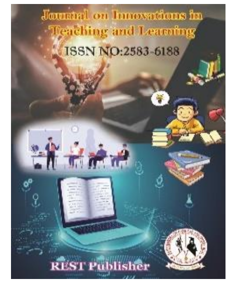
Journal on Innovations in Teaching and Learning

Vol: 2(1), March 2023

REST Publisher; ISSN:2583 6188

Website: http: //restpublisher.com/journals/jitl/

DOI: https://doi.org/10.46632/jitl/2/1/8



Student Success Performance Prediction Using Tree-Based Pipeline Optimization Method (TPOM)

***B. Kalaiselvi, P. Priyadharshini, K. Loheswaran, M. Rakkes**
Mahendra Engineering College, Namakkal, Tamil Nadu, India.
*Corresponding author email: kalaiselvib@mahendra.info

Abstract: *Predicting student success and analyzing student learning behaviors are critical components of improving online courses. By analyzing enormous amounts of clickstream data which tracks student behavior, instructors can get beneficial insights into the elements which impact students' academic results while recognizing opportunities for course development. In this research, a tree-based system is used in predicting student performance over academic for end-of-unit assignment through students' clickstream information. According to EDM Cup 2023 dataset, students with low capability are more probable to immediately ask for the response to an issue prior to intending a response, which has the potential to have detrimental effect on their learning results and might suggest attempts at manipulate the system. In the context of this purpose, we will anticipate growth in students while peering at the overall result through a device learning set of rules. The scholar's basic achievement over academic gets predicted using evaluating its earlier valuable data. When carrying out this prediction, the dataset utilized may assist in predicting student data. As a result, this study focuses on recognizing student success in academic performance utilizing the Tree-based Pipeline Optimization Method (TPOM) for Automatic Machine Learning (AutoML). In accordance with this pipeline, the Stochastic Gradient Descent (SGD) classifier performed better than other classifiers in the TPOM for AutoML in the EDM Cup 2023 dataset.*

Keywords: *Educational Data mining, student success, learning behavior, student performance prediction, Tree-based Pipeline Optimization Tool*

1. INTRODUCTION

In this current decade, large volumes of log data have been generated from student's who have attended through online courses will assist in providing significant details about their academics for analyzing student behavior as well as their effects on academic achievement. Modeling a student's understanding or capability with respect to their interactions with an educational system is a significant problem for modern learning methods. For example, a mastery-based tutoring system must evaluate student's understanding of a particular concept prior moving on to the subsequent one. In a similar way an examination preparation program must assess students' aptitude for their recommend appropriate items [1]. To offer appropriate vocabulary assessment, individualized language acquisition platforms must first analyze each student's forgetfulness. Furthermore, universities must monitor the academic achievement of student's for determining individuals who are at risk of falling remaining prompting instructors to intervene quickly. The commonly accepted solution, information tracking, addresses this challenge as a sequence forecasting issue, seeking to reliably forecast the accuracy of a student's response to the next question according to their prior experiences with the learning environment. Nevertheless, it is uncertain whether forecasting the accuracy of the student's answer to the following question is the ideal strategy to represent their capacity, as the accuracy of a student's respond may be affected by factors other than their proficiency. Insufficient involvement behaviors like fast inference may have an adverse effect on student's current performance, but they are not accurate indicators of low skills. As a result, if both models possess identical capability to predict students' abilities but a single can determine students' involvement using their best current answering questions performances whereas the other are unable the model that is capable of determining engagement is deemed superior in the understanding tracking job. This is true regardless of whether the two models are equally effective in representing the skills of students. It made an analogous case, emphasizing the framework over their immediate accomplishment inside it [2] [3]. They observed models that succeed at measuring the knowledge volume for the students keep once employing the learning environment frequently fail to do well at demonstrating the skills of learners that are determined using results from post-tests.

