



Interpretable Student Performance Mining from E-Learning Interaction Logs Using Feature Selection, Classification, and Clustering

C. Udayakumar¹, Dr. M. Chandran²

¹Research Scholar, Department of Computer Science, Sri Ramakrishna Mission Vidyalaya, College of Arts and Science, Coimbatore, Tamil Nadu.

²Associate Professor, Department of Computer Applications, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamil Nadu
E-mail id- onchandran@gmail.com.

Abstract

Educational assessment generally evaluates students using final marks, while technology-enhanced learning environments can record the learning actions that precede those marks. This research presents an Educational Data Mining framework for analyzing student performance from interaction logs generated by MathTutor, an e-learning application for mensuration. Five classifiers - Naive Bayes, Multilayer Perceptron, ZeroR, J48, and Random Forest - were compared. J48 and Multilayer Perceptron achieved 100% reported accuracy; J48 required only 0.03 seconds to build the model compared with 27.19 seconds for the Multilayer Perceptron and was selected as the preferred interpretable classifier. Expectation Maximization clustering produced five groups whose membership counts matched the five grade categories generated by J48. The findings further indicate that high-performing learners generally viewed more examples, practiced more exercises, and used fewer hints. The study demonstrates how interpretable data mining can supplement score-based assessment with evidence about learning behaviour and provide a basis for targeted remedial support.

Keywords: Educational Data Mining, Student Performance, J48, Expectation Maximization, Feature Selection, Information Gain, Chi-Square, E-Learning, Learner Profiling.

I. INTRODUCTION

Digital learning systems can capture far more information than traditional research-based assessment. In addition to final scores, a tutoring environment can record the number of examples viewed, exercises practiced, hints requested, mistakes made, time spent, and the sequence of problem-solving actions. Educational Data Mining applies machine-learning and data-mining methods to these records so that teachers can understand both student outcomes and the learning behaviours associated with those outcomes [1]-[3].

This research focuses on student-level performance mining. The goal is not merely to predict a grade, but to determine whether a compact set of interaction variables can produce an accurate and interpretable performance model. The study uses the MathTutor environment developed for teaching mensuration. Its learning sequence contains learning material, worked examples, exercises with assistance, and tests without hints. Student activity is automatically stored as a transaction log, making the dataset suitable for objective analysis rather than manual reporting.

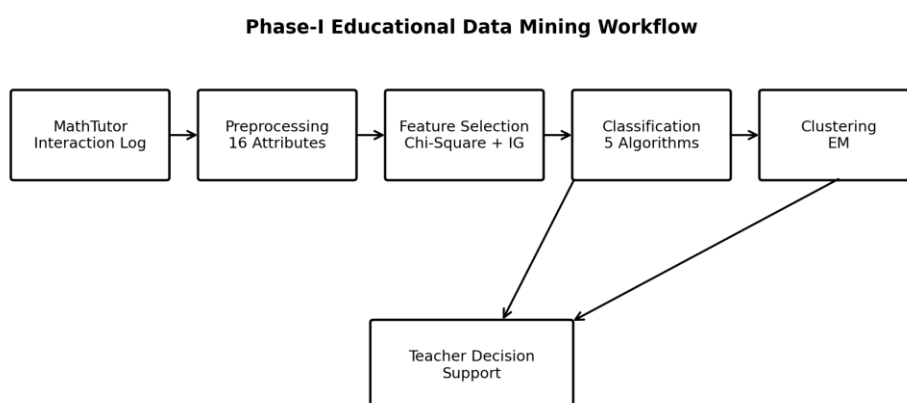


Figure 1. Proposed Phase-I workflow for performance mining

II. RELATED WORK

Educational performance prediction has been studied using decision trees, Bayesian classifiers, neural networks, clustering, association rules, and related machine-learning techniques. Earlier work often relied on semester marks, attendance, assignments, demographic variables, and institutional records. Such approaches are valuable for early risk prediction, but they do not directly describe how students interact with learning content. The

literature reviewed in the thesis also shows that clustering can be used to identify learner groups with similar characteristics and that interpretable classifiers are especially attractive when teachers must understand the basis of a prediction.

WEKA is suitable for this type of comparative experiment because it provides preprocessing, feature selection, supervised classification, unsupervised clustering, and visualization within a common environment [7]. The research design therefore combines attribute evaluation with multiple classification algorithms and an unsupervised cluster validation step.

