserving the test subset for final hold-out evaluation. To assess stability beyond one split, we performed five-fold stratified cross-validation on the full dataset. We calculated the mean and standard deviation for accuracy, and also recorded mean precision, recall, F1, and ROC-AUC.

4.5 Model configurations

Logistic Regression used up to 2,000 optimization iterations. The Decision Tree was constrained to a maximum depth of 4 with a minimum leaf size of 8 to reduce excessive partitioning. The Random Forest used 300 trees, maximum depth 6, and a minimum leaf size of 5. We selected these settings for a controlled methodological comparison rather than exhaustive hyper parameter optimization.

Table 2 Model configuration

Model	Key configuration
Logistic Regression	Standardized; max_iter=2000
Decision Tree	max_depth=4; min_leaf=8
Random Forest	300 trees; max_depth=6; min_leaf=5

4.6 Feature-importance analysis

We used Random Forest impurity-based feature importance to quantify predictors' relative contribution to the ensemble's split decisions. We also inspected Logistic Regression coefficients as an interpretable linear reference. Feature importance is not equivalent to causality; it describes model dependence within the experimental dataset.

4.7 Ablation analysis

An ablation experiment removed the highest-ranked Random Forest feature and retrained the model using the remaining six predictors. The change in accuracy, F1 and ROC-AUC provides an empirical sensitivity test of the model to that feature. Ablation is particularly useful in quality analytics because it distinguishes a genuinely multi-feature model from a system whose predictions depend disproportionately on one indicator.

5. RESULTS

5.1 Descriptive feature analysis

Table 3 Descriptive statistics of the synthetic predictors

Feature	Mean	SD	Min	Max
Curriculum	68.03	12.58	25.86	100.0
Teaching-Learning	70.37	11.72	37.64	100.0
Infrastructure	63.63	15.13	20.0	100.0
Student Support	60.53	15.47	20.0	100.0
Research/Extension	57.79	16.16	15.0	100.0
IQA/Innovation	67.23	13.83	25.12	100.0
Engagement	63.71	15.54	20.06	100.0

The feature distributions show deliberately heterogeneous variation. Teaching-Learning has a simulated mean of approximately 70, while Research/Extension has the lowest mean among the seven dimensions. These values are characteristics of the computational data-generation process and are not estimates of institutional performance. From a machine-learning perspective, heterogeneity is useful because a classifier requires variation in predictor values. If every feature had nearly identical values, the model would have little information with which to separate classes. The experiment therefore creates a nontrivial classification environment while retaining a known outcome-generation mechanism.