Assessing a student's prior educational actions, especially it replies to previous examinations may assist in forecast their success. An understanding tracking may offer understanding regarding student learning behavior, allowing administrators to tailor their methods of instruction to meet the requirements of their students [4]. Furthermore, this information enables learners to receive tailored feedback, which improves their educational experience and, eventually, their final results [5]. On keeping with this goal, the EDM Cup 2023 has been created as an occasion to forecast students' end-of-unit-assignment scores on math problems obtained from the Assessing online system. The EDM Cup gave students accessibility to a massive clickstream dataset containing millions of student activities and additional curricular data, allowing them to estimate their end-of-unit task grades according to their behaviors throughout in-unit tasks. EDM is a novel or improved DM technique that can be effectively utilized for education-related data. Data may have been obtained from previously utilized data, as well as operational data stored in educational organization databases. Data about students may include private data as well as academic achievement. It can also be accomplished with e-learning database platforms that incorporate an enormous quantity of information and data utilized by numerous organizations. It employs a variety of strategies to effectively execute data mining principles including K-means clustering as well as K-nearest neighbor. Employing association rules, categorization, and clustering, several types of data can be identified. This allows us to acquire data that represents student performance on exams as well as all of their specific details. The initial phase in sorting through this enormous quantity of data is cluster analysis, which is used for categorizing the initial information in an appropriate way.

EDM refers to the use of DM methods for collecting and analyzing massive data amount in educational sectors for improving the process of teaching and learning [6]. EDM is described as the extraction of novel knowledge from huge educational data amount acquired in the educational setting as well as kept in educational databases [7]. EDM is a field of research which concentrates on extracting useful details from complicated datasets using methods like DL, ML, and statistical analysis. Prediction algorithms are utilized for estimating the likelihood that students will pass or fail an examination, as well as finish or fail an assignment, course, or studying. In this scenario, a technique for classification may be applied. By learning the algorithm on past data, it discovers patterns and associations that enable it to generate probabilistic forecasts about the performance of students on exams. The linear regression approach is employed for predicting learners' academic achievement. Clustering algorithms indicate where learning resources ought to be enhanced as well as which instructional materials students might use to prepare for tests [8]. Identifying sequential patterns makes it possible to describe learner behavior patterns which contribute to specific learning outcomes. Data visualization is frequently used to demonstrate the speed at which a specific learning content is acquired and to better comprehend student learning trends, results, and consequently. It can also be feasible to combine models, like clustering a learners group afterwards applying classification for predicting a specific learner's performance [9].

Academic achievement may be anticipated using relationships with learners, questionnaires and evaluations, as well as EDM. Success in education can be described as a broad term encompassing academic accomplishments, participation in the learning process, fulfillment encountered over education, obtained knowledge and abilities throughout studying, conquering challenges with learning, extensive learning, positive professional advancement, and accomplishing educational objectives [10]. For past decade, researchers are developing a TPOM that autonomously constructs as well as optimizes ML pathways for an individual problem domain with no the need for human assistance. Generally, TPOM automating the development of software for computers by optimizing ML procedures using a variant of the developing computing technology referred to as Genetic Programming (GP). However, the valuable study demonstrates that GP as well as Pareto optimization collaborate to enable TPOM frequently design high accuracy, efficient pipelines, as well as frequently enhance basic ML investigations.
As a result, this study emphasizes ML approaches to design a successful model to recognize adult learners' interests in order to give them with specific rewards and conduct programs to increase their skill development, as well as providing additional assistance in enhancing student success prediction.

2. LITERATURE REVIEW

At present, data from click streams has been gaining popularity as a promising tool for predicting course-level success [11]. This data type records behavior of students' learning like accessing various educational materials or performing exams, and provides useful data on student interaction inside the online learning environment [12]. Asadi et al. suggested a technique for predicting performance among students using graph neural networks that analyzes raw, multi-dimensional clickstream data with no depending upon feature extraction. Researchers tested their methods on 23 MOOCs as well as discovered it performs comparably to models with human-engineered features [13]. In an earlier study, Al-azazi and Ghurab have created an approach known as Artificial Neural Network with Long Short-Term Memory (ANN-LSTM), that is an RNN method. They trained an approach on a MOOC's clickstream data for predicting student achievement prior to the course. The suggested ANN-LSTM attained an accuracy of over 70%, beating classic ML approaches like Decision Trees, K-nearest neighbours, as well as Deep Learning (DL) methods like RNNs and Gated Recurrent Units (GRU) [14].

