EVALUATING THE EFFECTIVENESS OF PROCEDURE INFORMATION USING DATA MINING ALGORITHMS

AUTHOR : Dr. SANTOSH KUMAR BYRABOINA.

M.Sc., M.Tech., Ph.D.

ASSOCIATE PROFESSOR,

WESLEY PG COLLEGE

SECUNDERABAD, TELANGANA, INDIA.

Absract

To illustrate the procedures, we use a sample of U.S. students' answers to problem-solving questions on the 2022 PISA (N = 426). Classifier development operations are carried out utilizing the shown methods after the creation and selection of concrete features. All of the methods yielded good results in terms of categorization accuracy. Recommendations for choosing classifiers are provided, taking into account research topics, classifier interpretability, and classifier simplicity. Both supervised and unsupervised learning approaches provide outcomes that are explained.With the advent of big data, it has become more difficult to glean insights from large datasets. This study aims to provide light on the possible benefits, drawbacks, and ramifications of using advanced analytical techniques to extract significant patterns, correlations, and trends from large datasets by examining the overlap between data mining and big data.The fundamental goals of this research are to determine the most effective data mining methods for large-scale data analysis, evaluate their scalability and computing efficiency, and get a knowledge of their potential to uncover previously unknown insights. In order to better understand the interplay between data mining approaches and the complexities of big data, this study will analyze the relevant literature and perform experiments. With the advent of the age of big data, there has been an unprecedented flood of data in many fields, requiring novel methods to be developed in order to glean useful insights. Knowledge discovery relies heavily on data mining, which has developed to meet the problems provided by big data's volume, velocity, and diversity.

Keywords: Data Mining, Big Data, Scalability, Patterns, Insights, Computational Efficiency.

Introduction

An extraordinary increase in data creation characterizes the current information environment, posing new difficulties and opening up new possibilities in many different areas of study. The need to glean useful insights from a growing mountain of data is only going to increase in importance. Two potent fields, data mining and big data, have come together thanks to this effort. Data mining, the process of discovering patterns, correlations, and trends within enormous datasets, provides an attractive approach to extracting useful insights from the sea of data known as big data.

Big Data's Explosive Growth:

"Big data" describes the huge datasets generated by a wide variety of modern digital activities such as social media, sensor networks, online transactions, and more. The sheer volume, pace, and variety of big data are only a few of the factors that contribute to its intrinsic complexity. Conventional data processing techniques can't handle the intricacies of big data since they were built for smaller datasets. Not only is it difficult to keep track of all of this information, but it's also difficult to draw conclusions from it that will lead to good choices.Data mining provides a methodical strategy for handling the complexities of large data analysis by drawing on methods from statistics, machine learning, and database systems. Its major goal is to unearth hidden patterns, correlations, and outliers in massive data sets. Data mining is a technique that uses sophisticated algorithms and computing power to uncover useful information that may be used to better decision-making, corporate strategies, and even the identification of new prospects. Scholars from both the United States and overseas have contributed several resources on data mining used a mix of data mining and dynamic behavior interception to unearth previously inaccessible data and check for the presence of a virus. He used it to locate Trojan viruses in computer networks. Machine learning (ML) and data mining (DM) techniques are introduced briefly in a tutorial format by Buczak and Guven . Xu et al. take a more holistic view of privacy concerns in the context of data mining, investigating a range of techniques for keeping personal data secure. He discussed cutting-edge techniques and proposed some early concepts for further study. After accounting for data mining, Yan and Zheng discovered that several basic signals are significant predictors of stock returns across industries. Addition to boosting data mining efficiency, it is able to effectively extract useful insights from the underlying data. Optimizing the research using the least mean square method, this study enhances the experimental effect.These studies cannot be trusted because they lack credibility due to

insufficient data and unfounded conclusions.Adding to boosting data mining efficiency, it is able to effectively extract useful insights from the underlying data. Optimizing the research using the least mean square method, this study enhances the experimental effect.

Review of Literature: Applying Data Mining Techniques over Big Data

The intersection of data mining techniques and the challenges posed by big data has garnered significant attention from researchers and practitioners alike. This section presents an overview of key literature that explores the application of data mining methods within the context of large-scale datasets.

1. Data Mining Techniques for Big Data:

Research by Han, et al. (2011) emphasizes the importance of data mining techniques in handling the complexities of big data. To guarantee effective pattern and insight extraction from massive datasets, the authors provide scalable approaches for tasks including classification, clustering, and association rule mining. This seminal book emphasizes the value of tailoring data mining techniques to the peculiarities of huge data.

2. Scalability and Performance:

Chen and Zhang (2014) delve into the challenges of scalability and performance when applying data mining techniques to big data scenarios. They emphasize the need for distributed and parallel processing techniques to overcome computational bottlenecks. The study discusses the trade-offs between accuracy and efficiency in the context of processing massive datasets and provides insights into optimizing algorithm performance.