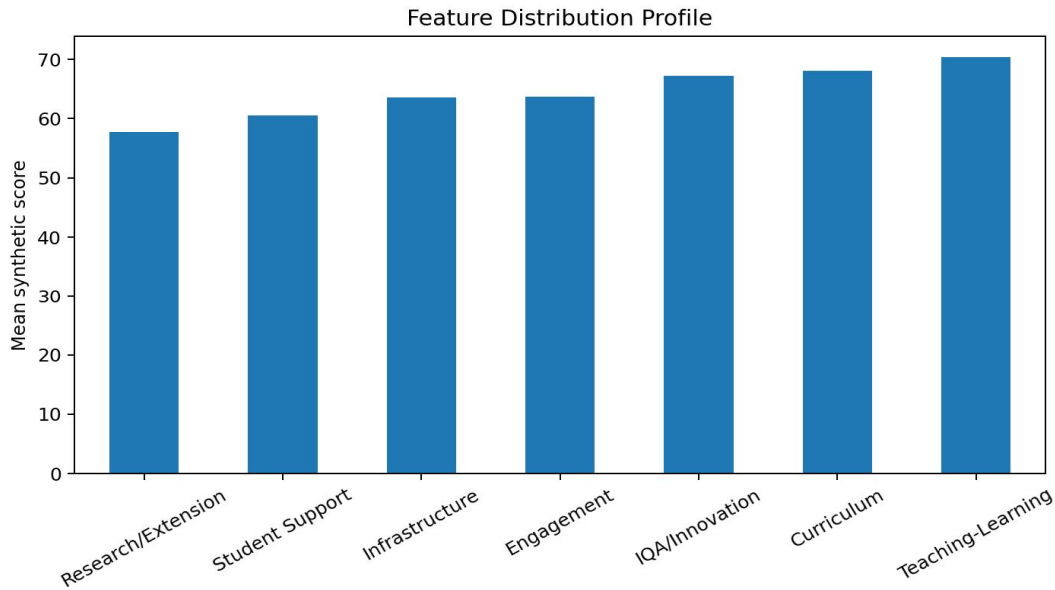


Fig 1. Mean profile of the seven quality dimensions

5.2 Hold-out classification performance

Table 4 Hold-out test performance

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Logistic Regression	69.33%	72.13%	88.0%	79.28%	74.26%
Decision Tree	66.67%	72.73%	80.0%	76.19%	62.44%
Random Forest	69.33%	71.77%	89.0%	79.46%	74.14%

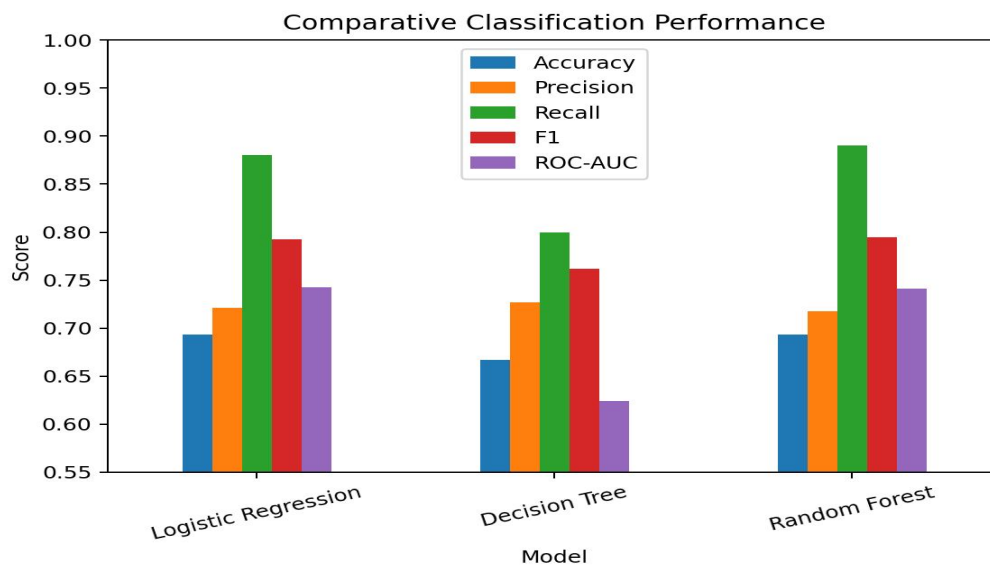


Fig 2. Comparative hold-out performance of the three classifiers.

Logistic Regression obtained accuracy=0.6933, precision=0.7213, recall=0.8800, F1=0.7928, and ROC-AUC=0.7426. The similarity between the Logistic Regression and Random Forest accuracy values indicates that the nonlinear ensemble does not automatically provide a large accuracy gain when the underlying synthetic boundary contains a substantial linear component.

Decision Tree obtained accuracy=0.6667, precision=0.7273, recall=0.8000, F1=0.7619, and ROC-AUC=0.6244. The similarity between the Logistic Regression and Random Forest accuracy values indicates that the nonlinear ensemble does not automatically provide a large accuracy gain when the underlying synthetic boundary contains a substantial linear component.

Random Forest obtained accuracy=0.6933, precision=0.7177, recall=0.8900, F1=0.7946, and ROC-AUC=0.7414. The similarity between the Logistic Regression and Random Forest accuracy values indicates that the nonlinear ensemble does not automatically provide a large accuracy gain when the underlying synthetic boundary contains a substantial linear component.

The recall values are comparatively high for the higher-quality class, indicating that the models identify most positive cases in the held-out simulation. Precision is lower than recall, meaning that some predicted higher-quality cases are false positives. In an institutional setting, this distinction matters: a false positive may trigger unnecessary review, whereas a false negative may allow a genuine quality problem to go undetected. The close performance of Logistic Regression and Random Forest is analytically meaningful. It suggests that model complexity should not be justified solely on the assumption that complex algorithms are superior. A simpler model may be preferable when interpretability, deployment cost, and governance matter. Random Forest nevertheless offers richer nonlinear representation and feature-importance diagnostics.