In accordance with the initial findings by Fensie et al, there is a lack of study on the adult educational process enrolled in distant education. Most of the research focuses on learner opinions yet experience as the main sources of data, but several papers make recommendations for communication with adult learners in distant learning with no offering any support to back up their claims. Language acquisition and continuing medical education are the two primary classifications utilized for describing and measure the processes of learning for adult learners. The final criteria for inclusion in scoping reviews have to be determined after multiple iterative searches and presentations assessments. Several students contributed their skills and expertise to this team's effort, resulting in a more thorough investigation and a deeper understanding of the literature [15]. As an outcome of the current global epidemic, Garretet al's use of distance education became widely recognized. Assessing the current state of observational as well as theoretical research on adult learning in distant education becomes more important as the nontraditional learner's amount grows and worldwide occurrences interrupt learning [16].

Palacios et al. have described the Student success Prediction (SPP), which aims to forecast a student's success prior to engaging in an educational program or passing a test [17]. Snjeˇzana Kriˇzanic has presented data gathering, problem formulation, practical uses, methodologies, as well as prediction. The five DM stages will be utilized to achieve the target. There are three primary kinds of education offered are offline, online, and blended learning. The purpose of the analyzing data technique known as DM is to identify hidden information within the source data. Decision trees and cluster analysis have been applied to analyze educational data from Croatian higher education institutes [18]. Cluster analysis groups students shared, while DT are used for more in-depth group assessments. A different approach of DM method provided by TismyDevasiaet al., classification, is used at Amrita VishwaVidyapeetham Mysore for predicting student division determined by past data. The naive Bayesian mining technique provides higher precision for extracting beneficial knowledge from data whereas the s tudent educational background has been considered [19]. Based on Wang et al.'s research on the amount of automating required by data scientists, developing an AutoML tool with a human inside the loop is preferred to completely automating ML operations [20]. Drozdal et al. have performed qualitative investigations to identify the elements which impact trust in ML models built with AutoML tools. They revealed that data scientists exhibit greater trust in AutoML solutions which incorporate non-functional factors such as performance measures, transparency, and visualization [21]. Furthermore, it investigated the results of student's in various issue kinds as well as themes to discover regions of difficulty for the students. Therefore, the learning behaviors of successful students have been contrasted with the behaviors of failing students for identifying different behavioral patterns among the two classes.

3. RESEARCH METHODOLOGY

Advances in technology is currently enhancing all students in the online courses individuality as well as quality of life through enabling individuals to form and maintain social relationships, receive access to care and services, promote their ongoing educational efforts, and even have leisure and entertain themselves. The application of the AutoML approach in this study helped to detect the interests of student academic success in the obtained dataset. This study concentrates on identifying the skills of particular students who have not succeeded in their academic as well as their enthusiasm in techniques of learning for supporting future progress towards succeeding in their academic educational lifestyle. TPOM was created as an AutoML to accomplish specified tasks. The tree-based AutoML tool is being invoked with multiple classifiers. The optimization pipeline connected with the tree type modeling is used for executing the analysis in determining the student's success.

This study's dataset obtain from the EDM Cup 2023 competition 2 that is hosted on Kaggle. This dataset is resulted with K-12 students using the Assignment-3 educational platform to complete their math's homework and exams. Table 1 depicts the mathematics course of study on Assignment, which is split into distinct units, each comprising a series of closely connected mathematical ideas. According to pedagogical process in each unit, teachers distribute homework to students whereas the dataset refers to as 'in-unit assignments'. After completing a unit, instructors evaluate student's understanding and achievement of unit competencies using unit-test assignments. The objective is to create a precise forecasting model that can forecast when a student will accurately reply on every item in the unit-test assignments determined by their performance on the associated in-unit assignments. The EDM Cup 2023 dataset includes clickstream data from the assessment platform as students completed in-unit assignments. This clickstream data consists mostly of time-stamped operations.

