IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 21 June 2023, accepted 6 July 2023, date of publication 12 July 2023, date of current version 20 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3294700

RESEARCH ARTICLE

Prediction of Student's Performance With Learning Coefficients Using Regression Based Machine Learning Models

PALLAVI ASTHANA[1], (Member, IEEE), SUMITA MISHRA[1], (Senior Member, IEEE), NISHU GUPTA[2], (Senior Member, IEEE), MOHAMMAD DERAWI[2], AND ANIL KUMAR[1]

[1]Amity School of Engineering and Technology, Amity University, Lucknow Campus, Uttar Pradesh 226028, India
[2]Department of Electronic Systems, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Corresponding author: Nishu Gupta (nishu.gupta@ntnu.no)

ABSTRACT Advanced machine learning (ML) methods can predict student's performance with key features based on academic, behavioral, and demographic data. Significant works have predicted the student's performance based on the primary and secondary data sets derived from the student's existing data. These works have accurately predicted student's performance but did not provide the metrics as suggestions for improved performance. This paper proposes the 'Learning Coefficients' evaluated through trajectory-based computerized adaptive assessment. Learning coefficients also provide quantified metrics to the students to focus more on their studies and improve their further performance. Before selecting the learning coefficients as the key features for student's performance prediction, their dependency on other key features is calculated through positive Pearson's coefficient correlation. Further, the paper presents comparative analysis of the performance of regression-based ML models such as decision trees, random forest, support vector regression, linear regression and artificial neural networks on the same dataset. Results show that linear regression obtained the highest accuracy of 97% when compared to other models.

INDEX TERMS Adaptive assessment, learning coefficients, machine learning models, regression based prediction, student's grade prediction.

I. INTRODUCTION

Educational institutes record student data as demographic, academic, and educational. Many programs conducted as part of teaching-learning activities also generate different data like study behaviors, study patterns, and participation in extracurricular activities. This vast amount of data can efficiently predict student's performance and establish a correlation between it's features through machine learning (ML) based algorithms when processed with proper tools [1]. These correlations can predict student's performance at the end of the program. Prediction of student's performance is significant in adapting the learning environment through customized academic assistance, academic guidance, advice mentoring, examining efficiency and effectiveness of learning methods along with meaningful feedback to modify the environment in order to improve the learning in students [2]. For various reasons like dropout rate, retention rate and assignment performance, most of the works have considered the metrics of student's performance as the acquired Cumulative Grade Point Average (CGPA) and course grade range at the end of the engineering program. Cumulative grades constitute a significant factor in determining a student's eligibility in various aspects of progression such as participation in placement activities or applying for higher studies, etc. [3]. Considerable research has been done in the past to predict student's performance due to two reasons; the availability of relevant information at educational institutes in digital format, and the development of ML tools that can model this information into data to make meaningful patterns through data mining and classification techniques [4]. By early identification and proper intervention, teachers and institutions can provide the

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino.

necessary support for a better understanding of courses [5] to improve student's knowledge and academic performance. ML models help discover patterns and relationships between data variables and analyze complex non-linear relationships for decision-making. They are also valuable in predicting student's performance based on various factors including their earlier and in-term performance in each course.

A. CONTRIBUTION

Earlier work relied on static predictors that only predicted the performance without providing the measures of improvement in the performance. This works used 'Learning Coefficients' as the dynamic predictors of student's performance. These are academic predictors and are calculated through adaptive assessment conducted as continuous evaluation during the semester. They bring uniformity in one of the predictors, purely as the evaluation of student's knowledge with respect to the course studied in the current academic semester. This way it provides a true measure of student's current performance and better performance in adaptive assessment suggests that they have a possibility of good performance in their final examination also.

A correlation feature selection approach is applied in this work to determine which features are most important for predicting student's performance. Correlation feature selection scores are presented using which the features were scored using this method. Besides this, the major contributions of this work are:

(1) To identify the ML techniques prominently adopted for accurately predicting student's academic performance.
(2) To understand the feature selection in ML based models.
(3) To propose quantified features as a uniform feature for the measurement of improvement in student's performance based on research gaps.

Apart from these contributions, the article reviews many standard ML based algorithms such as Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbour (KNN), Decision Tree (DT), Random Forests (RF), Multilayer Perception (MP), Linear Regression (LR), bagging and boosting that have provided accurate results while predicting performances [6], [7]. This data was pre-processed with supervised learning methods as shown in Figure 1.

B. ORGANIZATION

Section II presents a background of the ML-based student's performance prediction models, discusses the feature selection process for training and testing of these models and identifies the non-uniformity among these features. Section III discusses the proposed method and introduces 'learning coefficients' as quantitative metrics of student's performance calculated through adaptive assessment. In section IV, learning coefficients are established as one of the key features for prediction which is confirmed by calculating its correlation with other academic and demographic features. Further, this section presents the results and discusses the performance evaluation of various models. Finally, section V concludes the article.

II. BACKGROUND

This section discusses the prominent ML models used for student's performance prediction in recent research works. Table 1 summarizes the details of such works discussing ML-based models, feature selection and the adopted methodology based on the size of the dataset or the primary or secondary data selection. The sets of predictors is not the same in each work, bringing non-uniformity in the prediction results.

A. FEATURES FOR PREDICTION MODELS

Many factors like academic performance, demographic background, study behavior, selection of the courses, and extracurricular activities affect the performance of

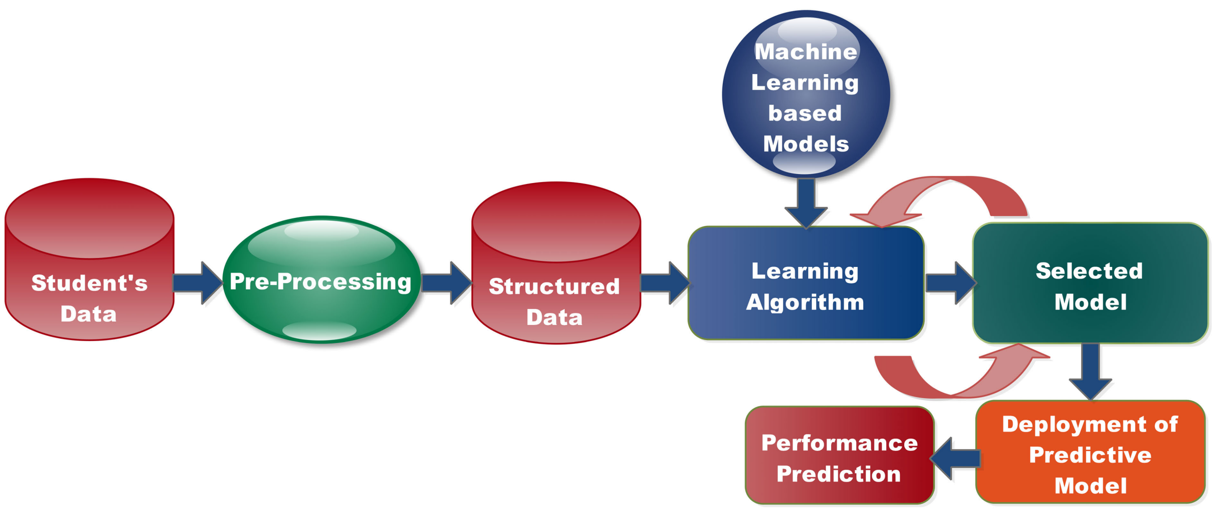

FIGURE 1. Prediction of student's performance based on ML models.