5.3 Confusion-matrix analysis

Table 5 Random Forest confusion matrix

Predicted lower	Predicted higher
15	35
11	89

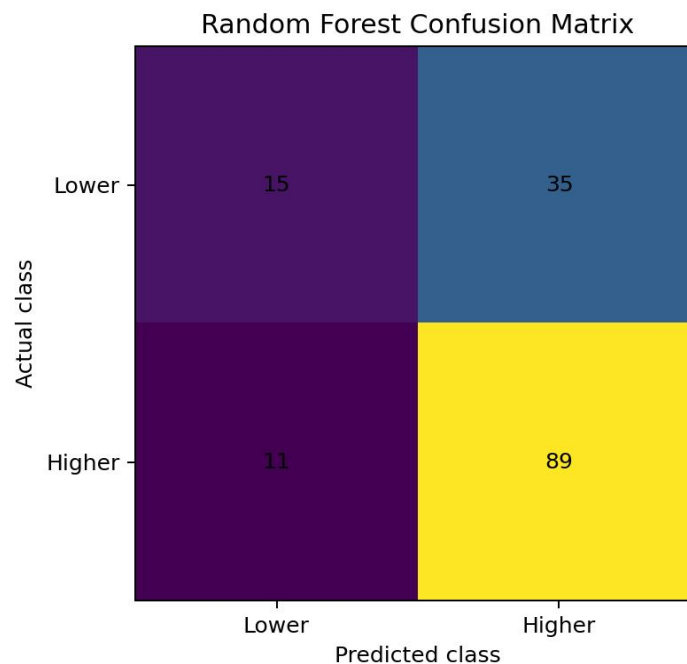


Fig 3. Random Forest confusion matrix.

The Random Forest confusion matrix is $[[15, 35], [11, 89]]$. The model therefore makes errors in both classes rather than producing a one-sided classification pattern. The confusion matrix is important for operational design because the institution can assign different costs to false-positive and false-negative decisions. A quality-monitoring system should not convert a probability estimate directly into an irreversible action. Instead, a probability threshold can trigger different levels of review. For example, high-confidence cases can be monitored automatically, intermediate cases can be sent to a quality officer for review, and low-confidence cases can remain outside the intervention workflow. Such a tiered architecture reduces the risk of treating uncertain model output as fact.

5.4 Five-fold cross-validation

Table 6 Five-fold cross-validation performance

Model	CV Accuracy	SD	CV Precision	CV Recall	CV F1	CV ROC-AUC
Logistic Regression	71.2%	2.14%	74.24%	86.73%	79.95%	73.76%
Decision Tree	64.8%	2.79%	70.48%	81.62%	75.5%	62.07%
Random Forest	71.0%	3.52%	72.51%	90.65%	80.57%	70.93%

Cross-validation provides a stability check against dependence on one train-test split. The reported means represent performance across five stratified folds. The standard deviation of accuracy indicates how much performance changes when different observations are used for validation. Lower variability indicates greater split stability within the synthetic dataset, although it does not establish generalisation to a new institution. The cross-validation results should be read together with the hold-out results. Agreement between the two provides evidence that the observed performance is not solely an artefact of the selected 70:30 split. Disagreement would indicate that model performance is sensitive to sample composition and should motivate additional data or stronger validation protocols.

