International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 3
May-Jun 2026, PP 01-09

Modified Teacher Learning Algorithm Based Privacy Preserving of Student Data for Grade Prediction

Sharjil Iqbal
Senior Data Engineer
Department of Marketing, University

Abstract. In educational organizations, early evaluation prediction is a significant place of enthusiasm as it permits educators to improve performance in their courses by giving uncommon consideration at the beginning stages. This paper has proposed a model that identifies the effective features from student's data for grade prediction. Paper has hide the sensitive information in the dataset to provide the privacy for the student and organization data. Feature selection is done by Modified Teacher Learning Algorithm. Work has utilized the selected feature for the training of error back propagation model for grade prediction. Experiment was done on real dataset and result shows that proposed work has improved the grade prediction accuracy.

Keywords: Data mining, Genetic Algorithm, Privacy Preserving, Neural Network.

I. Introduction

During academic years, students are evaluated on the basis of both academic and practical performance. Practical performance includes participation in scholarships, research competitions, project activities, and performance in different subjects. Students taking admission in the same college or university generally have similar scores at the initial stage, but as their academic journey progresses, noticeable differences in performance begin to appear due to several influencing factors.

Predicting student academic performance has become an important responsibility in the education sector because every educational institution aims to achieve better student outcomes. Student performance is affected by multiple factors such as demographic characteristics, personality traits, family and socioeconomic background, learning behavior, and surrounding environmental conditions [1, 2]. Understanding these factors helps educational institutions to identify weak areas and take necessary actions for improving student learning.

With the advancement of digital education systems, platforms such as learning management systems collect large volumes of student-related data [3]. This data can be effectively utilized to predict student grades and performance trends. Accurate prediction of student performance enables academic staff to identify students who may require additional support and allows timely corrective actions before final examinations. The primary objective of this research is to develop a predictive model for forecasting student grade performance and analysing the impact of various learning-related features on student academic success.

By applying different Machine Learning (ML) techniques, educational data can be effectively utilized for predicting student performance at various academic levels using multiple influencing factors [4]. The analysis of such data provides meaningful insights to teaching staff, helping them make informed decisions for improving student learning outcomes [5]. Predictive ML models offer an efficient solution by identifying academically at-risk students at an early stage, allowing teachers to provide timely guidance and support [6]. These predictions also help instructors modify teaching strategies, learning materials, and academic approaches according to the specific needs of individual students, thereby improving the overall effectiveness of the learning process.

II. Literature Survey

Olabanjo et al. [7] designed a Radial Basis Function Neural Network model for predicting the academic performance of secondary school students in Nigeria. For this purpose, they prepared a dataset using school archival records that contained raw examination scores and teacher evaluations of 1927 students from class one to class six. The prediction model considered several input parameters such as psychomotor assessment, average marks obtained in Mathematics, English, and other major subjects, along with teacher performance ratings to determine whether a student would successfully pass the West African Examination Council examination.

Abid et al. [8] presented a machine learning-based framework for identifying student academic performance with improved prediction accuracy. Their work focused on overcoming challenges faced in higher education systems. During analysis, they utilized the Kaggle dataset to study student academic activities. The collected data was first preprocessed using missing value replacement techniques to eliminate duplicate and incomplete entries. After preprocessing, various classification models such as Support Vector Machine, Random Forest, and Neural Networks were applied for student performance recognition. Additionally, Principal Component Analysis was used for feature selection, which significantly improved prediction accuracy.

Ani et al. [9] developed machine learning models for predicting student performance in Massive Open Online Courses (MOOCs). Their model used demographic details, academic background, and student interaction with course materials as major input parameters. The obtained results demonstrated that such predictive systems can effectively forecast student performance, identify factors responsible for slow academic progress, and support the implementation of suitable teaching strategies. This approach also enables educators to detect academically at-risk students at an early stage and provide timely intervention for performance improvement.