TABLE 1. Prominent ML models used for student's performance prediction.

Model	Selection of features	Size of dataset	Outcome and limitations
J48, REP Tree Hoeffding Tree [8]	Gender, Participation in extracurricular activities, Number of tuitions taken, Programming skills, Class test marks, Attendance, CGPA	850 students	J48 algorithm can recur on smaller subset in one class based on normalized gain of the best features of data. Precise calculation needed for the best splitting attribute
J48 Decision Trees [9]	Final grades in Secondary School Certificate data	1500 students	Highlighted that J48 algorithms are good for prediction in terms of speed, space and accuracy. Non-uniform prediction of grades
Decision Trees ID3, CHAID, and C4.5 [10]	Roll number, School marks, Undergraduate marks, Board, Communication skills, and Placement records	1342 students	Successfully implemented ID3 and CHAID for performance of students. Grades need to be converted into nominal values resulting in unnecessary normalization of features
Matrix factorization Collaborative Filtering and Restricted Boltzmann Machines [11]	Secondary School percentage, High School percentage, Entry test scores, and Interview	13 years of graduand data	Feature groups related to pre-requisite courses showed best performance Hard to map the pre-requisite courses of such a large data
C4.5 Decision trees, Artificial Neural Network Multilayer perception Naïve bayes, and SVM [12]	Family Size, Family income, Marks in Secondary School, and Mid-term	309 students	Emphasized on the prominent role of family background on academic performance. Data features heavily relied on non-academic features
SVM, KNN, Naïve Bayes [13]	Average grades of students	683 students	For the independent features like students grades, the SVM classifier performed better. Classifiers rely on consistency between exmas in different years due to variation in examination difficulty
J48 Algorithm [14]	Final grade pointer, Major courses, Nationality, Campus	236 students.	Decision tree based prediction to assess the effect of individual courses on student performance. Unbiased selection of root node
SVM, Discriminant Analysis [15]	Student's age, Course names, Grades	50 students	Identification of high performing students through prediction. Unavailability of large dataset for new university
L3 classifiers, C4.5 Decision trees, Multilayer Perceptron, Naïve Bayes, KNN, random Forest [16]	High School, Age of students, Course opted, Download of educational	5000 students	Application of association rules found useful association between predictive features, Yet to discover the effectiveness for examination performance prediction
Ensemble methods, Random forest, J48 Decision trees, Naïve Bayes, ANN [17]	Demographic details, Social details, academic details, Study habits	1227 students	Multi -boost technique combines weak learners to create a string of learners. Voting method combines the result of multiple classifiers for early performance prediction System is deficient in suggesting appropriate learning streams based on predicted performance
Exponentially weighted average forecaster (EWAF), Random Forest, Linear regression, KNN [7]	Student's Semester Grade Point Average (SGPA), High School Score, Course credits	1169 students	Validation of pre-requisite dependencies in student performance prediction Unable to recommend useful courses to students based on performance
Deep learning method, Random Forests, SVM, Linear regression [18]	Course grades	2014 to 2015 MOOC student records	Utilisation of Bayesian-based optimization architecture predicted the performance for unbalanced dataset without pre-processing Addition of multi features diminishes the optimizer's performance
Maximum-weight dependence trees (MWDT) [19]	Cognition and meta-cognition, Self regulatory survey questions	82 students	Prediction is based on cognitive and meta-cognitive features Prediction results needs cross-validation by inclusion of academic features
J48, SVM, Random Forest, Artificial Neural Network [20]	Facilities provided at educational Institutes like Library, Infrastructure, Laboratories	72 students	10-cross validation technique provided results on small dataset. Larger dataset will further validate the utility of extracted features
Robust Hybrid ensemble model Naïve Bayes, Multilayer Perceptron, K Nearest Neighbour, Decision trees [21]	Demographic information, Motivational behaviour, Academic information	319 students	Utilization of single classifier based model, ensemble classifier based model and classifier based hybrid ensemble models for prediction Lack of meta-analysis o be regarded as decision support to identify most efficient model
Decision Trees, Linear Support Vector Machine, Random Forest, Gradient Boosting [22]	Academic information: student-specific features, Course Specific features	13 years of under graduand data	Extraction of indicators of students performance before the selection of course. Utilization of single model for the groups with distinct features
K Nearest neighbor, Naïve Bayes, Neural Network, SVM, Random Forest [23]	Student's Grade points, Internal assessment, Moodle activity, Video interaction in Flipped classes	772 students	Determination of impact of feature selection on classification in virtual learning environment. Less accuracy in prediction of poor performing students
SVM, Ensemble filtering [24]	Program structure, Gender, Race, Sponsorship	447 students	Utilization of novel features for performance prediction Using only linear function may improve the accuracy

engineering students. When any model is developed based on this data, it provides ideal conditions which may not exist for many students as they have different kinds of background and motivational levels. Table 1 shows that most of the work focus on implementing and validating the effectiveness of the predictive algorithms. Researchers have selected different features in their work for inclusiveness and to increase the range of features. Most of the features are static in nature and suffer from many limitations such as:

(i) Academic data: the most reliable data to predict student's academic performance. It consists of student's CGPA, internal assessment, internal examination scores and courses selected by the students. Major problem with this kind of data is non-uniformity. Different institutes have their own assessment methods and any model that has shown accuracy on a dataset of any institute may not provide the same results for another institute. Moreover, the applied methods may differ for different courses. For example, with the secondary school data, few institutes have a higher cut-off for admission. Hence, there will be less variation in this data, whereas, in the institutes with a lower cut-off, this variation will be high. Therefore, the reliability of the same features would be different in both the settings [25], [26].

(ii) Demographic and personality data: such details include the age, gender and personality data about student's background like parent's education, parent's income, emotional intelligence, student's interests, level of motivation, communication, interest in sports, hobbies and ethnicity. This data is vital because student's motivational level heavily rely on these factors. However, accurate data is often missing, and analysis with such kind of data can lead to biases against students with a specific background. Also, this kind of data would require permission concerning ethics [27].