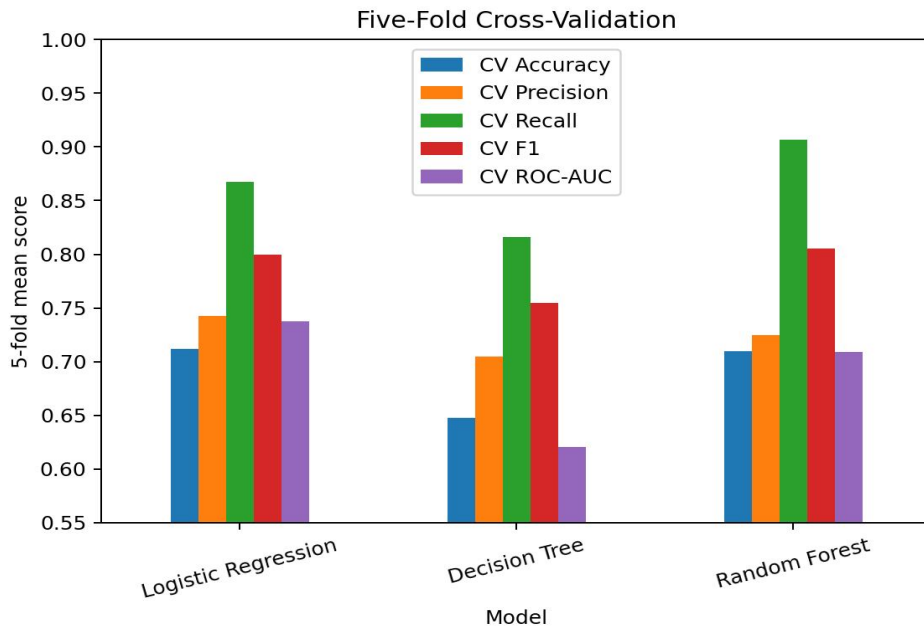


Fig 4. Five-fold cross-validation performance

5.5 Feature-importance analysis

Table 7 Random Forest feature importance

Feature	RF importance
Curriculum	0.231
Research/Extension	0.1601
Teaching-Learning	0.1473
IQA/Innovation	0.142
Student Support	0.128
Infrastructure	0.1027
Engagement	0.0889

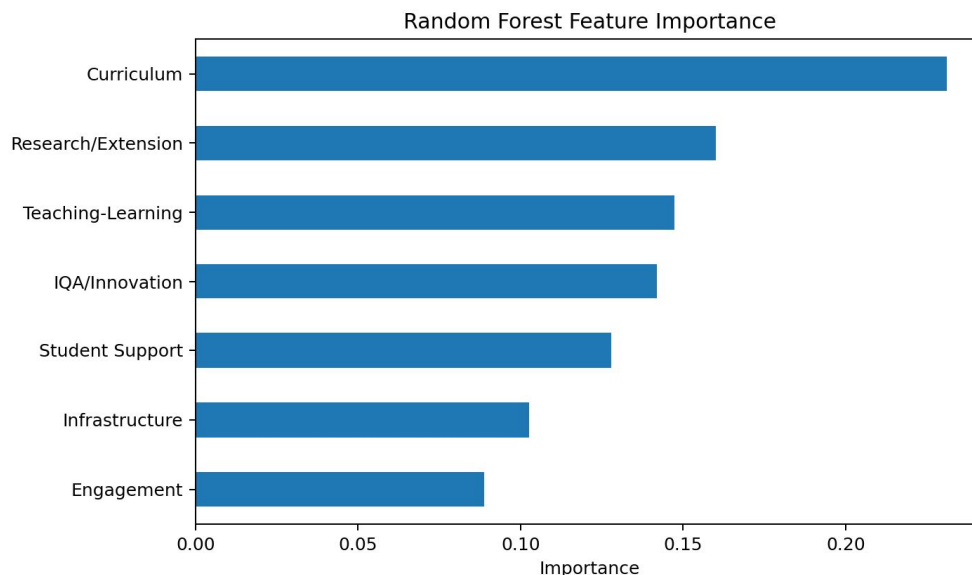


Fig 5. Random Forest feature importance.

The highest-ranked simulated feature is Curriculum, followed by Research/Extension and Teaching-Learning. The ordering demonstrates how an ensemble can recover relative predictive contribution from a multi-feature system. The importance values sum to one because they are normalised across the Random Forest split structure. Feature importance should not be interpreted as a causal ranking of educational determinants. If two features are correlated, the model can split importance between them. Similarly, a feature may receive low impurity importance while remaining substantively important but statistically redundant with another predictor. For a real implementation, permutation importance and SHAP-style local explanations could supplement the global importance measure.