TABLE 1. Student response in in-Unit assignments

Student Response	Action in response (%)
Correctly Respond	61.15
Wrongly Respond	26.92
Requested answer	10.28
Requested hint	1.26
Requested explanation	0.35
Skill associated video requested	0.03
Requested for live tutor	0.02

The descriptive statistics for dataset contains 34,652 students, 53,615 test assignments and 607,236 in-unit assignments. The in-unit assignments include 57,235 discrete instance, whereas the unit-test assignments comprise an additional 1835 instances. There is no overlap among in-unit as well as unit-test assignments. Clickstream data is available for in-unit assignments yet unavailable for unit-test assignments. As a result, while predicting if a student will accurately respond to the question on a unit-test assignment, researchers solely employ clickstream data from the relevant in-unit assignments as an input features. The tasks included in the dataset have been separated into three categories namely in-unit assignments, unit-test assignments with instance scores, and unit-test assignments without instance scores. The competition administrator gave test assignments with instance scores for models for practice throughout the competition. The competition host scored models on the public as well as private leaderboards using test assignments that did not include item scores. Researchers organized students' activities when responding a specific question chronologically, resulting in a procedure sequence.

TABLE 2. Frequency of Student response in in-Unit assignments

Student Response	Action in response (%)	Cumulative (%)
Correctly Respond	67.52	67.52
Wrongly Respond, Correctly Respond	9.60	77.12
Requested answer, Correctly Respond	5.95	83.07
Wrongly Respond, Requested answer, Correctly Respond	4.55	87.62
Wrongly Respond X 2, Correctly Respond	2.34	89.96
Wrongly Respond X 2, Requested answer, Correctly Respond	2.48	92.44
Wrongly Respond X 3, Requested answer, Correctly Respond	1.87	94.31
Wrongly Respond X 3, Correctly Respond	0.60	94.91
Wrongly Respond	0.51	95.42
Help Requested, Correctly Respond	0.38	95.80
Others	4.20	100

Table 2 shows that 67.52% of students properly responded to an instance on their initial attempt, while 9.60% found out the solution following a single failure. Surprisingly, the ten most frequently occurred with response sequences make up 95.80% of overall response sequences found in this research dataset. As per data preprocessing the onehot encoder is utilized as predominant sequence that convert each feature input into unique binary vector in which single bit is on as 1 and off as 0. Each unit vector represents one particular action sequence.

4. WORKING OF TPOM

AutoML refers to the process of completely automating the application of ML to real-world situations. The key challenge with TPOM is that the suggested algorithm must identify an acceptable set of operations for every stage in the ML pipeline to eliminate bias. Based on analytical, equation 1 depicts TPOM.

$$O_P C_{O_S} + 2^{N.G(f1,f2)} P_{N_{Max}} + \sum_{m' \in Max} \sum_{r \in R} P\left(\langle m'.r\mathrm{I}m\rangle P\big(r + y.v(m')\big)\right) \tag{1}$$

Where,

O_P = operation set of Pre-defined as default
O_S = operations selection of an algorithm
$G(f1, f2)$ = Novel features development using generator function
N = Number of selected features
N_{Max} = Number of maximum selected features

The automated data pre-processing may be carried out using a action set selected (OS) from the usual operating set (OP) and executed in the dataset. The feature extraction occurs by selecting relevant features (2N) from the dataset utilizing identification as well as generating current dependant pairings as (G (f1, f2)). Selecting a model and hyperparameter use are achieved by optimization in determining the optimal parameter configuration from an endless search area or by learning from previously created models for specific objectives. The final part in the equation represents the stochastic learning strategy that was used for several decades to confine the configuration space.