3. Unstructured Data Mining:

The authors explore sentiment analysis, text mining, and image recognition, demonstrating how data mining can transform unstructured data into actionable insights. This research underscores the versatility of data mining methods in addressing the diverse formats and types of data within big data repositories.

4. Stream Mining and Real-Time Analytics:

In this age of big data, Gama et al. (2014) stress the significance of real-time analytics and stream mining. The study discusses the challenges of handling high-velocity data streams and

proposes adaptive algorithms that can continuously update models and patterns. 5. Case Studies and Industry Applications:

Numerous case studies demonstrate the tangible impact of applying data mining techniques to big data across various industries. For instance, in healthcare, data mining has been used to predict disease outbreaks (Savova, et al., 2010). In finance, data mining aids in fraud detection (Phua, et al., 2010), while in marketing, it enables personalized recommendations (Linden, et al., 2003). These case studies showcase the versatility and efficacy of data mining techniques in addressing industry-specific challenges within the realm of big data.

Objectives of the Study:

Data mining methods and the difficulties posed by huge data are the focus of this research. This study intends to shed light on the efficacy of data mining methods by analyzing their implementation in the context of large-scale datasets. Through empirical evaluation and case studies, the study will elucidate the benefits, limitations, and implications of applying data mining methodologies to big data scenarios.

Research and Methodology

Selected 429 American pupils at random from the PISA 2022 dataset. The pupils' average age was 15 years and 3 months, which is representative of American teenagers of that age.. After eliminating three pupils whose records were lacking both student and school IDs, the final count was 426. There were no blanks in the data. A total of 320 samples (75.12%) were used to create a training set, whereas 106 (24.88%) were used to create a test set. To improve prediction accuracy (see also Sinharay, 2016; Fossey, 2017), training datasets are often two-to-three times the size of test datasets.In PISA 2022, students are asked to solve 42 problems across 16 different topics. Organization for Economic Co-operation and Development states that these questions measure mental processes in simulated computer environments where real-world issues must be solved. In this research, TICKETS task2 (CP038Q01), a problem-solving item, was examined. This question is a level 5 (out of a possible 6) and demands advanced knowledge and analytical skills to solve Planning and carrying out are the major mental operations required for this activity. Students are expected to formulate a solution to the issue presented, put it to the test, and make any necessary adjustments based on the results. When students need to make a modification, they may click the "CANCEL" button that appears before the "BUY" button. For this assignment to be completed successfully, students will need to weigh the pros and drawbacks of these two options, compare their

respective prices, and settle on the most economically viable option.This item receives one of three possible polytomous scores: 0, 1, or 2. Partial credit is awarded to students who derive just one answer without comparing it to the other. Students who fail to come up with any of the two answers and instead purchase the incorrect ticket will not get any points for this assignment. Figure 1's last image shows the price of 72 zeds for four full-fare tickets on a rural train. Actions like "COUNTRY TRAINS" and "FULL FARE" are not directly connected to this item since they are not required to complete the job at hand. Unrelated activities are acceptable for scoring purposes so long as students compare options and ultimately purchase the proper ticket. Student identification was aided by the ID variable, and features were generated using the event_value and time variables. Unfortunately, the log file did not include each student's score; these had to be manually coded and validated against the grading rubric to ensure accuracy. Of the 426 pupils, 121 (28.4%) received a passing grade, 224 (52.6%) received a failing grade, and 81 (19.0%) received no grade at all. Codes of 2, 1, and 0 were assigned to indicate full, partial, and no credit, respectively.