(iii) Institutional data: it pertains to the facilities dependent on the institutes, the condition of the laboratories, the state of experiments, infrastructure, teaching methodology, transportation facility and communication medium [26], [27]. This data is contextual and dependent on factors such as the availability of proper resources for selected courses that would affect the accuracy of the predictive models.

(iv) Behavioral data: it includes the study pattern, attention span, rate of downloading of study material in case of flipped classes, social interactions, time spent on social sites and playing educational computer games. Feature-related behavior of students is crucial for the analysis of their performance prediction. Most of this data is subjective, based on surveys and questionnaires. Researchers use this as primary data or extract secondary information from this data for prediction purposes. This data relies heavily on the student's responses which can have many diversification and needs to precisely map various features like the demographic and academic profile of the students [26], [27].

B. GAPS IN THE PRESENT RESEARCH

The main features utilized for the student's performance prediction are demographic details, institutional details and educational details which are static in nature. Students have different backgrounds, study habits and variable academic benchmarks. These factors can predict the student performance but as these features are not sufficient to record the student's learning during the semester, there is a need to devise an effective tool to predict students' performance before the final assessment that may improve learning outcomes. Therefore, there is a need to devise an effective tool to predict and improve students' performance. In this work, we aim to predict their performance during the semester to provide with the opportunities of improvement [26]. This can be done with the inclusion of a dynamic feature along with other features to ensure the measured improvement.

III. PROPOSED METHOD

To strengthen the predictive models, 'learning coefficients' calculated through adaptive assessment [28] are introduced in this work. They are the quantified performance metrics for a group of courses including a course and its pre-requisite courses. It has been established that performance in the pre-requisite course is a significant predictor of performance in the successive courses [18]. This academic feature depends on the knowledge-building of the students thereby bringing uniformity in the dataset. In addition, adaptive assessment can be conducted multiple times during a semester, thus providing more opportunities for improvement. To assess the performance of learning coefficients based student's performance prediction model, a study was conducted on the students of undergraduate Computer Science and Engineering (CSE) program at Amity University, Lucknow, India. The conceptual framework of the proposed methodology is shown in figure 2. The university follows Choice Based Credit System (CBCS). All the students need to complete a minimum number of credit units in each semester. The program structure defines the courses in each semester under four different categories namely, Core Courses (CC), Domain Elective (DE), Value Added Courses (VAC) and Open Elective (OE). CC and DE include courses that are related to the branch a student is enrolled in, while OE and VAC allow the students to study topics of interest that are beyond their core domain. DE courses are taught in collaboration with other engineering departments. CSE program has domain elective course on Microprocessors (MP) in the seventh semester which is offered by electronics and communication engineering department. This experiment is conducted on 91 students who opted MP as DE course. Two prerequisite courses for the MP course as mentioned in the program structure are (a) Computer Organization and Architecture (COA) and (b) Basics of Electronics Engineering (BEE). Students study these prerequisite courses in the third and fourth semester respectively.

A. TRAJECTORY BASED ADAPTIVE ASSESSMENT

Adaptive assessments are tailored specifically to each student by providing a customized set of questions where the difficulty level of every succeeding question depends on

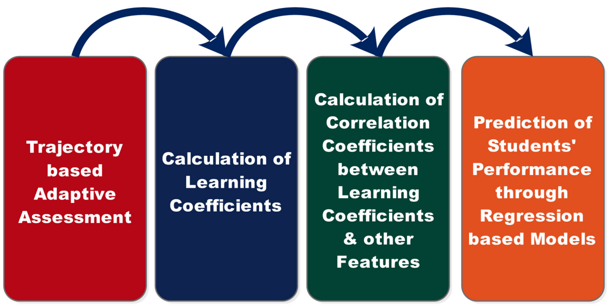

FIGURE 2. Adopted framework.

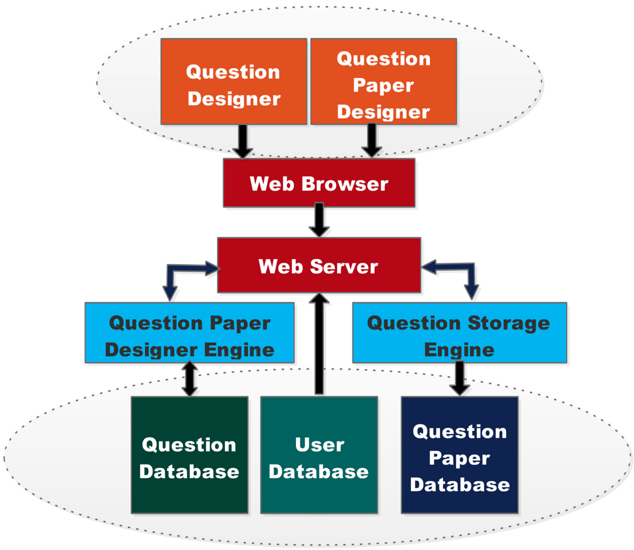

FIGURE 3. Designing of a computerized assessment.

the student's performance in the preceding question. The trajectory-based adaptive assessment works in this pattern.

Automated method of adaptive question generation is still in the design phase but its basic building blocks are shown in figure 3. This system will require the blended learning for the implementation, where questions will be uploaded on web server and accessed through Web browser.

B. PARTICIPANTS AND DATABASE

This study was conducted on 91 undergraduate students of CSE department. Their records were grades in Microprocessor course, their SGPA for seven semesters, cumulative grade pointers at the course completion, and board of secondary education. The student's names and enrolment numbers were removed to maintain the privacy of data and for the analysis within the university's data privacy obligations.

C. CALCULATION OF LEARNING COEFFICIENTS AS EVALUATION METRIC

In order to calculate the learning coefficients, we present an algorithm. The question paper consists of 10 sets of multiple choice-based options, each set of three questions in the order of appearance: BEE → COA→ MP. A correct answer will award score 1, otherwise 0. The assessment is formative. The process of assessment of scores is shown in figure 4.