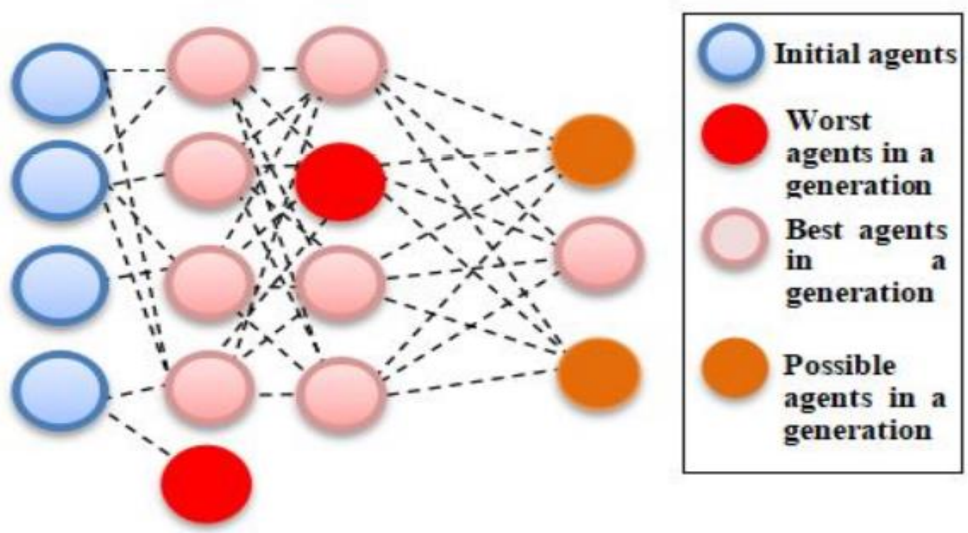


FIGURE 1. Working of evolutionary TPOT algorithm

Figure 1 depicts the implementation of TPOM with an enhanced strategy in the optimization process using GP to discover the ideal ML pipeline. In general, the creation of mathematical functions in trees has been optimized using fitness criteria such as classification accuracy. Every tree generation has been created through random changes of the tree's structure or the operation that occurs in each node of the tree. Iterating this procedure with respect to the quantity of training generated yields the best tree. The iteration approach to the training iteration number yielded an ideal tree. This can be used to create optimized ML pipelines that help improve and outperform typical supervised ML algorithms. Thus, the suggested TPOM is evaluated in relation to the classifier accuracy of each iteration. Selection, mutation, and crossover procedures have been used to improve the GP approach for determining the optimum pipeline. Thus, the research activity entails assessing the TPOM efficiency in relation to determined hyperparameters for predicting student success in academic performance and the reliability of student's skills for future development.

TPOM is a computational tool that does intelligent search using an ML pipeline. It includes feature selection methods, supervised classification models, a preprocessor, and certain estimators or transformers that follow the scikit-learn API. TPOM is built using a variety of packages, including NumPy, SciPy, update_checker, scikit-learn, stopit, pandas, tqdm, xgboost, and joblib. Furthermore, prior to import the AutoML model, the packages must be installed in Python using pip install tpot. There are numerous pipeline combinations obtained from TPOM that are made up of data transformers accessible in the scikit-learn Python library, such as Min-Max Scaler, Max Abs Scaler, Normalizer, Standard Scaler (SS), polynomial feature expansion, and Binarizer, for preprocessors, as well as Recursive Feature Elimination (RFE), Variance Threshold, and Select Percentile (SP) for selectors. Therefore, TPOM has provided several custom features such as hot encoder, stacking estimator (SE), sklearnstransformer applications and zero counts. Moreover, the TPOM time complexity is indicated as $\partial(k)$. Where, k = Tree depth, Whenever the subalgorithm has temporal complexity is represented as $\partial(1)$. The tree has formed with depth k in all types of targets, including binary variable and multivariable targets. Assume the sub-algorithm has a temporal complexity of $\partial(1)$ owing to n iterations, and the TPOM executes for $\partial(k)$.