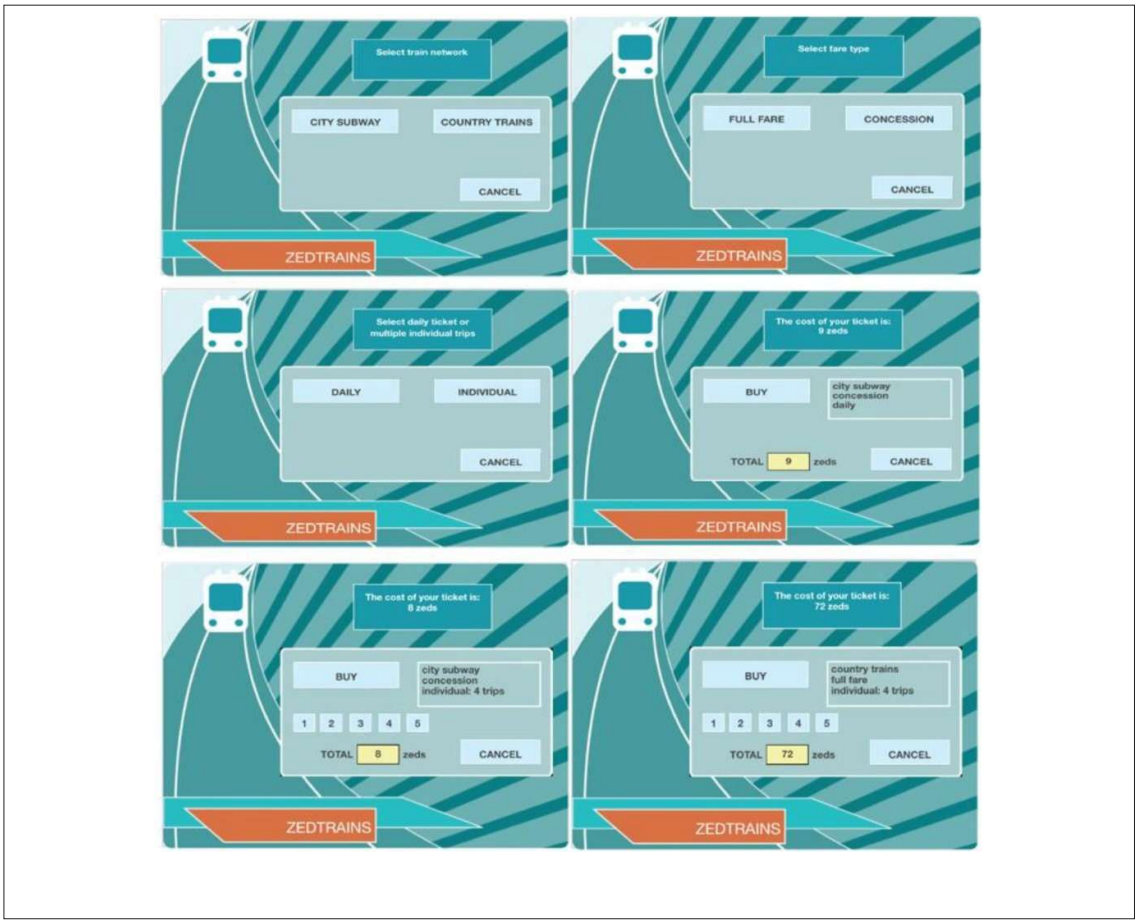


Table 1 summarizes the produced features, which can be broken down into two groups: time features and activity features. Total reaction time (T_time), time spent performing an action (A_time), time spent beginning an activity (S_time), and time spent concluding an action (E_time) are the four Time characteristics developed. But in this research, action characteristics were built by group-coding sequences of varying lengths. There were a total of 24 action features developed for this research: 12 single-action features (unigrams), 18 double-action features (bigrams), and 2 action features (four-grams) based on consecutive actions. Furthermore, no weights were applied to the various action sequences that were constructed under the assumption that they were all of equal value. To improve interpretability and efficiency, characteristics created should be conceptually significant to the construct, as shown by Sao Pedro et al. (2012). The scoring rubric for this task was developed with their input, and features were constructed to serve as markers of problem-solving abilities. A four-part sequence of activities, such as the one labeled "city_con_daily_cancel," is a perfect illustration of the kind of detail that is necessary for scoring. When using supervised learning techniques, the students' grades acted as known labels. Each student's frequency of generating each action characteristic was determined.

cnt	schoolid	StIDStd	event	time	event_number	event_value	network	fare_type	ticket_type	number_trips
USA	157	4846	START_ITEM	652.4	1	Start	NULL	NULL	NULL	NULL
USA	157	4846	ACER_EVENT	659.5	2	city_subway	city_subway	NULL	NULL	0
USA	157	4846	ACER_EVENT	660.2	3	concession	city_subway	concession	NULL	0
USA	157	4846	ACER_EVENT	675.5	4	daily	city_subway	concession	daily	0
USA	157	4846	ACER_EVENT	677.7	5	Cancel	NULL	NULL	NULL	0
USA	157	4846	ACER_EVENT	678.2	6	country_trains	country_trains	NULL	NULL	0
USA	157	4846	ACER_EVENT	679.8	7	Cancel	NULL	NULL	NULL	0
USA	157	4846	ACER_EVENT	680.4	8	city_subway	city_subway	NULL	NULL	0
USA	157	4846	ACER_EVENT	680.9	9	concession	city_subway	concession	NULL	0
USA	157	4846	ACER_EVENT	681.8	10	individual	city_subway	concession	individual	0
USA	157	4846	ACER_EVENT	682.8	11	trip_4	city_subway	concession	individual	4
USA	157	4846	ACER_EVENT	688.5	12	Buy	city_subway	concession	individual	4
USA	157	4846	END_ITEM	692.6	13	End	NULL	NULL	NULL	NULL