Alshaikh and Hewahi [10] proposed a student performance prediction framework based on Convolutional Neural Networks (CNN) to enhance the effectiveness of intelligent tutoring systems. Their research focused on integrating advanced learning methods for improving prediction efficiency. The study utilized data obtained from Kaggle, where decision tree analysis and Principal Component Analysis were applied to identify the most relevant features from the dataset. These selected features were then provided as input to the CNN classifier. Additionally, a resampling mechanism was implemented to handle data imbalance problems, resulting in improved classification performance.

Li et al. [11] introduced a support vector machine-based approach for converting qualitative student feedback into quantitative information for course design optimization. Their system also incorporated Graph Convolutional Neural Networks to recommend suitable educational resources based on the cognitive abilities of individual students, thereby improving academic outcomes. The proposed framework was experimentally evaluated in School A to assess its effectiveness in recommending relevant course topics during the teaching process.

Li et al. [12] developed a complete end-to-end deep learning framework for academic performance prediction. Their model was designed to automatically extract useful features from heterogeneous student behavioral data collected from multiple sources. By integrating these diverse datasets, the system was able to predict student academic performance with improved efficiency and reduced dependency on manual feature engineering.

Zhai et al. [13] presented an interpretable early warning prediction model for identifying academic crises among college students. The main objective of their work was to detect students likely to perform poorly at an early stage and to examine the factors that positively or negatively influence academic achievement. The proposed approach provided clear insights into the relationship between different student-related factors and grade performance, enabling educational institutions to take timely corrective actions.

III. Proposed Methodology

In order to work on student data for grade prediction this work proposed a model term as Student Grade Prediction by Modified Teacher Learning Model (SGPMTLM). Explanation of model is each paragraph and fig. 1 shows models block diagram. In order to maintain the privacy of the work original data was suppressed in the work. In this proposed work input data was pre-processed and feature optimization was done by modified teacher learning optimization algorithm. Optimized feature were used for the traiiing of machine learning mathematical model.

Data Pre-Processing The student datasets collected from different locations and servers undergo several cleaning operations before knowledge extraction. Dataset cleaning is performed to remove sensitive information from the data tables such as student names, roll numbers, contact numbers, and school names [3, 6]. Further, all textual data are transformed into numerical values to make them suitable for machine learning processing. Student marks are converted into grade categories for efficient prediction analysis.

SPD←Data_Pre_Processing(IRD)----Eq. 1

Where IRD is Input raw data and IPD is input pre-processed data.
Further data is pass for the data privacy. To maintain privacy, direct identifying attributes such as city pin codes are transformed into generalized urban/rural classes using superclass substitution. All categorical activity-based features are converted into binary values, where 1 indicates that the student performs the activity and 0 indicates non-participation.

Certain sensitive attributes such as age, salary, and postal code are anonymized because these values can directly reveal student identity. This anonymization is achieved by converting exact values into predefined ranges. A randomization function is applied to generate values within fixed intervals, and the original values are replaced with these intervals to ensure privacy preservation while maintaining data usability for analysis.