(i) The first question in the set is from the course BEE; if a student answers it correctly, the score is 1. If it is wrong, the score is 0. This score is '*a*'. Cumulative a's are '*J*'.
(ii) If a=1, the next question will appear from the course COA; if a=0, the next question is from BEE again. For a correct answer, the score is 1 otherwise 0. This score is '*b*'. Cumulative b's are *K*.
(iii) If b=1, the next question will appear from the course MP; if b=0, the next question is from COA again. For a correct answer, the score is 1, otherwise 0. This score is *c* Cumulative c's are *L*.
(iv) Either c=0 or c=1, the next question will appear from BEE for second set of questions.
(v) This process will end when L adds ten scores. Learning coefficients J, K, and L are the average values of a_i, b_i and c_i respectively as shown in the equations (1), (2) and (3) respectively.

$$J = (a_1 + a_2 + \ldots\ldots.a_{10})/10 \quad (1)$$
$$K = (b_1 + b_2 + \ldots\ldots b_{10})/10 \quad (2)$$
$$L = (c_1 + c_2 + \ldots\ldots.c_{10})/10 \quad (3)$$

where $a_i \rightarrow$ Score in the questions of course BEE $b_i \rightarrow$ Score in the questions of course COA $c_i \rightarrow$ Score in the questions of course MP. Scores are calculated for all the students to compare the performance in each course in the trajectory. Evaluated learning coefficients are shown in figure 5. Average values of the learning coefficients for 91 students are calculated. Average value of J is 0.94, K is 3.3 and L is 6.2. These are the metrics for the measurement of the academic knowledge when taught by trajectory based teaching pedagogy, so these are purely academic features for the prediction.

IV. RESULTS AND DISCUSSIONS

Data analysis was conducted in two phases (i) Establishment of 'learning coefficients' as key features for prediction by finding the correlation with academic and demographic features and (ii) Prediction of student's CGPA based on the academic features and learning coefficients.

Experiments ran on open development called collaboratory (colab) developed by Google for analysis, and supports full Python syntax. It is an efficient method to run ML-based algorithms. Colab provides the facility to plot graphs that facilitate the graphical representation and is easier to comprehend. Results are saved on Google drive directly. Data Pre-processing: Microsoft excel file is saved in .csv format for processing colab. Critical stages of data pre-processing are: Dataset Cleaning: removing features such as student's names, enrolment number, course descriptions, batch, institution and date of birth. University maintains the records of all students on an online portal. Hence, no data was missing and all the records were complete.

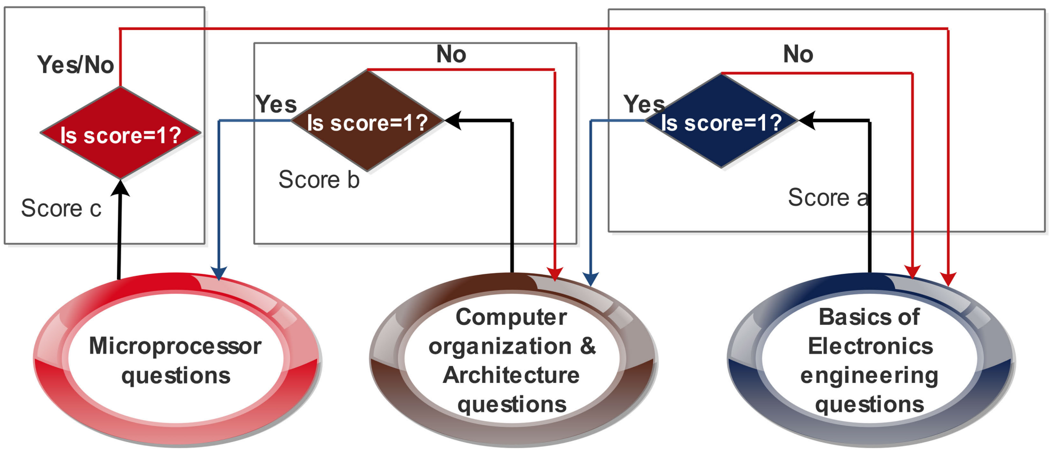

FIGURE 4. Calculation of scores in adaptive assessment.

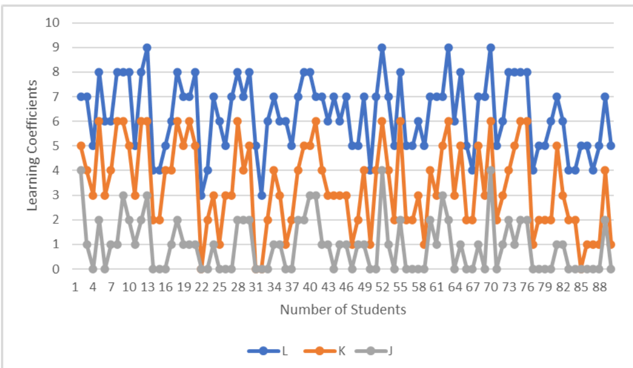

FIGURE 5. Values of 'learning coefficients' J, K and L.

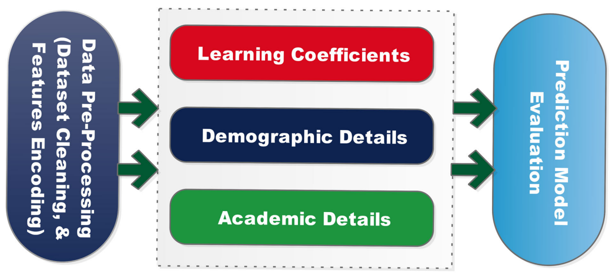

FIGURE 6. Methodology for the evaluation of the prediction model.

A. FEATURE ENCODING

All the data types are converted to numeric data type because the algorithm used for this work supports numeric data type as shown in Table 2. Grade pointers and scores were not rounded-off to integers but taken as original values for the prediction with values up to two places after the decimal. Few features such as gender and board of examination (high school, intermediate) were converted to numeric values making it convenient to run on Python.

B. FEATURES SELECTION AND DATA PRE-PROCESSING

Before the evaluation phase, pre-processing included removing extra columns from dataset and converting them into numerical values to make it usable for the analysis. Figure 6 shows the methodology considered for the evaluation of the prediction model.

C. CORRELATION BETWEEN LEARNING COEFFICIENTS AND OTHER KEY FEATURES

Learning coefficients are correlated with academic features and other demographic features for establishing them as predictor of performance. Academic data consists of student's pointers in class 10th and 12th, SGPA of semester I to VII, marks obtained in the end semester exam of the Microprocessor course, and gender as a demographic feature. The dependency between the learning coefficients and response variable was calculated through a correlation matrix as shown in figure 7. Correlation coefficients are calculated through Scikit, a software machine learning library used for Python programming language. The colab tool was used for the evaluation of the work that supports Scikit which further calculates the correlation coefficient by Pearson's correlation coefficient (r). The correlation coefficient is calculated as shown in equation 4:

$$r = \frac{\Sigma(x_i - x)(y_i - y)}{\sqrt{\Sigma(x_i - x)^2 \Sigma(y_i - y)^2}} \tag{4}$$

Correlation matrix shows the correlation between learning coefficients and other key features. Positive correlation coefficient value (r) shows the dependency among features and a negative correlation value indicates independence in features. The correlation coefficient was calculated for learning coefficients and the remaining features. Calculated values of 'r' are favorable for the grade pointers in the course Microprocessor ranging between .70 to .78; CGPA of all the seven semesters between .23 to .82; CGPA achieved in final semester as .73 to .84; percentage of 10th as .31 and .36 in the 12th examination. Positive correlation coefficient values

TABLE 2. Selection of key attributes.