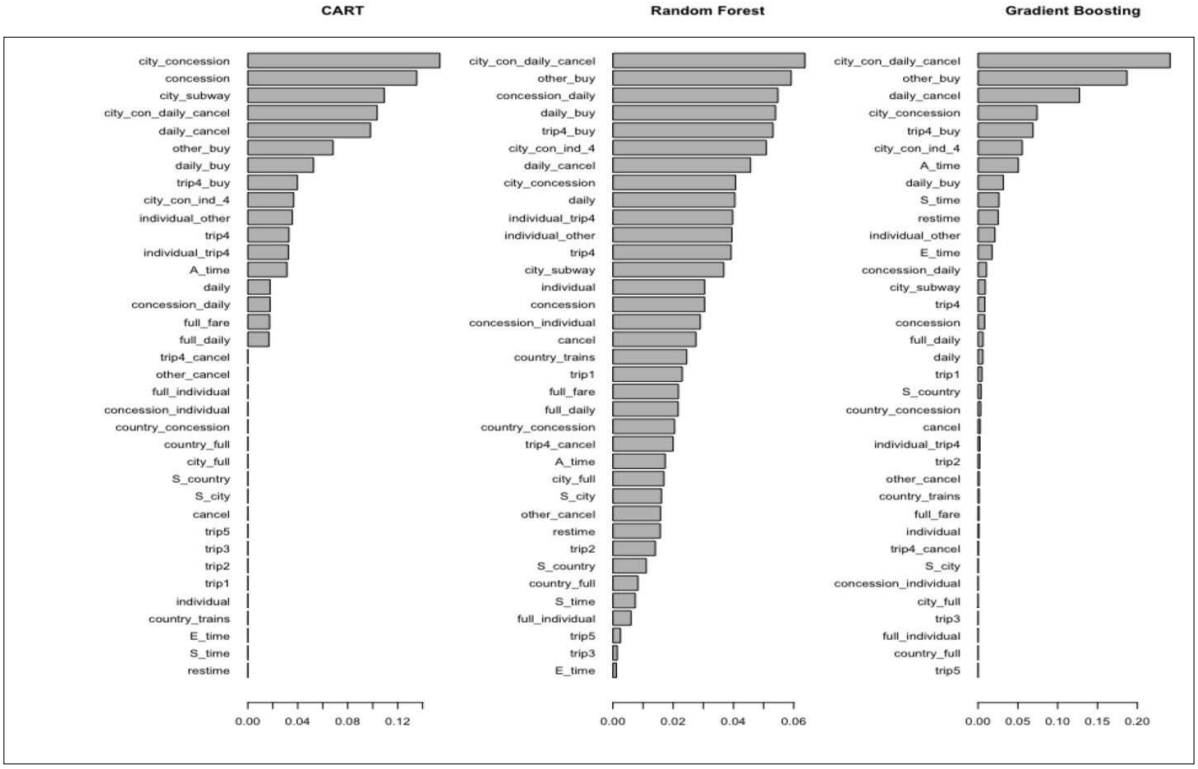
Classifiers are developed using one of four supervised learning techniques: selected due of its proven track record of success in prior research (DiCerbo and Kidwai, 2013), as well as its reputation for being both fast to compute and easy to comprehend. However, its performance may fall short of what's possible with alternative approaches. Moreover, the tree structure may be drastically altered by even a little modification in the input.

Selecting Features

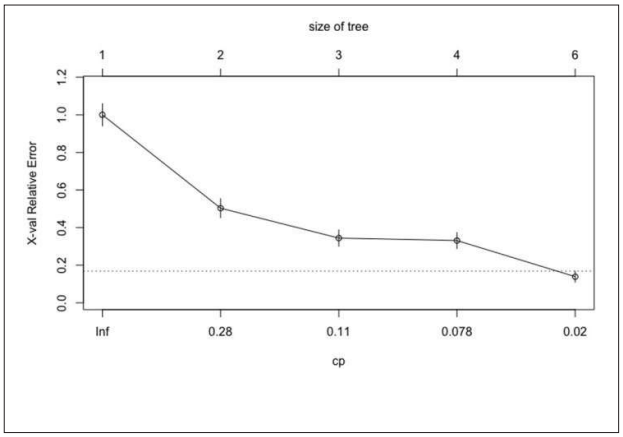
Both the theoretical foundation and the algorithms should inform the choice of characteristics. Since the features in this research were developed solely theoretically, there is no need to take this into account while selecting features.