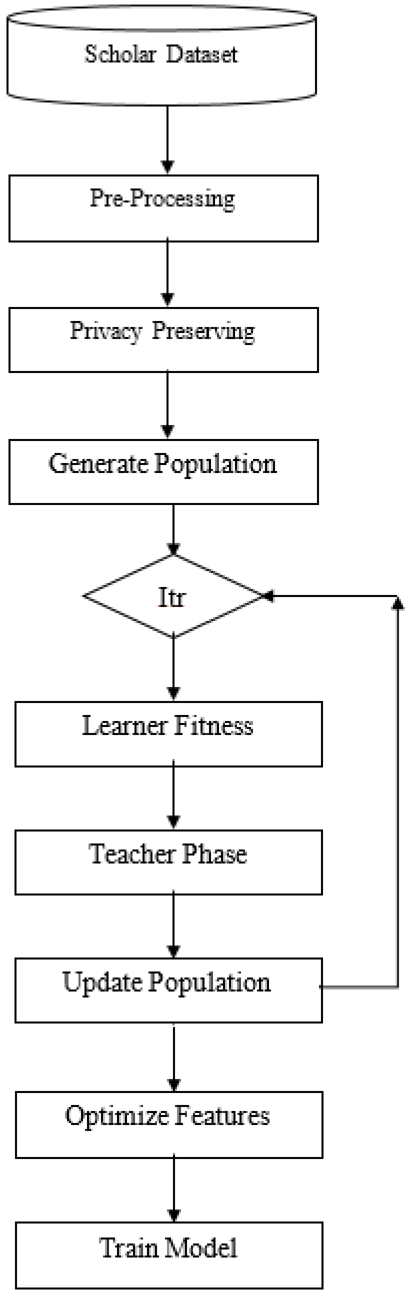

Fig. 1 Block diagram of proposed SGPMTLM model.

Modified Teacher Learning Optimization Algorithm (MTLA): The proposed work employs a Modified Teacher Learning Optimization Algorithm (MTLA) for identifying the most effective features that influence student grade prediction [14]. Teacher Learning Based Optimization (TLBO) is a population-based optimization technique inspired by the teaching and learning process in a classroom environment. It consists of two main phases: the teacher phase and the learner phase.

In the proposed modified version, adaptive feature updating and dynamic selection mechanisms are introduced to improve convergence speed and feature selection accuracy. Each candidate feature subset is treated as a learner, while the best-performing solution in the population is considered as the teacher. The objective of MTLA is to identify the optimal subset of student-related features that maximizes prediction accuracy while minimizing redundant and irrelevant attributes.

Generate Population: In MTLA, each learner represents a possible feature subset. The population consists of multiple learners generated randomly. Each learner is represented as a binary vector where:

1 indicates the feature is selected
0 indicates the feature is not selected

$$L=\{f0, f1, f0, f1, f1 \ldots\ldots fn\}$$

where:

fn denotes available features
N denotes total number of learners
The learner generation process is represented as:

$$PL \leftarrow Generate\_Learner(fn, N)$$

Fitness Function

The fitness value of each learner is evaluated based on student grade prediction accuracy using the selected feature subset. The fitness evaluation procedure is carried out by first filtering the input features according to the learner representation generated by the Modified Teacher Learning Algorithm. The selected feature subset is then used to train the Error Back Propagation Neural Network (EBPNN). After training, the neural network is tested to evaluate its prediction performance on student grade classification. The predicted grades are compared with the actual grades to measure the effectiveness of the selected feature subset. Based on this comparison, classification accuracy is computed for each learner. The learner achieving the highest prediction accuracy is considered the optimal solution and is selected as the best feature subset for student grade prediction.

$$F \leftarrow Fitness(PL, SPD)$$

Teacher Phase: In this phase, the best learner in the population acts as the teacher and attempts to improve the performance of other learners [15]. The teacher updates learner solutions by shifting feature selection toward the optimal feature subset. The feature update equation is represented as:

$$Lnew \leftarrow crossover(Li, Lt)$$

Here Li is ith learner in population and Lt is teacher Learner in current iteration as per fitness value. In crossover operation algorithm modify the learner as per best teacher leaner feature set. Here random set of features from the teacher are replicate into learner.

Learner Phase In the learner phase, learners interact with each other to improve their solutions. Two learners are selected randomly and their fitness values are compared.
If learner Li performs better than learner Lj, then Lj updates its feature subset. Otherwise, learner Li updates its feature subset. This interaction enhances exploration of the feature search space. In this model modification is that not all set of learners participate in the Learner phase, low fitness value learners are not further treat in the iteration, this directly reduces the cost of the model.

Population Updates
After both teacher and learner phases, updated learners are evaluated. If the updated learner achieves higher fitness, it replaces the previous learner. Otherwise, the original learner is retained. The process continues until the maximum number of iterations is reached. The final best learner provides the optimized feature subset for grade prediction.