Attributes	Data Types	Nomenclature	Attribute Details
Student's Id	Ordinal	1, 2, 3,................91	Roll number of students
Semester marks in specified course (Microprocessor)	Ordinal	1	Marks obtained by the student in specified course (Floating point number)
SGPA	Ordinal	2, 3, 4, 5, 6, 7, 8	Weighted average of all the subjects undertaken in a semester. Floating number up to the seventh semester
CGPA	Ordinal	CGPA	Floating number final cumulative grades
High School grades	Ordinal	10^{th}	Floating number
Intermediate grades	Ordinal	12^{th}	Floating number
Board of the school	Nominal	10^{th} board and 12^{th} board	0 for Central Board of Secondary Education (CBSE) board, 1 for Indian Certificate of Secondary Education (ICSE) board, and 2 for state boards
Gender	Nominal	Female and Male	1 for females, and 0 for Male
Learning Coefficient prerequisite 1	Ordinal	J	Mean score of the course 'Basics of Electronics Engineering' in 'Trajectory based assessment'
Learning Coefficient prerequisite 2	Ordinal	K	Mean score of the course 'Computer Organisation' in 'Trajectory based assessment'
Learning Coefficient for the specified course (Microprocessor)	Ordinal	L	Mean score of the course 'Microprocessor' in 'Trajectory based assessment'

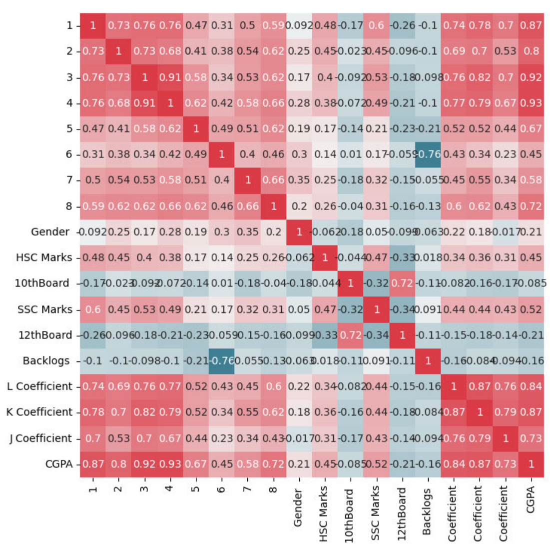

FIGURE 7. Correlation matrix.

prove that learning coefficients are also significant factors in predicting the student's performance.

D. EVALUATION OF THE MACHINE LEARNING MODELS FOR PREDICTION

The dataset was split into training and testing data in 90% and 10% respectively of the total inputs. As the dataset is small, regression-based ML models are most suitable to run the prediction. Deep Learning based models are prone to over-fitting on a small dataset. After pre-processing, the dataset is run for the following ML models: LR, DT, RF, and SVR on the colab tool which are briefly discussed below.

1) *Linear Regression*
LR is used to quantify the linear relationship between an explanatory variable and response variable. In the event, if there are more than one explanatory variables, then it is called *Multiple Linear Regression*. This kind of relationship predicts the response variable or dependent variable, and the variables for predicting the variable are called the explanatory or independent variables [29], [30]. The difference between the true and predicted values is known as the residual. Predicting a response variable based on the explanatory variable is known as regression analysis, and a straight line that describes the prediction is known as the regression line. The calculation of the best fit line is based on the method of least squares which minimises the sum of the vertical distance between all of the data points and the line of best fit. Mathematical representation of multiple linear regression is given in equation 5.

$$y_i = \alpha_0 + \alpha_1 x_i^1 + \alpha_2 x_i^2 + \ldots .\alpha_m x_i^m + \epsilon_i \quad (5)$$

where x_i denotes explanatory variable, y_i is the response variable, $\alpha_0, \alpha_1 \ldots .\alpha_n$ are model parameters wherein α_0 is also known as the intercept and ϵ_i is the residual. The subscript 'i' refers to the $i^{þ}$ data instances in the dataset and 'm' refers to the number of explanatory variables.

2) *Decision Tree*
DTs classify data into good performing and poor performing based on the relationship between the features and their comparative significance. DT regression trains a model in the form of a tree to predict data in the future [31]. They are simple algorithms for prediction structured as root to a leaf node in a tree. Node represents the test of features indicating their possible value

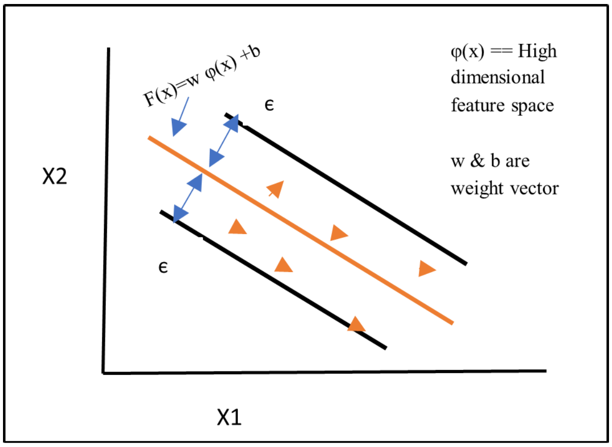

FIGURE 8. Support Vector Regression: hyperplane line with boundary epsilon.

and creates the split based on the calculated value of the feature of the decision. This split creates new nodes and the process continues until the formation of pure leaf nodes.

3) *Random Forests*
Random forests have many DTs that take different features as a root and then based on it Entropy and GINI Index are calculated. Each tree may have a different decision. An algorithm based on RF selects the decision chosen by most DTs based on the majority voting system [18]. To apply RF algorithm,
   - M number of data instances are chosen from the dataset.
   - A DT associated to these M data instances is then built.
   - Both the steps listed above are repeated until N number of DTs are generated.

For a new input data point, each tree generates the prediction value of response variable and assigns that data point to the average across all the predicted values.

4) *Support Vector Regression (SVR)*
SVR is used to model non-linear relationships between variables to adjust the model's robustness through estimated hyperplane functions. A hyperplane is the best fit line with maximum points fitted within a threshold value. They are the decision boundaries to predict the continuous output based on a set of mathematical functions known as kernels. Popular kernels used in SVR are linear, non-linear, polynomial, radial basis function (RBF) and sigmoid [32]. Figure 8 shows the hyperplane between two variables X_1 X_2. ϵ is a tunable parameter that determines the width of the plane around the hyperplane. Points that fall inside this plane are correct predictions. SVR is reasonable for small datasets as it has good generalization capability.