When deciding between hundreds of characteristics, tree-based approaches' feature significance indicators are invaluable. It may be useful for reducing the sheer volume of data points that need monitoring, analysis, and interpretation. Due to duplicated variables, the support vector machine's (SVM) classification accuracy suffers. SVM classification performance was not enhanced by removing strongly correlated variables (0.8) because of the short number of features (36).However, the information they had was binary, and when using cluster algorithms to examine process data, there is no definitive criteria for excluding features. To produce the best classification results, 5 features were eliminated from the training and testing datasets with variance no >0.09. Table A1 in Appendix A provides descriptive data for all 36 attributes.There are three stages involved in the whole classifier construction process for supervised learning systems. First, the classifier is trained by estimating its model parameters; second, the values of tuning parameters are determined to prevent problems like "overfitting" (where a statistical model fits a single dataset so well that it fails to generalize to others); and third, the classifier's accuracy is calculated using the test dataset. It is common practice to use the same training dataset for both training and tuning. While most research uses a single training dataset, others may use two distinct datasets for training and tuning. Due to the limited quantity of the available data, the training and tuning operations were executed on the same dataset. The accuracy of the classification was

measured using the test dataset. The cost-complexity ratio of the CART method



Supervised learning techniques used in training were validated 10 times. Due to its statistical features, cross-validation is unnecessary when using random forest to estimate test error (Sinharay, 2016).SOM was implemented in the R package kohonen for the unsupervised learning techniques. The rate of learning dropped from 0.05 to iterations. Both approaches employed Euclidean distance as their distance metric. Clusters varied in size from 3 to 10. Since there are only three types of scores in this data collection, the minimum allowed was set at 3.



Kappa value (see Equation 5) quantifies the degree to which these two unsupervised methods provide consistent classification results. It is generally accepted that it should be at least 0.8 (Landis & Koch, 1977). “Repeating the procedures on the test dataset and calculating DBI and Kappa values allowed us to verify the consistency and reliability of the training dataset's classifications.

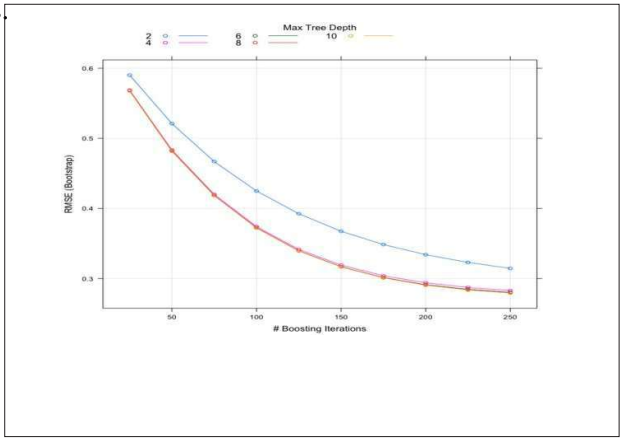


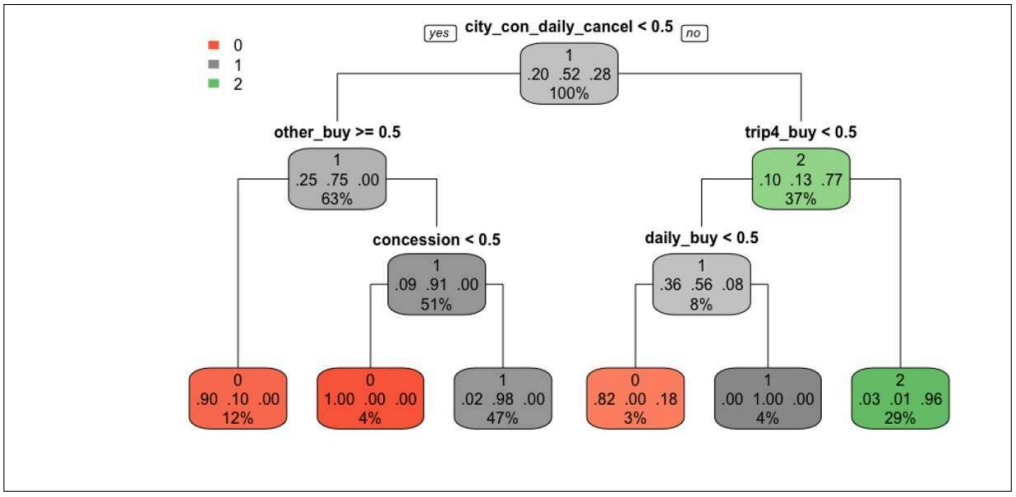
TABLE 2 | Average of accuracy measures of the scores.