III. RESEARCH METHODS

A. Dataset and Learning Environment

The MathTutor lesson considered in this study teaches metric measures within mensuration. The system contains 60 problems distributed across example, exercise, and test categories. The experimental analysis focuses on 120 sixth-grade students who completed 20 test problems: 10 mental problems, 5 simple problems, and 5 big problems. Each correctly solved test problem contributes one point, producing a maximum test score of 20. The interaction log contains approximately 7,200 entries together with student status, examples viewed, exercises solved, hints, problem-solving steps, test outcomes, and related activity information.

Table 1. Student performance grade definition used in the experiment

Score range	Grade	Interpretation
19-20	A	Excellent
16-18	B	Very Good
12-15	C	Good
8-11	D	Not Bad
0-7	E	Not Satisfactory

B. Preprocessing and Feature Construction

The source log was transformed into a student-level dataset of 120 instances and 16 attributes. These attributes represented Student ID, examples viewed, maximum examples, exercise problems solved, maximum exercises, hints requested, maximum hints, scores and maximum scores for mental, simple, and big problems, total score, maximum total score, and grade. Feature reduction was then performed so that the mining model used only variables that contributed useful information to the class label.

C. Feature Selection

Two independent attribute evaluators were used: Chi-Square and Information Gain. Both methods produced the same ordering for the seven highest-ranked attributes, increasing confidence that the reduced feature set was stable for the experiment. For a class variable G and candidate feature X, Information Gain is expressed as the reduction in entropy obtained after observing X.

$$IG(G, X) = H(G) - H(G | X)$$

Table 2. Ranked attributes selected for performance modeling

Rank	Attribute	Chi-Square	Information Gain
1	Score	465.946	2.0397
2	Big problems solved	292.821	1.3754
3	Hints requested	279.799	1.3453
4	Simple problems solved	165.729	0.7863
5	Mental problems solved	134.829	0.7842
6	Examples viewed	133.524	0.6280
7	Exercise problems solved	92.337	0.4741

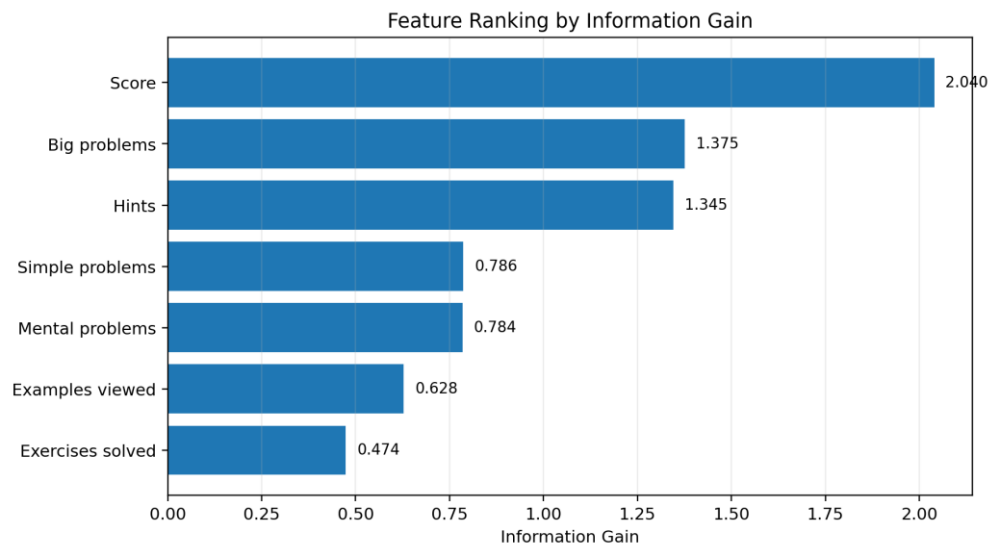


Figure 2. Information Gain values of the selected attributes

D. Classification Model

Five classification algorithms were evaluated: Naive Bayes, Multilayer Perceptron (MLP), ZeroR, J48, and Random Forest. The model-selection procedure compares correctly and incorrectly classified instances, accuracy, Kappa statistic, error measures, and model-building time. Accuracy is defined as the proportion of correctly classified instances among all test instances.

Accuracy = (Correctly Classified Instances / Total Instances) x 100

J48 is a decision-tree classifier derived from the C4.5 family. Its principal advantage in this educational setting is interpretability: a teacher can inspect the generated decision rules and understand which student actions contribute to a predicted grade. This is important because an accurate but opaque classifier provides less instructional value than a model that also explains the conditions associated with performance.