It is not possible for a regression-based model to predict the exact value of a continuous variable. A regression

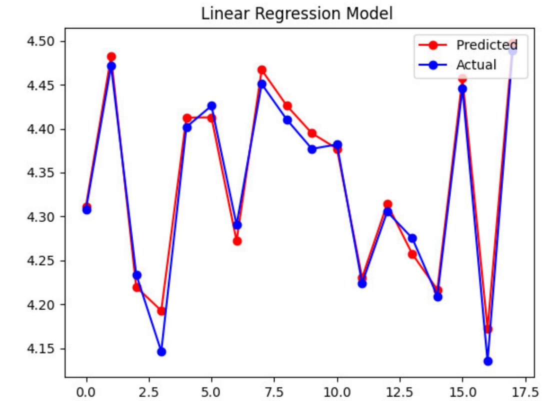

FIGURE 9. Comparison between the true values and predicted values in LR.

model predicts values that are either lower or higher than the actual value. Therefore, we determine the model's accuracy through evaluation metrics based on residuals. Residuals are the difference between the actual and the predicted values. Important parameters for the models such as accuracy RMSE, MAE, and MSE [33], [34] have been evaluated. Evaluation metric used to measure accuracy of best performing model is R^2 score. It is a simple but efficient metric to compare the developed models. R^2 score is calculated using the formula as in equation 6.

$$R^2 = 1 - \frac{R_{SS}}{T_{SS}} \tag{6}$$

where R^2 is the coefficient of determination, R_{SS} is the sum of the squares of the residuals and T_{SS} is the total sum of the squares. Here, $R_{SS} = \sum \left(y_i - \hat{y}_i\right)^2$ where y_i is the actual value and $\hat{y}_i$ is the predicted value; $T_{SS} = \sum (y_i - \bar{y})^2$ where y_i is the actual value and $\bar{y}$ is the mean value of the variable/feature.

In order to determine various parameters, ML models were developed using the default value settings of most of the hyper-parameters defined during the experiment and are presented in Table 3.

TABLE 3. Parameter settings for ML algorithms.

ML Algorithm	Parameters	Value
	criterion	"friedman_mse"
RF	random_state	0
	estimators	32
DT	splitter max_depth	'Best'
		5
	random_state	0
SVR	Degree	8
	Kernel	Poly
LR	Intercept	True

To assess the performance of ML models, apart from R^2 Score, MAE (Mean Absolute Error), MSE (Mean Square Error), and RSME (Root Mean Square Error) are also evaluated.

RMSE, which is the standard deviation of the differences between the predicted values p_i' and observed (true) values p_i', is calculated in equation 7.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(p_i' - p_i)^2}{n}} \tag{7}$$

MAE is the magnitude of the difference between the prediction of observation and the actual value of observation. MAE is the average of the errors in the entire group, used as a loss function for regression-based ML algorithms to provide quantifiable measure of the errors as shown in equation 8.

$$MAE = \frac{\sum_{i=1}^{n}\left|p_i' - p_i\right|}{n} \tag{8}$$

MSE is the loss function calculated as squared errors over the entire group as shown in equation 9. The main advantage of calculating MSE is to ensure that there are no outliers in the model.

$$MSE = \frac{\sum_{i=1}^{n}\left|p_i' - p_i\right|^2}{n} \tag{9}$$

Prediction models based on LR, RF, DT and SVR were trained and tested on the Scikit tool that supports Python. DT and RF are also regression models of ML with a similar loss function as MSE and MAE.

TABLE 4. Calculated values of Accuracy RMSE, MAE and MSE.

Machine Learning Model	Accuracy	RMSE	MAE	MSE
LR	.97	.0149	.0183	.0003
RF	.87	.0356	.0264	.0013
DT	.75	.0539	.0332	.0029
SVR	.74	.0836	.0754	.0069

Table 4 shows the calculated evaluation metrics from different models, and it shows that LR outperforms the other models for the test data. It is expected as complex models perform well on training data but due to over-fitting they, perform poorly over the test data. The table also shows the calculated values of accuracy of RMSE, MAE, and MSE for the above LR, RF, DT and SVR algorithms. LR provides the maximum accuracy of 97%, values of the residuals are least, and hence the values of errors are also small as .0149 .0183, and .0003 for RMSE, MAE and MSE respectively. As we can see from the formula, MAE measures the average magnitude of the errors without considering their direction. It's the average of the absolute differences between predictions and actual values over the test sample. Apart from determining how close the prediction is to the actual value on an average, RMSE also indicates the effect of large errors. Scikit-learn evaluation metric library used in this work does not have RMSE function. So, to get the RMSE, we use the Numpy square root method to find the square root of mean squared error. RMSE function values are examined to determine if there are any large errors in the model developed. We can see from table 4 that the RMSE value is larger than the MAE, particularly in the DT model. This is a result of some large

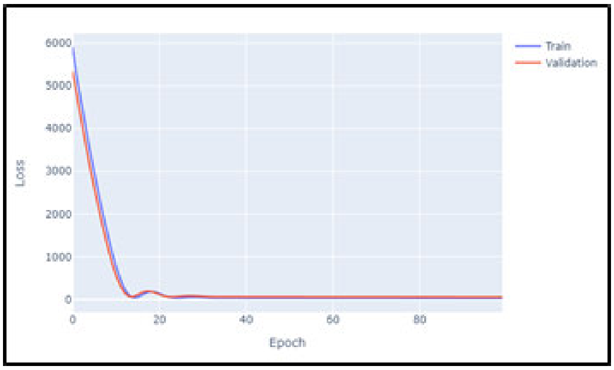

FIGURE 10. Training loss for neural network.

errors due to over-fitting. maximum tree depth for DT model was initially kept at 3 then it was increased to 5.

Figure 9 shows the comparison between true values and predicted values plotted o test data. Predicted values are not deviated much from the true values, thus giving better accuracy.

Apart from the models shown in Table 4, the given dataset also ran on the Artificial Neural Network, which is also a regression-based model. For this work, an ANN was trained with four layers one input layer, two hidden layers, and one output layer. The mean square error is the loss function with the optimizer 'adam' and 'ReLU' activation[36]. The model ran for 100 epochs and provided an accuracy of 79% with a mean absolute error of .05, as shown in figure 11.