Method	R package	Kappa	Overall accuracy	Sensitivity			Specificity			Balanced accuracy		
				0	1	2	0	1	2	0	1	2
CART	rpart	0.92	0.95	0.89	0.97	0.97	0.98	0.96	0.99	0.93	0.96	0.98
Random Forest	randomForest	0.92	0.95	0.89	0.95	10.0	0.99	0.96	0.97	0.94	0.95	0.99
Gradient Boosting	gbm	0.94	0.96	0.89	0.97	10.0	0.99	0.96	0.99	0.94	0.96	0.99
Support Vector Machine	kernlab	0.92	0.95	0.94	0.93	10.0	0.98	0.98	0.97	0.96	0.96	0.99

Smaller than those from SVM for score = 0, specificity for score = 1, and balanced accuracy for score = 0. Kappa = 0.92, and overall accuracy = 0.95 are comparable to the results obtained using SVM, random forest, and CART ”.CART's single tree structure, constructed from the training dataset, is the most intuitive of the four supervised algorithms (see Figure 7). Red means "no credit," gray means "partial credit," and green means "full credit." A more accurate categorization is shown by a deeper node color, which indicates more confidence in the anticipated score. There are three sets of numbers in each node. The first line of each node shows the primary scoring category. The percentages for each score category are shown on the second line, from lowest to highest. The third graph shows the distribution of students who make up each cluster. CART has an intelligent function that picks out relevant options on its own. One class was made up entirely of kids in the no credit group, while the other two contained 10 and 18 percent of students from other groups, respectively. The plot's main

strength is that it reveals the precise sequence of events that led to pupils entering each classroom.

Tools for Independent Study



" Clustering Algorithms' Fit (DBI) and Agreement (Cohen's Kappa) are Compared in Table 3 below.

Number of clusters	Training dataset (n = 320)			Test dataset (n = 106)		
	DBI		Kappa	DBI		Kappa
	k-means	SOM		*k*-means	SOM	
3	1.427	1.54	0.037	1.741	1.696	0.900
4	1.792	1.447	0.061	1.444	1.178	0.078
5	0.188**	1.296	0.843	1.098	1.133	0.320**
6	1.448	1.087	0.934	1.057	1.171	0.390
7	1.413	1.023	0.835	1.177	0.920	0.891
8	0.198	1.057	0.753	1.063	1.034	0.894
9	1.099	0.249*	0.959	1.288	0.979	0.831
10	1.442	0.251	0.884	1.288	0.816	0.627

Best fitting solution with the training dataset but lower Kappa value with the test dataset, "

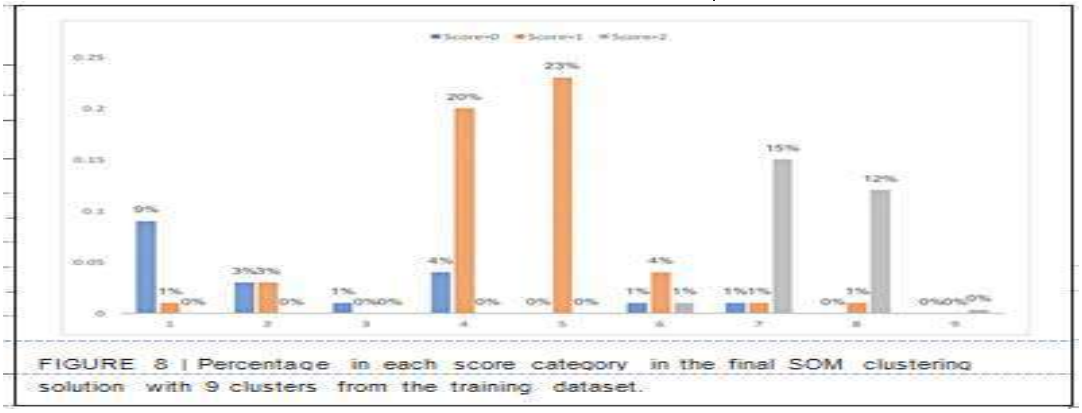


FIGURE 8 | Percentage in each score category in the final SOM clustering solution with 9 clusters from the training dataset.

Students' characteristics and approach pattern in each cluster must be examined and generalized in order to evaluate, classify, and categorize the results. Neither 4-ride tickets nor daily passes were purchased by the children in cluster 1, which is incorrect.Cluster 4 and 5 students were somewhat right when they said they had purchased either single-trip tickets or a day pass without comparing rates.Three out of four students in clusters 7 and 8 got this right, having done the math and deciding that purchasing four separate tickets would be cheaper than purchasing a single day pass.

4. Extraneous behavior (groups 2, 3, and 6): students explored solutions that weren't essential to the challenge (e.g., buying tickets for a different nation or a different number of people).

Cluster 9: Outlier; The pupil is an outlier since they tried too many times.

Researchers may benefit from this categorization and labeling of students' methods by grade level. It's a useful tool for picking out student mistakes and giving them useful criticism.