E. Clustering Model

After removing the known grade label, Expectation Maximization (EM) clustering was applied with five clusters, corresponding to the five grade categories. The unsupervised grouping was then compared with the supervised J48 class distribution. Agreement in the number of members across the five groups was used as supporting evidence that the performance structure discovered by the classifier was also present in the interaction features.

IV. RESULTS AND DISCUSSION

A. Classification Performance

Table 3. Reported performance of the evaluated classifiers

Classifier	Correct	Incorrect	Accuracy (%)	Kappa	MAE	RMSE	Build time (s)
Naive Bayes	118	2	98.333	0.9779	0.0112	0.0719	0
MLP	120	0	100.000	1.0000	0.0048	0.0076	27.19
ZeroR	35	85	29.000	0.0000	0.3021	0.3882	0.30
J48	120	0	100.000	1.0000	0.0000	0.0000	0.03
Random Forest	116	4	96.667	0.9557	0.0083	0.764	0.80

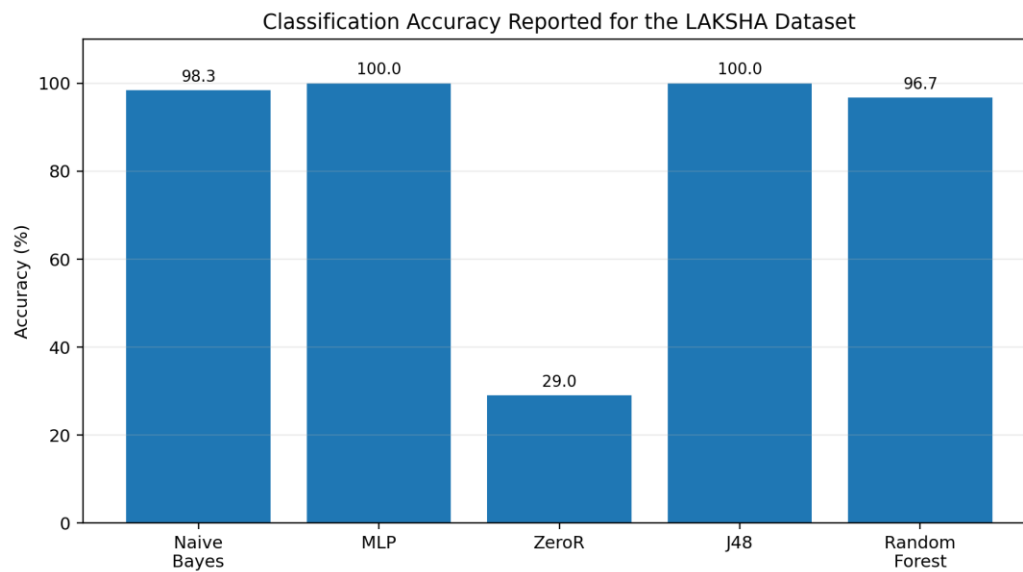


Figure 3. Comparative classification accuracy

The results show that both J48 and MLP classified all 120 instances correctly in the reported experiment. However, the J48 model required only 0.03 seconds to build, whereas the MLP required 27.19 seconds. J48 also produced zero reported mean absolute error and RMSE in this dataset. The combination of perfect reported accuracy, low computational cost, and explicit decision rules supports its selection as the preferred classifier for this study. Naive Bayes and Random Forest also performed strongly, whereas ZeroR served as a weak baseline.

B. Class Distribution and Clustering Agreement

Table 4. EM clusters compared with the J48 grade distribution

EM Cluster	Grade	Number of students	Percentage
0	E	9	8%
1	D	11	9%
2	A	30	25%
3	B	35	29%
4	C	35	29%

The five EM clusters contained exactly the same numbers of instances as the five J48 grade groups in the reported analysis. This does not by itself establish equivalence of every individual assignment, but it provides useful evidence that the selected interaction attributes encode a natural performance structure consistent with the supervised grade labels. In practical use, clustering can therefore support learner grouping even when explicit grade labels are unavailable.

C. Learning Behaviour Associated with Performance

The J48 tree and EM visualizations provide an important educational interpretation beyond the final score. High-performing learners generally viewed more worked examples, completed more exercise problems, solved more complex problems successfully, and requested fewer hints. The thesis reports that only one hint was used by two top A-grade learners, while most B-grade learners viewed more than 17 of the 20 available example problems. These observations suggest that practice behaviour and help dependence contain meaningful information about achievement.

A useful interpretation of the Phase-I model is therefore: student performance is jointly related to outcome evidence and learning-process evidence. Score identifies the level reached; examples viewed and exercises solved describe practice exposure; hints requested indicate assistance dependence; and performance on simple and big problems reflects the ability to apply knowledge under increasing complexity. This allows the teacher to distinguish low performance caused by insufficient practice from low performance associated with persistent difficulty even after assistance.