Error Back Propagation Neural Network
The optimized feature subset obtained from MTLA is provided as input to the Error Back Propagation Neural Network for student grade prediction [16]. The neural network is trained using the selected student attributes to classify student performance into grade categories.

Proposed SGPMTLM Algorithm:
Input: IRD // Input Raw Dataset
Output: SGPM // Student Grade Prediction Model

1. SPD←Data_Pre_Processing(IRD)
2. SPD←Privacy_Preserving(SPD)
3. PL←Generate_Learner(fn, N)
4. Loop i to itr
5. F←Fitness(PL, SPD)
6. Loop 1 to i
7. PL←crossover(Li,Lt)
8. EndLoop
9. PL←Learner_Phase(PL)
10. EndLoop
11. Lt←Fitness(PL, SPD)
12. TFS←Filter_Feature(SPD,Lt) // TFS: Train Feature Set
13. SGPM ←Train(SGPM,TFS)

The proposed SGPMTLM model takes the Input Raw Dataset (IRD) and performs data preprocessing followed by privacy-preserving operations to prepare secure student data for analysis. An initial learner population is then generated, and feature optimization is performed iteratively using the Modified Teacher Learning Algorithm through fitness evaluation, crossover, and learner phase operations. After completing all iterations, the best learner is selected based on maximum prediction accuracy. The optimized feature subset is extracted to form the training feature set, which is finally used to train the Student Grade Prediction Model (SGPM) for accurate grade prediction.

V. Experiment and Results

The implementation of the proposed model was carried out using MATLAB 2016a software. The experimental analysis was performed on a system equipped with an Intel Core i3 6th generation processor and 4 GB RAM. Comparison was done with existing model KPPDERT proposed in [17].

Evaluation Parameters

In order to compared proposed model with other existing models following set of parameters were evaluated.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F\_Measure = \frac{2 * Prcision * Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Results

Table 1 Student grade detection models precision value.

Dataset Percent	I Grade		II Grade	
	KPPDERT	SGPMTLM	KPPDERT	SGPMTLM
20	0.3525	0.4904	0.2155	0.385
40	0.5912	0.4833	0.3091	0.38
60	0.7879	0.478	0.4204	0.4083
80	0.2391	0.4	0.5497	0.499

Table 1 presents the precision values obtained by the existing KPPDERT model and the proposed SGPMTLM model for student grade detection under different dataset percentages for both Grade I and Grade II classification. For Grade I, the proposed SGPMTLM achieved a higher precision value of 0.4904 at 20% dataset compared to 0.3525 obtained by KPPDERT. The results indicate that the proposed SGPMTLM model provides competitive and stable precision performance for student grade prediction across varying dataset sizes.

Table 2 Student grade detection models recall value.

Dataset Percent	I Grade		II Grade	
	KPPDERT	SGPMTLM	KPPDERT	SGPMTLM
20	0.2997	0.3879	0.1773	0.2175
40	0.2733	0.3816	0.191	0.2159
60	0.3708	0.4134	0.255	0.2248
80	0.0236	0.3561	0.3067	0.4286

Table 2 shows that proposed SGPMTLM has increases the recall values in all set of testing dataset. It was shown in table that use of teacher learning model has increases the work efficiency. Table 2 shows that recall value average was increase by 27.7% as compared to KPPERT model proposed in [17].s

Table 3 Student grade detection models f-measure value.

Dataset Percent	I Grade		II Grade	
	KPPDERT	SGPMTLM	KPPDERT	SGPMTLM
20	0.0188	0.4332	0.2997	0.278
40	0.3738	0.4265	0.2361	0.2754
60	0.5042	0.4433	0.3423	0.2899
80	0.043	0.3771	0.4353	0.4612

Table 3 shows the f-measure value of the SGPMTLM model where f-measure was improved by 24.5% as compared to KPPDERT model. It was shown that use of optimized features by teacher algorithm for error back training has increases the model f-measure value.