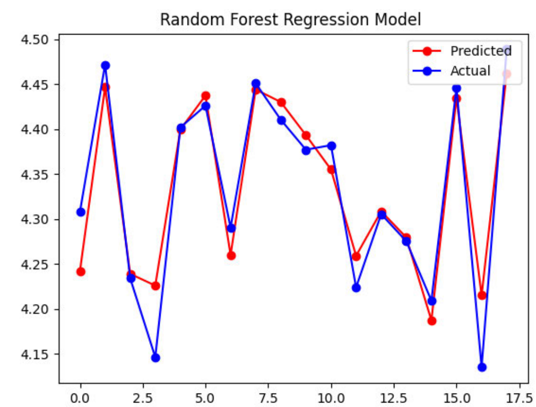

FIGURE 11. Comparison between true values and predicted values in RF.

These parameters show that along with the traditional primary or secondary feature, the proposed learning coefficient is also an effective feature that can predict the student's performance. Along with prediction, it is also a quantified measure that suggests improvement metrics. Existing works focus on predicting the performance of students, without providing evidential measures of improvement [35]. One of the significant limitations of our work is small dataset having only 91 records. Research with more data may result in better feature engineering and noteworthy insights. Proposed method provides a valuable tool for course instructors to modify teaching practices for imparting quality education. for instance, there may be provision of additional teaching support to poor performing students at early stage.

V. CONCLUSION

The work proposed in this article utilized regression based ML models for student's performance prediction. Out of the various models that have been evaluated, 97% prediction accuracy has been achieved by linear regression model. Data sets consist of the academic data including student's CGPA for seventh semester, grades pointers in class 10^{th} and 12^{th} examinations, and gender of the students. We employed a novel set of variables namely learning coefficients for determining the student performance, which have not been used in prior works. The learning coefficients were calculated for the course of Microprocessor taught to the seventh-semester students, calculated in trajectory based adaptive assessment of three inter-related courses. Traditional key features are based on the student's academic data scored at the end-term examination. However, the learning coefficients are evaluated during the continuous evaluation conducted throughout the semester. Hence, their inclusion as the key features provides a quantitative metric to the students so that they can focus more on their studies and secure better grades in the end semester examination. Whereas larger dataset would certainly be a better choice, future extension of this work will include a bigger sample size to obtain the improved overall performance of the proposed ML models. This work is still in the research stage and proposed 'Learning Coefficients' are the innovative evaluation metrices of student's performance. Purpose of this work is to corelate the learning coefficient with other academic predictors and use it to predict student's performance. This work was conducted in only one course that limited the data. For the less data, simpler ML based model such as Linear regression has provided accurate results. But as, the learning coefficient will be calculated for more courses and programs, it will enlarge the dataset, and more complex Machine learning based models will be required for the prediction.

REFERENCES

[1] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, p. 552, Sep. 2021.

[2] C. Romero and S. Ventura, "Guest editorial: Special issue on early prediction and supporting of learning performance," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 145–147, Apr. 2019.

[3] A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," *Educ. Inf. Technol.*, vol. 26, no. 1, pp. 205–240, Jan. 2021.

[4] M. Al-kmali, H. Mugahed, W. Boulila, M. Al-Sarem, and A. Abuhamdah, "A machine-learning based approach to support academic decision-making at higher educational institutions," in *Proc. Int. Symp. Netw., Comput. Commun. (ISNCC)*, Oct. 2020, pp. 1–5.

[5] F. Ofori, E. Maina, and R. Gitonga, "Using machine learning algorithms to predict students' performance and improve learning outcome: A literature based review," *J. Inf. Technol.*, vol. 4, no. 1, pp. 33–55, 2020.

[6] B. Sekeroglu, K. Dimililer, and K. Tuncal, "Student performance prediction and classification using machine learning algorithms," in *Proc. 8th Int. Conf. Educ. Inf. Technol.*, Mar. 2019, pp. 7–11.

[7] J. Xu, K. H. Moon, and M. van der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 5, pp. 742–753, Aug. 2017.

[8] M. I. Hoque, A. K. Azad, M. A. H. Tuhin, and Z. U. Salehin, "University students result analysis and prediction system by decision tree algorithm," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 3, pp. 115–122, 2020.

[9] B. Khan, M. S. H. Khiyal, and M. Daud Khattak, "Final grade prediction of secondary school student using decision tree," *Int. J. Comput. Appl.*, vol. 115, no. 21, pp. 32–36, Apr. 2015.

[10] T. Jeevalatha, N. A. N. Ananthi, and D. S. Kumar, "Performance analysis of undergraduate students placement selection using decision tree algorithms," *Int. J. Comput. Appl.*, vol. 108, no. 15, pp. 27–31, Dec. 2014.

[11] A. Polyzou and G. Karypis, "Feature extraction for classifying students based on their academic performance," in *Proc. 11th Int. Conf. Educ. Data Mining*, Buffalo, NY, USA, Jul. 2018, pp. 356–362.

[12] A. Acharya and D. Sinha, "Early prediction of students performance using machine learning techniques," *Int. J. Comput. Appl.*, vol. 107, no. 1, pp. 37–43, Dec. 2014.

[13] T. Anderson and R. Anderson, "Applications of machine learning to student grade prediction in quantitative business courses," *Glob. J. Bus. Pedagog.*, vol. 1, no. 3, pp. 13–22, 2017.

[14] M. A. Al-Barrak and M. Al-Razgan, "Predicting students final GPA using decision trees: A case study," *Int. J. Inf. Educ. Technol.*, vol. 6, no. 7, pp. 528–533, 2016.

[15] L. M. A. Zohair, "Prediction of student's performance by modelling small dataset size," *Int. J. Educ. Technol. Higher Educ.*, vol. 16, no. 1, pp. 1–18, Dec. 2019.

[16] L. Cagliero, L. Canale, L. Farinetti, E. Baralis, and E. Venuto, "Predicting student academic performance by means of associative classification," *Appl. Sci.*, vol. 11, no. 4, p. 1420, Feb. 2021.

[17] A. Siddique, A. Jan, F. Majeed, A. I. Qahmash, N. N. Quadri, and M. O. A. Wahab, "Predicting academic performance using an efficient model based on fusion of classifiers," *Appl. Sci.*, vol. 11, no. 24, p. 11845, Dec. 2021.

[18] W. Wunnasri, P. Musikawan, and C. So-In, "A two-phase ensemble-based method for predicting learners' grade in MOOCs," *Appl. Sci.*, vol. 13, no. 3, p. 1492, Jan. 2023.

[19] A. Zollanvari, R. C. Kizilirmak, Y. H. Kho, and D. HernáNdez-Torrano, "Predicting students' GPA and developing intervention strategies based on self-regulatory learning behaviors," *IEEE Access*, vol. 5, pp. 23792–23802, 2017.

[20] T. M. Alam, M. Mushtaq, K. Shaukat, I. A. Hameed, M. U. Sarwar, and S. Luo, "A novel method for performance measurement of public educational institutions using machine learning models," *Appl. Sci.*, vol. 11, no. 19, p. 9296, Oct. 2021.