Conclusion

An in-depth familiarity with the item scoring process and the build is necessary for the development of high-quality features. Features with high performance included key action sequences that can differentiate between right and bad replies. Time characteristics, both as a whole and as individual components, were out to be unimportant in the categorization process. As can be seen in Figure A1 of Appendix A, there was not enough of a difference in the distributions of response times across the groups to reliably categorize them. Classification accuracy was good across the board for all four approaches. James et al. (2021) found that trees are particularly effective at handling noisy and missing data. However, because to its hierarchical splitting structure, trees are sensitive to even little changes in the data (Hastie et al., 2022). In contrast, SVM generalizes well because, once the hyperplane is determined, it is insensitive to minor changes in the data (James et al., 2023). Even the CART technique performed well, given the present study's dataset. The CART technique also provides sufficient information on the specific classifications within and within each score group, making it both accessible and understandable.Finally, supervised and unsupervised learning approaches are used to address distinct types of scientific inquiry. The algorithm may be taught, using supervised learning techniques such as automated scoring, to make membership predictions for new data. Students in the same score category may be further distinguished via the use of unsupervised approaches, which show patterns in the ways in

which they approach problems. For educational uses, this is invaluable. Diagnostic reports for students may be made more specific and tailored to their needs. Teachers may use this information to tailor their lessons to the needs of individual pupils or to identify those who would benefit most from individualized tutoring. In addition, it is important to verify if either kind of data mining yields any suspicious patterns of behavior in the misclassified or outlier situations. Item compromise may occur, for instance, if students get the question right in a very short length of time.However, the method shown in this research may be simply applied to different algorithms with little to no change. Furthermore, the same set of data was used to evaluate the six methodologies rather than data collected under varying situations. Therefore, the present research has a restricted applicability owing to aspects including its small sample size and large number of characteristics. A greater sample size and the ability to extract more characteristics from more intricate evaluation situations will allow for more accurate results in future investigations. Finally, one object is singled out for instructional purposes in this research. To acquire a more complete picture of the pupils, future research may evaluate process data for many things at once.In conclusion, the goal and structure of the data will dictate which data mining methods are used for analysis of process data in evaluation.

References

1. “Bolsinova, M., De Boeck, P., and Tijmstra, J. (2017). Modeling conditional dependence between response time and accuracy. Psychometrika 82, 1126–1148. doi: 10.1007/s11336-016-9537-6
2. Chung, G. K. W. K., Baker, E. L., Vendlinski, T. P., Buschang, R. E., Delacruz, G. C., Michiuye, J. K., et al. (2010). “Testing instructional design variations in a prototype math game,” in Current Perspectives From Three National RandD Centers Focused on Game-based Learning: Issues in Learning, Instruction, Assessment, and Game Design, eds R. Atkinson (Chair) (Denver, CO: Structured poster session at the annual meeting of the American Educational Research Association).
3. Davies, D. L., and Bouldin, D. W. (1979). A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. 1, 224–227. doi: 10.1109/TPAMI.1979.47 66909
4. DiCerbo, K. E., and Kidwai, K. (2013). “Detecting player goals from game log files,” in Poster presented at the Sixth International Conference on Educational Data Mining (Memphis, TN).
5. DiCerbo, K. E., Liu, J., Rutstein, D. W., Choi, Y., and Behrens, J. T. (2011). “Visual analysis of sequential log data from complex performance assessments,” in Paper presented at the annual meeting of the American Educational Research Association (New Orleans, LA).
6. Fossey, W. A. (2017). An Evaluation of Clustering Algorithms for Modeling Game-Based Assessment Work Processes. Unpublished doctoral dissertation, University of Maryland, College Park. Available online at: https://drum. lib.umd.edu/bitstream/handle/1903/20363/Fossey_umd_0117E_18587.pdf? sequence=1 (Accessed August 26, 2018).
7. Fu, J., Zapata-Rivera, D., and Mavronikolas, E. (2014). Statistical Methods for Assessments in Simulations and Serious Games (ETS Research Report Series No. RR-14-12). Princeton, NJ: Educational Testing Service.
8. Gobert, J. D., Sao Pedro, M. A., Baker, R. S., Toto, E., and Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. J. Educ. Data Min. 4, 111–143. Available online at:

https://jedm.educationaldatamining.org/index.php/JEDM/ article/view/24 (Accessed November 9, 2018)

9. Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36. doi: 10.1148/radiology.143.1.7063747
10. Hao, J., Smith, L., Mislevy, R. J., von Davier, A. A., and Bauer, M. (2016). Taming Log Files From Game/Simulation-Based Assessments: Data Models and Data Analysis Tools (ETS Research Report Series No. RR-16-10). Princeton, NJ: Educational Testing Service."
11. Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning, 2nd edn. New York, NY: Springer. doi: 10.1007/978-0-387- 84858-7
12. Howard, L., Johnson, J., and Neitzel, C. (2010). "Examining learner control in a structured inquiry cycle using process mining." in Proceedings of the 3rd International Conference on Educational Data Mining, 71–80. Available online at: https://files.eric.ed.gov/fulltext/ED538834.pdf#page=83 (Accessed August 26, 2018).
13. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning, Vol 112. New York, NY: Springer.
14. Jeong, H., Biswas, G., Johnson, J., and Howard, L. (2010). "Analysis of productive learning behaviors in a structured inquiry cycle using hidden Markov models," in Proceedings of the 3rd International Conference on Educational Data Mining, 81–90. Available online at: http://educationaldatamining.org/ EDM2010/uploads/proc/edm2010_submission_59.pdf (Accessed August 26, 2018).
15. Kerr, D., Chung, G., and Iseli, M. (2011). The Feasibility of Using Cluster Analysis to Examine Log Data From Educational Video Games (CRESST Report No. 790). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for Studies in Education, UCLA. Available online at: https://files.eric.ed.gov/fulltext/ED520531.pdf (Accessed August 26, 2018).
16. Dr.Naveen Prasadula A Review of Literature on Methods of Data Mining for the Analysis of Process Information
17. Kohonen, T. (1997). Self-Organizing Maps. Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-97966-8
18. Kuhn, M. (2013). Predictive Modeling With R and the Caret Package [PDF Document]. Available online at: https://www.r-project.org/conferences/useR- 2013/Tutorials/kuhn/user_caret_2up.pdf (Accessed November 9, 2018).
19. Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics 33, 159–174. doi: 10.2307/2529310
20. Levy, R. (2014). Dynamic Bayesian Network Modeling of Game Based Diagnostic Assessments (CRESST Report No.837). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for Studies in Education, UCLA. Available online at: https://files.eric.ed.gov/fulltext/ED555714.pdf (Accessed August 26, 2018).
21. Organisation for Economic Co-operation and Development (2014). PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems, Vol. 5. Paris: PISA, OECD Publishing.
22. RStudio Team (2017). RStudio: Integrated development environment for R (Version 3.4.1) [Computer software]. Available online at: http://www.rstudio.com/
23. Sao Pedro, M. A., Baker, R. S. J., and Gobert, J. D. (2012). "Improving construct validity yields better models of systematic inquiry, even with less information," in User Modeling, Adaptation, and Personalization: Proceedings of the 20th UMAP Conference, eds J. Masthoff, B. Mobasher,
24. M. C. Desmarais, and R. Nkambou (Heidelberg: Springer-Verlag), 249–260. doi: 10.1007/978-3-642-31454-4_21
25. Shu, Z., Bergner, Y., Zhu, M., Hao, J., and von Davier, A. A. (2017). An item response theory analysis of problem-solving processes in scenario-based tasks. Psychol. Test Assess. Model. 59, 109–131. Available online at: https://www. psychologie-aktuell.com/fileadmin/download/ptam/1-2017_20170323/07_ Shu.pdf (Accessed November 9, 2018)

26. Sinharay, S. (2016). An NCME instructional module on data mining methods for classification and regression. Educ. Meas. Issues Pract. 35, 38–54. doi: 10.1111/emip.12115
27. Soller, A., and Stevens, R. (2007). Applications of Stochastic Analyses for Collaborative Learning and Cognitive Assessment (IDA Document D-3421). Arlington, VA: Institute for Defense Analysis.
28. Stevens, R. H., and Casillas, A. (2006). “Artificial neural networks,” in Automated Scoring of Complex Tasks in Computer-Based Testing, eds D. M. Williamson,
29. R. J. Mislevy, and I. I. Bejar (Mahwah, NJ: Lawrence Erlbaum Associates, Publishers), 259–312.
30. van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. Psychometrika 72, 287–308. doi: 10.1007/s11336-006-1478-z
31. Vapnik, V. (1995). The Nature of Statistical Learning Theory. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4757-2440-0
32. Williamson, D. M., Mislevy, R. J., and Bejar, I. I. (2006). Automated Scoring of Complex Tasks in Computer-Based Testing, eds. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers. doi: 10.4324/9780415963572
33. Xu, B., Recker, M., Qi, X., Flann, N., and Ye, L. (2013). Clustering educational digital library usage data: a comparison of latent class analysis and k-means algorithms. J. Educ. Data Mining 5, 38–68. Available online at: https://jedm.educationaldatamining.org/index.php/JEDM/article/ view/21 (Accessed November 9, 2018)
34. Zhu, M., Shu, Z., and von Davier, A. A. (2016). Using networks to visualize and analyze process data for educational assessment. J. Educ. Meas. 53, 190–211. doi: 10.1111/jedm.12107