V. Conclusion

This paper has proposed a model that predicts the student grades based on the demographic and other academic features. In order to improve the prediction accuracy learning of the model was done by the optimized features from the modified teacher learning model. It was found that optimized features train the learning model efficiently. Experiment was done on real educational dataset and result shows that proposed model has improve the prediction precision by 1.66% while recall was enhanced by 27.7%. In future researcher can expand the feature set from society and other factor to further enhance the prediction accuracy.

References

1. AKukkar, R. Mohana, A. Sharma, and A. Nayyar, "A novel methodology usingRNN + LSTM + ML for predicting student's academic performance," Educ. Inf.Technol., Jan. 2024,
2. M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," Smart Learn. Environ., vol. 9, no. 1, p.11, Dec. 2022
3. Asselman, M. Khaldi, and S. Aammou, "Enhancing the prediction ofstudent performance based on the machine learning XGBoost algorithm,"Interact. Learn. Environ., vol. 31, no. 6, pp. 3360–3379, Aug. 2023,
4. IK. Nti, S. Akyeramfo-Sam, B. Bediako-Kyeremeh, and S. Agyemang, "Prediction of social media effects on students' academic performance using MachineLearning Algorithms (MLAs)," J. Comput. Educ., vol. 9, no. 2, pp. 195–223, Jun.2022.
5. S. Hussain and M. Q. Khan, "Student-Performulator: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning," Ann. Data Sci., vol. 10, no. 3, pp. 637–655, Jun. 2023.
6. K. M. Hasib, F. Rahman, R. Hasnat, and M. G. R. Alam, "A machine learning andexplainable ai approach for predicting secondary school student performance,"

in2022 IEEE 12th Annual Computing and Communication Workshop and Conference(CCWC), IEEE, 2022, pp. 0399–0405.

7. O.A. Olabanjo, A.S. Wusu, M. Manuel A machine learning prediction of academic performance of secondary school students using radial basis function neural network Trends Neurosci. Educ. (2022), Article 100190 Sep 23.
8. Abid, K., Aslam, N., Fuzail, M., Maqbool, M.S., Sajid, K.: An efficient deep learning approach for prediction of student performance using neural network. VFAST Trans. Softw. Eng. 11(4), 67–79 (2023).
9. Ani A, Khor E T. Development and evaluation of predictive models for predicting students performance in MOOCs[J]. Education and Information Technologies, 2023: 1-24.
10. Alshaikh, F., Hewahi, N.: Convolutional neural network for predicting student academic performance in intelligent tutoring system. Int. J. Comput. Digit. Syst. 15(1), 239–258 (2024).
11. Li, Y., Shao, D., Sun, H.: Optimizing higher education curriculum design and teaching content recommendation system using data mining algorithms. Appl. Math. Nonlinear Sci. 9(1), 2427 (2024).
12. Li Xiao yong, Zhang Yong, Cheng Huimin, Li Mengran, Yin Baocai. Student achievement prediction using deep neural network from multi-source campus data. Complex & Intelligent Systems, 2022, 8(6).
13. Zhai Mingyu, Wang Sutong, Wang Yanzhang, Wang Dujuan. An interpretable prediction method for university student academic crisis warning. Complex & Intelligent Systems, 2021(prepublish).
14. M. M. Puralachetty, V. K. Pamula, L. M. Gondela and V. N. B. Akula, "Teaching-learning-based optimization with two-stage initialization," 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), 2016
15. R. Venkata Rao, Vivek Patel, An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems, Scientia Iranica, Volume 20, Issue 3, 2013.
16. Ashwani Mathur "Hybrid Combination of Error Back Propagation and Genetic Algorithm for Text Document Clustering" International Journal of Computer Trends and Technology 68.11 (2020):64-68.
17. A Aminifar, M. Shokri, F. Rabbi, V. K. I. Pun and Y. Lamo, "Extremely Randomized Trees With Privacy Preservation for Distributed Structured Health Data," in IEEE Access, vol. 10, pp. 6010-6027, 2022.