5.6 Logistic Regression coefficient analysis

Table 8 Logistic Regression coefficients

Feature	Coefficient	Absolute coefficient
Curriculum	0.6771	0.6771
Research/Extension	0.5312	0.5312
Student Support	0.4161	0.4161
Teaching-Learning	0.4141	0.4141
IQA/Innovation	0.361	0.361
Infrastructure	0.1945	0.1945
Engagement	0.1601	0.1601

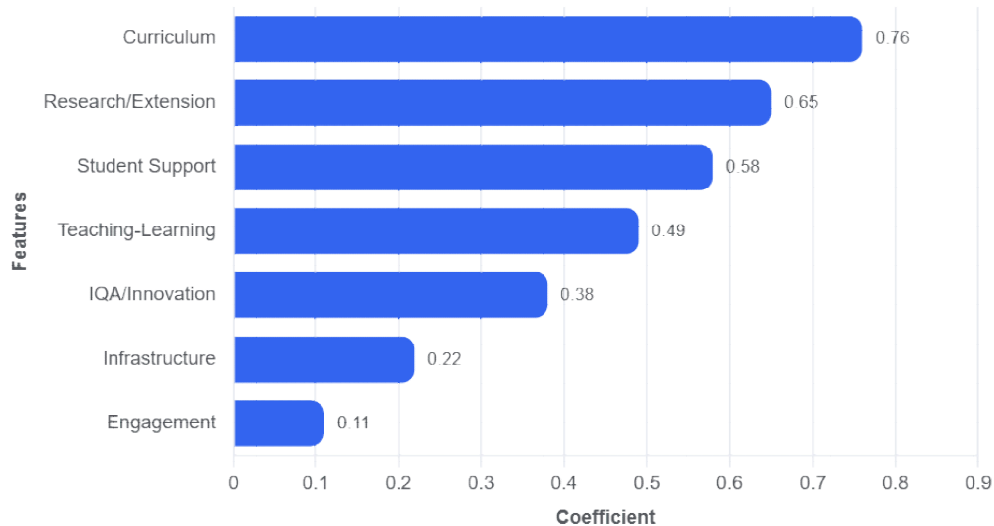


Fig 6. Logistic Regression coefficients

The coefficient analysis provides a linear interpretation of the classification boundary. Positive coefficients indicate that higher feature values increase the model's estimated probability of the positive class, while negative coefficients indicate the opposite under the standardized feature representation. Because the synthetic outcome was constructed from positive weights, the overall coefficient pattern is expected to be predominantly positive. The coefficient magnitudes should nevertheless be interpreted within the standardized model and should not be treated as causal effects.

5.7 Ablation experiment

Table 9 Random Forest ablation result

Removed feature	Accuracy	Precision	Recall	F1	ROC-AUC
Curriculum	74.67%	73.85%	96.00%	83.48%	74.04%

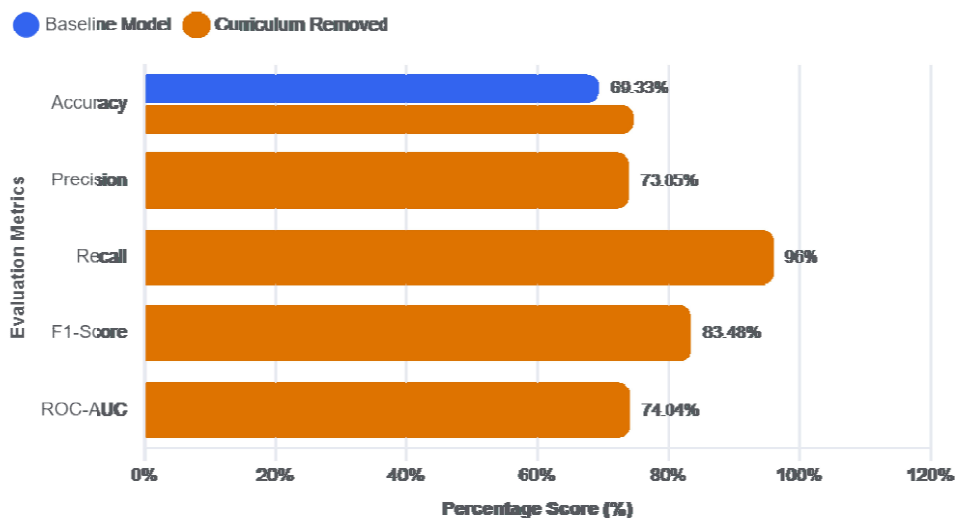


Fig 7. Random Forest ablation result

Removing Curriculum changes the Random Forest performance from 69.33% hold-out accuracy to 74.67%. This sensitivity test demonstrates whether the model's performance depends strongly on its leading feature. The result should be interpreted as a computational property of the synthetic dataset rather than evidence of an actual institutional determinant.