5. RESULT AND DISCUSSION

The goal of this suggested TPOM model with updated hidden layer is to determine the learner's interest by using a variety of tree-based optimizers. The optimizer's goal is to commence a higher learning rate (alpha =0.01) during training. Table 1 shows the many TPOM classifier models, including the SGD classifier, Random Forest Tree (RFT) classifier, XGBoost classifiers and DT classifier. The output layer with sigmoid has three distinct target values low, medium, and high. The low is represented as "0", the medium as "1", and the high as "2". Student academic performance is tested using six different questionnaires, as well as test scores. From the various TPOM models shown, the SGD classifier determined the optimum TPOM classifier model for detecting student success. The currently best Cross Validation (CV) is calculated with the highest accuracy, is defined in this experimental investigation, and the accuracy is presented in figure 2.

```
Generation 1 - Current best internal CV score: 0.9185521885521885
Generation 2 - Current best internal CV score: 0.9185521885521885
Generation 3 - Current best internal CV score: 0.9357239057239057
Generation 4 - Current best internal CV score: 0.9357239057239057
Generation 5 - Current best internal CV score: 0.9357239057239057
```

FIGURE 2. 5fold CV score for SGD classifier method in TPOM

Figures 3 demonstrate the SGD Classifiers that have been used to evaluate the model's accuracy using the confusion matrix class and associated metrics. The value of each class in the TPOM classifier.

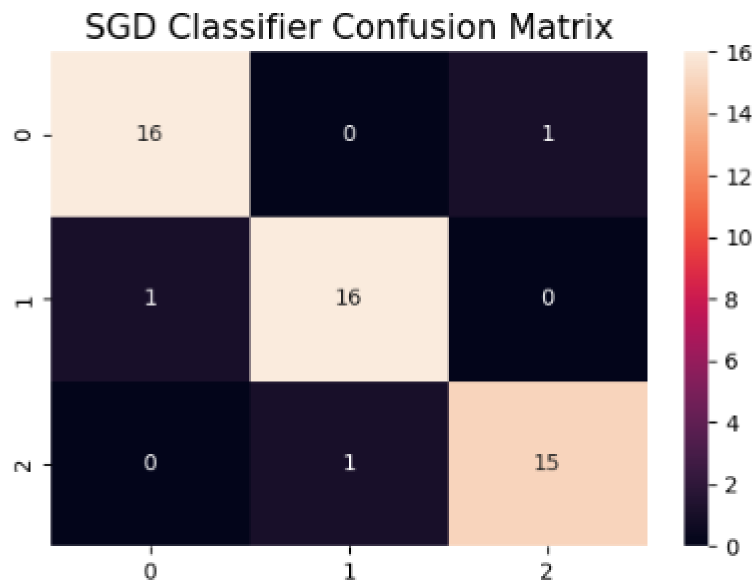


FIGURE 3. Confusion matrix of SGD classifier for student success prediction

TABLE 3. TPOM with various classifier methods

Classification of TPOM	Confusion matrix metrics					
	Micro Precision	**Micro Recall**	**Micro F1-Score**	**Weighted Precision**	**Weighted Recall**	**Weighted F1-Score**
SGD Classifier	95.06	95.06	95.06	94.0	94.0	94.0
RFT Classifier	86.17	86.17	86.17	86.32	86	86.01
DT Classifier	85.89	85.89	85.89	86.54	86	85.98
XGBoost Classifier	92.52	92.52	92.52	90.54	92	91.99

Table 2 shows the evaluation of the TPOM model using 5 generations of cross validation and squared_hindge as the loss parameter. The accuracy of a multivariable classifier may be measured using micro precision, micro recall, and micro F1-Score, however the SGD classifier has a high accuracy of 95.06% when compared to other TPOM classifiers.