[21] S. Sakri and A. S. Alluhaidan, "RHEM: A robust hybrid ensemble model for students' performance assessment on cloud computing course," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 388–396, 2020.

[22] A. Polyzou and G. Karypis, "Feature extraction for next-term prediction of poor student performance," *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 237–248, Apr. 2019.

[23] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques," *Appl. Sci.*, vol. 10, no. 11, p. 3894, Jun. 2020.

[24] Y. Baashar, Y. Hamed, G. Alkawsi, L. F. Capretz, H. Alhussian, A. Alwadain, and R. Al-amri, "Evaluation of postgraduate academic performance using artificial intelligence models," *Alexandria Eng. J.*, vol. 61, no. 12, pp. 9867–9878, Dec. 2022.

[25] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao, "Predicting academic performance: A systematic literature review," in *Proc. Companion 23rd Annu. ACM Conf. Innov. Technol. Comput. Sci. Educ.*, Jul. 2018, pp. 175–199.

[26] M. Bilal, M. Omar, W. Anwar, R. H. Bokhari, and G. S. Choi, "The role of demographic and academic features in a student performance prediction," *Sci. Rep.*, vol. 12, no. 1, p. 12508, Jul. 2022.

[27] F. Qiu, G. Zhang, X. Sheng, L. Jiang, L. Zhu, Q. Xiang, B. Jiang, and P.-K. Chen, "Predicting students' performance in e-learning using learning process and behaviour data," *Sci. Rep.*, vol. 12, no. 1, p. 453, Jan. 2022.

[28] P. Asthana, S. Tanwar, A. Kumar, and S. Mishra, "Students' assessment for quantitative measurement of course learning outcomes in online class of power plant instrumentation," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Jan. 2022, pp. 1–5.

[29] M. Gadhavi and C. Patel, "Student final grade prediction based on linear regression," *Indian J. Comput. Sci. Eng.*, vol. 8, no. 3, pp. 274–279, 2017.
[30] B. Sravani and M. M. Bala, "Prediction of student performance using linear regression," in *Proc. Int. Conf. Emerg. Technol. (INCET)*, Jun. 2020, pp. 1–5.
[31] V. Matzavela and E. Alepis, "Decision tree learning through a predictive model for student academic performance in intelligent m-learning environments," *Comput. Educ., Artif. Intell.*, vol. 2, Oct. 2021, Art. no. 100035.
[32] C. Yan, X. Shen, F. Guo, S. Zhao, and L. Zhang, "A novel model modification method for support vector regression based on radial basis functions," *Struct. Multidisciplinary Optim.*, vol. 60, no. 3, pp. 983–997, Sep. 2019.
[33] S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, and N. A. Md. Ghani, "Multiclass prediction model for student grade prediction using machine learning," *IEEE Access*, vol. 9, pp. 95608–95621, 2021.
[34] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, p. 237, Dec. 2020.
[35] Y. Meier, J. Xu, O. Atan, and M. van der Schaar, "Predicting grades," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 959–972, Feb. 2016.

PALLAVI ASTHANA (Member, IEEE) is currently working as a full time Assistant Professor in the Department of Electronics and Communication Engineering at Amity University, Uttar Pradesh, India. She has done her M.Tech. in Digital Systems and is pursuing Ph.D. in Computer Science and Engineering. Her area of interest includes Instrumentation, machine learning, Control Systems etc. Currently, she is working in the field of Engineering Education and developing the tools for the quantitative measurement of Course learning Outcomes. She has published more than 25 research papers in Reputed Conferences and Journals. She has also written 6 book chapters. Pallavi Asthana is a member of the IEEE and IETE, UK. She has been a key contributor in organizing various International Conference, and workshop at Amity University. She is also a member of Internal Quality Assurance Cell at Amity University.

SUMITA MISHRA (Senior Member, IEEE) received the M.Tech. degree in optical communication from the Shri Govindram Seksaria Institute of Technology and Science, Indore, India. She received Ph.D. degree from Dr. Ram Manohar Lohia Avadh University, Faizabad, India. She is currently affiliated with the Electronics and Communication Engineering Department, Amity School of Engineering and Technology, Amity University, Lucknow Campus, Uttar Pradesh, India. Dr. Mishra is a member of IET (UK), IAENG and served as Senior Member of IEEE. She has more than 50 publications in international journals and prestigious conferences. She has published two patents and serves as reviewer and editorial board member of many highly reputed journals. Dr. Mishra has had key responsibilities in organizing several International Conferences and various international events. She has also delivered Invited Speech at national and international forums. Her current research interests include deep learning, visible light communication, and Internet of things and is guiding several Ph.D. students in this area.

NISHU GUPTA (Senior Member, IEEE) received the Ph.D. degree in electronics and communication engineering from MNNIT, Allahabad, India, in 2016. He is currently a Postdoctoral Fellow (ERCIM Alain Bensoussan Fellowship) with the Department of Electronic Systems, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology (NTNU), Gjøvik, Norway. He was a member of the Zero Trust Architecture Working Group of MeitY-C-DAC-STQC Project under e-Governance Standards and Guidelines, Ministry of Electronics and Information Technology (MeitY), Government of India. His Ph.D. research was devoted to the development of MAC protocols for safety applications in VANETs. He is the author and the editor of several books and book chapters with publishers, including Springer, Taylor & Francis, Wiley, and Scrivener Publishing. His research interests include intelligent vehicles, the IoT, smart city and transportation, and augmented cyber-security. He was a recipient of the Best Paper Presentation Award from NTU, Singapore, in 2019. He serves as a reviewer for various high quality journals.

MOHAMMAD DERAWI received the Diploma, B.Sc., and M.Sc. degrees in computer science engineering from the Technical University of Denmark, in 2007 and 2009, respectively, and the Ph.D. degree in information security from the Norwegian Information Security Laboratory (NISLab). He is currently a Full Professor and received the title as the Youngest Professor of Norway within Electronic Engineering and is specialized within information security, biometrics, smart devices, programming, and micro-controllers. He was active in several EU-projects, including the 7th Framework European project TrUsted Revocable Biometric IdeNtitiEs (TURBINE) and Erasmus Mundus. His research interests include smart technologies, biometrics with specialization on both behavioral biometrics (gait recognition and ECG) and biological biometrics (fingerprint, face) in mobile devices, pattern recognition, and e-health.

ANIL KUMAR is currently the Deputy Pro-Vice Chancellor with Amity University, Lucknow, India, and the Director of the Amity School of Engineering and Technology, Amity University. He has a vast experience of over 30 years in the areas of modeling and simulation, navigation, Kalman filtering, advanced control theory, INS, extended Kalman filters, optimal estimation, flight dynamics, and flight simulation. He has published more than 40 technical articles.