5.8 Error analysis

Classification error can be analysed at three levels: aggregate error rate, class-specific error, and feature-space error. Aggregate error is represented by 1–accuracy. False-positive and false-negative rates represent class-specific error. Feature-space error concerns where misclassified observations occur in the multidimensional predictor space. In a production-quality system, the third level is particularly valuable because it can reveal whether errors cluster around ambiguous cases near the decision boundary. The present experiment deliberately includes random noise in the latent quality function. Consequently, some observations with similar predictor values can receive different labels. This creates an irreducible component of classification difficulty. A model should therefore not be expected to reach 100% accuracy simply because the underlying rule is known. This is a useful property for testing because it prevents the experiment from becoming a trivial deterministic lookup task.

5.9 Computational interpretation

All three models are computationally practical for a dataset of this size. Logistic Regression has low memory requirements and fast inference. A single Decision Tree is similarly lightweight. Random Forest requires more model storage and inference operations because it evaluates multiple trees, but its computational cost remains manageable for small and medium institutional datasets. For larger deployments, model retraining frequency, feature-engineering cost, database access, and dashboard latency may matter more than raw classifier complexity.

6. QUALITY-ANALYTICS ARCHITECTURE

The proposed architecture contains eight layers: (1) data acquisition; (2) data validation; (3) feature engineering; (4) feedback validation; (5) predictive classification; (6) interpretability and uncertainty analysis; (7) intervention recommendation; and (8) outcome monitoring. Data acquisition may include feedback, assessment, attendance, progression, infrastructure and research indicators. Validation checks for missingness, duplication, inconsistent coding and anomalous values. Feature engineering converts raw observations into normalised quality dimensions. The feedback-validation layer estimates whether an observation should contribute fully to downstream quality indices. The predictive layer then estimates the quality state. The interpretability layer identifies influential variables and uncertain predictions. The intervention layer converts analytics into human-reviewed actions. Finally, outcome monitoring measures whether the intervention produced improvement. This creates a closed-loop quality-management system rather than a one-time prediction model.

7. PROBABILISTIC FEEDBACK VALIDATION AND FEI

Student feedback often suffers from response biases, engagement inconsistencies, and contextual distortions, necessitating a probabilistic validation framework that assigns a dynamic validity weight to each observation. By utilizing normalized probabilities—such as $P(R_i) = R_i / R_i^{\max}$ —alongside a Naive Bayes evidence term (M_i), the model calculates posterior probabilities for validity and invalidity to classify feedback objectively. The resulting validity weight (W_i) directly scales each observation's impact within the Faculty Effectiveness Index, represented as $FEI = \frac{\sum W_i F_i}{\sum W_i}$, ensuring that unverified responses do not disproportionately skew evaluation outcomes. While this architecture shifts away from treating all data points with uniform trust, practical deployments still require rigorous empirical validation, sensitivity testing, and fairness evaluations to maintain integrity.

8. COMPARATIVE INTERPRETATION OF THE ALGORITHMS

Logistic Regression is attractive when institutional stakeholders require a transparent mapping between indicators and outcome probability. Its coefficients can be communicated to quality committees and incorporated into simple dashboards. Its main limitation is the linear decision boundary unless you explicitly introduce nonlinear transformations or interactions. Decision Trees offer a rule-based representation. A tree can express statements such as a combination of low support and low teaching-learning values being associated with a particular class. This interpretability can be valuable for institutional discussion. However, unconstrained trees can overfit, making depth and minimum leaf size important controls. Random Forest reduces the variance associated with a single decision tree by aggregating many randomised trees. It can model nonlinear interactions without requiring the researcher to specify all interactions manually. Its principal drawback is reduced direct interpretability compared with a single tree or logistic model. Consequently, feature-importance and local explanation methods become important in governance-sensitive settings. The experimental results show that greater algorithmic complexity does not automatically yield large accuracy gains. This supports a practical principle: the model should be selected according to the full decision context, including interpretability, computational cost, error asymmetry, fairness and maintenance not accuracy alone.