V. PROPOSED PERFORMANCE-MINING MODEL

The Phase-I model can be expressed as a sequence of transformations rather than a single weighted score. Let R denote the raw student-tutor log. Preprocessing produces a student-level feature vector X . Feature selection retains the seven high-ranked variables X^* . J48 then estimates the supervised performance class G , while EM independently estimates an unsupervised learner cluster C .

$R \rightarrow X \rightarrow X^* \rightarrow \{ J48(X^*) = G, EM(X^*) = C \} \rightarrow \text{Teacher Decision}$

This architecture avoids arbitrary weights and retains interpretability. If the predicted grade is low and the learner also appears in a cluster characterized by fewer examples and exercises or greater hint usage, the system can recommend additional practice or teacher intervention. If performance is high with low assistance, the learner can progress to the next lesson. Thus, the model converts mining results into an instructional decision without reducing learning behaviour to a single opaque score.

CONCLUSION

This research presented an interpretable Educational Data Mining approach for student performance analysis using interaction logs from a mathematics e-learning environment. The original 16 attributes were reduced to seven influential variables using Chi-Square and Information Gain. Among five evaluated classifiers, J48 and MLP achieved 100% reported accuracy, while J48 was substantially faster and easier to interpret. EM clustering generated five groups with the same membership counts as the five grade categories. Most importantly, the analysis showed that learning actions such as viewing examples, practicing exercises, solving difficult problems, and requesting hints provide useful evidence beyond the final score. The Phase-I model therefore establishes a practical foundation for learner profiling and remedial decision support. A subsequent stage can extend this student-level analysis to fine-grained knowledge-component modeling so that the system identifies not only which students need support, but also which mathematical skills require additional learning opportunities.

REFERENCES

- [1] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, vol. 33, no. 1, pp. 135-146, 2007.
- [2] R. S. J. d. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3-17, 2009.
- [3] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 40, no. 6, pp. 601-618, 2010.
- [4] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User Modeling and User-Adapted Interaction*, vol. 4, no. 4, pp. 253-278, 1995.
- [5] H. Cen, K. R. Koedinger, and B. Junker, "Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement," in *Proc. 8th International Conference on Intelligent Tutoring Systems*, pp. 164-175, 2006.
- [6] K. R. Koedinger, R. S. J. d. Baker, K. Cunningham, A. Skogsholm, B. Leber, and J. Stamper, "A Data Repository for the EDM Community: The PSLC DataShop," in *Handbook of Educational Data Mining*, CRC Press, 2010.
- [7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10-18, 2009.
- [8] J. C. Stamper and K. R. Koedinger, "Human-Machine Student Model Discovery and Improvement Using DataShop," in *Proc. AIED 2011*, pp. 353-360, 2011.
- [9] H. Cen, K. R. Koedinger, and B. Junker, "Is Over Practice Necessary? Improving Learning Efficiency with the Cognitive Tutor through Educational Data Mining," in *Proc. AIED 2007*, pp. 511-518, 2007.
- [10] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "Implementation of E-Learning Package for Mensuration - A Branch of Mathematics," *IEEE*, ISBN 978-1-4799-2876-7, pp. 219-221, 2014.
- [11] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "Educational Data Mining Applications," *Operations Research and Applications: An International Journal*, vol. 1, no. 1, pp. 23-29, 2014.
- [12] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "Performance of Classification Algorithms on Students' Data - A Comparative Study," *International Journal of Computer Science and Mobile Applications*, vol. 3, no. 9, pp. 1-8, 2015.
- [13] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "Analysing Students Performance Using Educational Data Mining Methods," *International Journal of Applied Engineering Research*, vol. 10, no. 82, pp. 667-671, 2015.
- [14] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "A Study on Learning Factor Analysis - An Educational Data Mining Technique for Student Knowledge Modeling," *IOSR Journal of Computer Engineering*, vol. 17, no. 6, pp. 95-101, 2015.
- [15] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "Assessing Students' Performance Using Learning Curves," *International Journal of Engineering and Techniques*, vol. 2, no. 1, pp. 6-16, 2016.
- [16] S. Lakshmi Prabha and A. R. Mohamed Shanavas, "ELATE - A New Student Learning Model Utilizing EDM for Strengthening Math Education," *International Journal of Computer Sciences and Engineering*, vol. 4, no. 3, pp. 10-14, 2016.