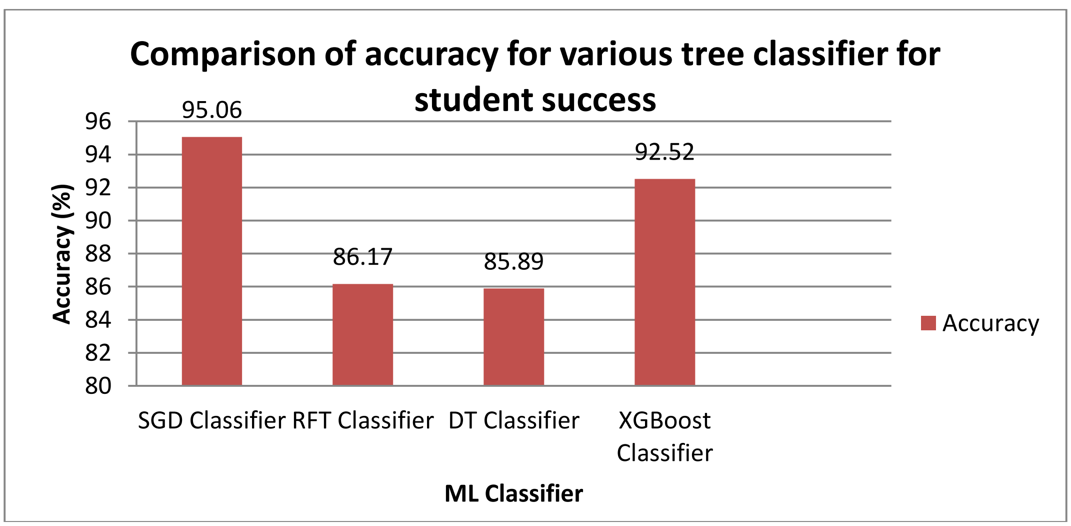


FIGURE 4. Comparison of accuracy for various TPOM models

Figure 4 shows the accuracy for four different TPOM models. SGD has a micro F1-Score of 95.06%, which is greater than other TPOM models. The proposed TPOM model of SGD classifier performs better in categorizing student success prediction from their center.

6. CONCLUSION

The concept of development has accumulated knowledge of people who have progressed through many periods of their lives and have performed important and diverse social and cultural roles and responsibilities. Certain provisions must be considered while identifying techniques for recognizing and responding to the set of practices and situations encountered by persons outside of the classroom. Simultaneously, the facility needed to understand the student i online course interests and development practices, as well as apply pressure on the learner to meet the staff's expectations about the appropriate model of teaching and behavior for the student success in order to enhance the learner's interest in learning more and spending valuable time. As a result, when compared to existing TPOM models, the proposed TPOM model with SGD classifier has a higher learning rate and an accuracy of 95.06% after five rounds of cross validation. The TPOM model is recommended since it has the most feature selection options for an improved comprehension of the model. This TPOM model with SGD classifier is critical in resisting further pressure on non-interested learners to increase their skills. On the contrary, the student is directed to focus on basic studies in order to stimulate their interest by altering the manner of instruction for the student to be successful in their academic.

REFERENCES

[1]. LOH, H., SHIN, D., LEE, S., BAEK, J., HWANG, C., LEE, Y., CHA, Y., KWON, S., PARK, J., and CHOI, Y. 2021. Recommendation for effective standardized exam preparation. In LAK21: 11th International Learning Analytics and Knowledge Conference. LAK21. Association for Computing Machinery, New York, NY, USA, 397–404.

[2]. SCRUGGS, R., BAKER, R., AND MCLAREN, B. 2020. Extending deep knowledge tracing: Inferring interpretable knowledge and predicting post-system performance. In Proceedings of the 28th International Conference on Computers in Education, H. J. S. et al., Ed. Australia: Asia-Pacific Society for Computers in Education.

[3]. SCRUGGS, R., BAKER, R. S., PAVLIK, P. I., MCLAREN, B. M., AND LIU, Z. 2023. How well do contemporary knowledge tracing algorithms predict the knowledge carried out of a digital learning game? Educational technology research and development 71, 3, 901–918.

[4]. WANG, F., HUANG, Z., LIU, Q., CHEN, E., YIN, Y., MA, J., AND WANG, S. 2023. Dynamic cognitive diagnosis: An educational priors-enhanced deep knowledge tracing perspective. IEEE Transactions on Learning Technologies 16, 3, 306–323.

[5]. JANG, Y., CHOI, S., JUNG, H., AND KIM, H. 2022. Practical early prediction of students' performance using machine learning and explainable AI. Education and Information Technologies 27, 1–35.

[6]. Batool, S.; Rashid, J.; Nisar, M.W.; Kim, J.; Kwon, H.Y.; Hussain, A. Educational data mining to predict students' academic performance: A survey study. Educ. Inf. Technol. 2023, 28, 905–971. [CrossRef]

[7]. Romero, C.; Ventura, S. Educational data mining and learning analytics: An updated survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2020, 10, e1355.