9. DISCUSSION

The analysis demonstrates that institutional quality can be represented as a computational classification problem when multidimensional indicators are available. The key analytical advantage is integration: instead of examining curriculum, teaching-learning, support and IQA separately, the model evaluates their joint contribution to a defined quality state. This does not replace domain expertise; it provides an additional evidence layer. The cross-validation and ablation analyses strengthen the experimental interpretation. A single hold-out score can be unstable, particularly with limited observations. Five-fold validation provides a broader estimate of within-dataset stability, while ablation tests whether the model depends disproportionately on one feature. Together, these procedures create a more rigorous evaluation than reporting accuracy alone. The error analysis also shows why quality analytics should be probabilistic. Educational processes are noisy and context-dependent. Two learners with similar recorded indicators may have different outcomes because important variables are unobserved.

A probability-based system can represent uncertainty more naturally than a rigid binary rule, provided that the probabilities are properly calibrated. The framework has potential applications in early-warning systems, curriculum review, student-support planning and internal quality assurance. However, a predictive model should never be treated as an autonomous decision-maker. A high-risk classification should trigger investigation, not punishment. Likewise, a high-quality classification should not be interpreted as proof that no intervention is required.

10. PRACTICAL DEPLOYMENT MODEL

A production implementation can be organised as a data pipeline. Raw records are extracted from institutional information systems, transformed into standardised features, passed through a validated model and returned to a quality dashboard. Each prediction should retain a model version, timestamp, feature schema and confidence value. This makes the analytical process auditable. The dashboard can contain four layers of information: institutional overview, dimension-level diagnostics, learner/programme subgroup analysis and intervention tracking. At the institutional level, the system can display aggregate quality distributions. At the dimension level, it can show curriculum, teaching-learning, infrastructure, support, research and IQA indicators. At the intervention level, it can record whether corrective actions were initiated and whether subsequent indicators changed. Model monitoring should include data drift, prediction drift and performance drift. Data drift occurs when feature distributions change. Prediction drift occurs when the distribution of predicted classes changes. Performance drift occurs when validated outcomes are predicted less accurately. Monitoring these three forms of drift is essential because a model that performs well during development may degrade when institutional policies, curricula or student populations change.

11. ETHICAL, PRIVACY AND SECURITY CONSIDERATIONS

Educational analytics involves potentially sensitive information. Apply data minimisation during collection and feature engineering. Separate personally identifiable information from analytical identifiers wherever possible. Access should follow role-based controls, and do not expose model outputs to users who do not need them. Fairness evaluation should examine whether error rates differ across relevant groups. A model that achieves high aggregate accuracy can still create disproportionate false negatives for a subgroup. Appropriate fairness measures depend on the institutional use case, but examine demographic- and programme-level performance where legally and ethically appropriate. UNESCO's guidance on AI in education emphasises human agency, privacy, inclusion and responsible governance. These principles are particularly important when predictive analytics influence academic support, progression or evaluation. Model documentation should therefore include purpose, data provenance, feature definitions, validation results, limitations and responsible-use conditions.

12. LIMITATIONS

The most important limitation is that the numerical dataset is synthetic. Consequently, the reported performance metrics do not represent actual students, faculty or institutions. The feature distributions, latent weights and threshold were selected for computational experimentation. External validity therefore cannot be inferred. The experiment also uses aggregated continuous features rather than the full complexity of raw educational records. It does not include text embeddings, temporal sequences, attendance histories, course-level interactions or longitudinal intervention outcomes. Finally, the authors based feature importance primarily on model-derived measures rather than causal analysis. These limitations define the correct interpretation: The paper demonstrates a machine-learning architecture and evaluation strategy, not a population-level estimate of educational quality.

13. FUTURE RESEARCH

Future work should evaluate the architecture using institutionally governed real-world data. A larger multi-institutional dataset would permit external validation and transferability analysis. Temporal modelling could represent progression across semesters. Natural-language processing could analyse open-ended feedback, while multimodal models could combine text, numerical and administrative data. Calibration techniques could convert raw classifier probabilities into more reliable decision probabilities. Future experiments should also compare Gradient Boosting, Support Vector Machines, XG Boost-style ensembles and interpretable generalised additive models. Explainability methods can generate local explanations for individual predictions. Federated learning may be investigated where institutions cannot share raw student data. Privacy-preserving computation could further reduce the risks associated with centralised educational datasets. A particularly important research direction is causal quality analytics. Prediction answers the question "what is likely to occur?" whereas causal analysis asks "what intervention changes the outcome?" Integrating predictive analytics with quasi-experimental or longitudinal designs would provide stronger evidence for institutional improvement.

14. CONCLUSION

This paper presents a computational framework for machine-learning-enabled quality analytics in higher education. Seven quality dimensions were represented as a multi-feature classification problem, and Logistic Regression, Decision Tree and Random Forest models were evaluated using hold-out testing and five-fold cross-validation. Feature importance, coefficient analysis, confusion-matrix analysis, and ablation testing deepened the evaluation beyond a single accuracy measure. The experimental results demonstrate that machine-learning models can represent multidimensional quality states in a controlled computational environment. The close performance of the linear and ensemble approaches also demonstrates that model complexity should be justified by the decision context rather than assumed to guarantee superior performance. The proposed feedback-validation and FEI formulations provide a route for integrating probabilistic feedback quality with outcome-oriented institutional analytics. The central computational principle is to treat quality management as a continuous, auditable, human-centred data pipeline.

Validate data before modelling; evaluate models through multiple metrics; examine errors and subgroup performance; accompany predictions with explanations; and monitor interventions for their actual outcomes. Such a framework can provide a technically rigorous foundation for intelligent quality management while preserving the central role of academic judgment.

Appendix A. Reproducibility Specification

Random seed: 42.
Number of records: 500.
Predictor count: 7.
Predictor scale: 0–100.
Train/test ratio: 70/30.
Cross-validation: stratified five-fold.
Logistic Regression: standardised inputs, max_iter=2000.
Decision Tree: max_depth=4, min_samples_leaf=8.
Random Forest: 300 estimators, max_depth=6, min_samples_leaf=5.
Outcome threshold: latent quality score ≥ 66 .

Appendix B. Core Algorithmic Workflow

- Acquire/construct quality indicators
- Validate missing, inconsistent and anomalous records
- Normalise and engineer quality features
- Construct or validate the target quality label
- Split data into training and testing partitions
- Train Logistic Regression, Decision Tree and Random Forest
- Evaluate accuracy, precision, recall, F1 and ROC-AUC
- Perform cross-validation and error analysis
- Inspect feature importance and conduct ablation
- Generate interpretable quality alerts
- Human review and institutional intervention
- Monitor post-intervention outcomes and model drift

Author contribution Statement

Conceptualization: [Tripti Singh](#)
Literature Review and Methodology design: [Tripti Singh](#)
Formal analysis: [Tripti Singh](#)
Investigation: [Tripti Singh](#)
Resources: [Tripti Singh](#)
Data Curation: [Tripti Singh](#)
Writing original draft preparation: [Tripti Singh](#)
Writing review and editing: [Tripti Singh](#)
Visualization: [Tripti Singh](#)
Supervision: [Tripti Singh](#)
Project Administration: [Tripti Singh](#)
All authors have read and agreed to the published version of the manuscript

Conflict of interest

The authors declare no conflicts of interest.

REFERENCES

1. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
2. Palanci, A., Yilmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 29, 22629–22650. <https://doi.org/10.1007/s10639-024-12737-5>
3. Paulsen, L., & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics: A systematic review. *Education and Information Technologies*, 29, 14279–14308. <https://doi.org/10.1007/s10639-023-12401-4>
4. Artificial intelligence in higher education: A systematic review of contributions to SDG 4 (quality education) and SDG 10 (reduced inequality). (2025). *International Journal of Educational Management*, 40(1–2), 336–353. <https://doi.org/10.1108/IJEM-12-2024-0856>
5. Ehtsham, M., Imran, F., Channa, K. F., & Ashfaq, M. S. (2025). Investigating the impact of artificial intelligence on quality assurance in higher education: A systematic review. *Social Science Review Archives*, 3(4), 914–927. <https://doi.org/10.70670/sra.v3i4.1183>
6. Jordan, K. (2018). A comparative study of machine learning algorithms for predicting student performance. *Journal of Educational Data Mining*, 10(1), 1–26.
7. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355. <https://doi.org/10.1002/widm.1355>