[8]. Križani´c, S. Educational data mining using cluster analysis and decision tree technique: A case study. Int. J. Eng. Bus. Manag. 2020, 12, 1847979020908675.

[9]. Almasri, A.; Alkhawaldeh, R.S.; Celebi, E. Clustering-based EMT model for predicting student performance. Arab. J. Sci. Eng. 2020, 45, 10067–10078.

[10]. Arulkadacham, L.; McKenzie, S.; Aziz, Z.; Chung, J.; Dyer, K.; Holt, C.; Mundy, M. General and unique predictors of student success in online courses: A systematic review and focus group. J. Univ. Teach. Learn. Pract. 2021, 18, 7.

[11]. ROHANI, N., GAL, K., GALLAGHER, M., AND MANATAKI, A. 2023. Early Prediction of Student Performance in a Health Data Science MOOC. In Proceedings of the 16th International Conference on Educational Data Mining, M. Feng, T. K¨aser, and P. Talukdar, Eds. International Educational Data Mining Society, 325–333.

[12]. BAKER, R., XU, D., PARK, J., YU, R., LI, Q., CUNG, B., FISCHER, C., RODRIGUEZ, F., WARSCHAUER, M., AND SMYTH, P. 2020. The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes. International Journal of Educational Technology in Higher Education 17, 1, 1–24.

[13]. ASADI, M., SWAMY, V., FREJ, J., VIGNOUD, J., MARRAS, M., AND K¨ASER, T. 2023. Ripple: Concept-based interpretation for raw time series models in education. In Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence, B. Williams, C. Yiling, and J. Neville, Eds. Vol. 37. AAAI Press, 15903–15911.

[14]. AL-AZAZI, F. A. AND GHURAB, M. 2023. ANN-LSTM: A deep learning model for early student performance prediction in mooc. Heliyon 9, 4. Article e15382.

[15]. Fensie, Anne & Gill, Parm&Rogowski, Aubrey &Bellnier, Karen &Stidham, Sharon & Jones, Melissa & Wiley, Linda & Pierre, Teri &Alicea, Megan &Wehr, Katrina & Clark-Stallkamp, Rebecca & Krieger, Meryl &O'Mahoney, Aoife &Stalford, Lauren &Budhrani, Kiran&Mazlan, Nurul&Avci, Hulya. (2022). The Study of Adult Learners in Distance Education: A Scoping Review of the Literature. 10.13140/RG.2.2.32441.44644.

[16]. Garret, R., Legon, R., Simunich, B., &Fredericksen, E. E. (2021). CHLOE 6: Online learning leaders adapt for a post-pandemic world (The Changing Landscape of Online Education). Quality Matters and Eduventures.http://qualitymatters.org/qaresources/resource-center/articles-resources/CHLOE-project

[17]. C. A. Palacios, J. A. Reyes-Suárez, L. A. Bearzotti, V. Leiva, and C. Marchant, "Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile", Entropy, vol. 23, no. 4, pp. 1–23, 2021.

[18]. Snjeˇ zanaKriˇ zanic, "Educational Data Mining Using Cluster Analysis And Decision Tree Technique: A Case Study", International Journal of Engineering Business Management, 2020.

[19]. Ms.TismyDevasia, Ms.Vinushree T P, Mr.VinayakHegde, "Prediction Of Students Performance Using Educational Data Mining", ResearchGate, 2020.

[20]. D. Wang, Q. V. Liao, Y. Zhang, U. Khurana, H. Samulowitz, S. Park, M. Muller, and L. Amini, "How much automation does a data scientist want?" arXiv preprint arXiv:2101.03970, 2021.

[21]. J. Drozdal, J. Weisz, D. Wang, G. Dass, B. Yao, C. Zhao, M. Muller, L. Ju, and H. Su, "Trust in automl: exploring information needs for establishing trust in automated machine learning systems," in Proceedings of the 25th International Conference on Intelligent User Interfaces, 2020, pp. 297–307.