JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

IDENTIFY DIFFERENTIATED LEARNERS TO ENHANCE LEARNING OUTCOMES

[1] Atharva Pandharikar, [2]Deepti Dabal, [3]Rina Gholap, [4]Harshada Sonawane, [5]I Priyadarshini, [6]Ganesh Jadhav

Department of Computer Engineering
KK Wagh Institute of Engineering Education and Research, Nashik, Maharashtra, India,
Affiliated to Savitribai Phule Pune University

Abstract: — The aim of this paper is to investigate the impact of various input parameters on engineering students' first-semester grades and identify influential input parameters. The decline in students' academic performance is a major issue in engineering education nowadays. Hence based on academic performance the present study aims to develop a Machine learning based classifier which classifies the learner as slow learner, average learner and fast learner. Every learner has different learning needs and learning styles. Learning styles play an important role in helping learners retain learned concepts longer and improve their understanding of concepts .In addition to type of learner each student is suggested his or her preferred method of study based on the results of a learning style test.

IndexTerms - — Learning Style, Support Vector Classifier (SVC), Multi Class Classification, students' academic performance.

I. INTRODUCTION

First-year engineering students confront difficulties since concepts are more difficult to comprehend than in prior 12th grade schooling. The major reason is a lack of preparedness for the high level of academic rigors in engineering. Hence there is a need to check level of preparedness. Identification of low-performing students at an early stage is critical so that their faculties and administration can give appropriate help. As a result, it is necessary to identify the parameters that reduce engineering students' performance and to develop a suitable educational technique for predicting low-performing students at an early stage so that necessary assistance can be provided. There are numerous parameters that can influence engineering students' performance. The parameters can be 12th marks, CET/JEE marks, Psychometric test marks, their continuous assessment marks, mid semester exam marks and so on. The current study has two major research goals:

1) To investigate the impact of various input parameter on engineering students' first-semester grades in order to identify influential input parameters,

2) To identify and determine an individual's preferred way of learning i.e. learning style of students.

In this paper, different machine learning classifiers like Support Vector Classifier (SVC) ,Random Forests Classifiers(RFC),Decision Tree and K Nearest Neighbors(KNN) are used to predict students into one of 3 categories: slow, average or fast. The performance of each of the classifiers is compared.

The rest of the paper is organized as follows. Related work is reviewed in Section II. The data that is collected is described in Section III. The methodology adopted is described in Section IV. The Data Visualization techniques are described in Section V. The Model Evaluation Metrics are described in Section VI.

II. RELATED WORKS

Several research have been conducted by various authors to predict students' academic achievement using supervised (classification) . Swati Verma, Rakesh Kumar Yadav,and Kuldeep Kholiya[1] suggest that background parameters(family income, parent occupation, gender, medium/language of study),academic parameters(marks obtained in various tests),psychological parameters(interest in Engineering study) are important parameters in deciding academic performance. The study compared different classifiers like Naïve Bayes Classifier, KNN, Decision Tree to classify students into Grade A and Grade B.

Feldman, Juan and Monteserin Ariel and Amandi, Analía [2] suggest techniques for automatically determining the learning style of students. The study says the learning style of students can be automatically determined by using machine learning techniques like Bayesian Networks, Decision Trees and Neural Networks.

Eyman Alyahyan & Dilek Düştegör[3] suggest use of data mining techniques for student performance prediction .The review suggest using exploratory Data Mining techniques for students' performance prediction with year by year prediction of students performance and reviews different algorithms for performance prediction.\

Radhika R Halde[4] suggests various Psychological factors influence the performance of students .The psychological factors include Motivation level, Concentration Level ,Time management skills, Information Processing, Skills. The students score in these psychological factors is calculated based on the questionnaire circulated to students. Also previous year scores are also taken into account. The previous year results data and the results of questionnaire data is integrated into a single dataset. The correlation coefficient is found between the dataset and the final semester result.

Yamini Joshi, Kaushik Mallibhat, Vijayalakshmi M.[5] predicted students' performance based on different factors like Academic Performances(Continuous Assessment Marks), Interaction with Peers(number of GitHub commits made by student) , Demographic factors(Gender) and interactions with LMS named Moodle which is used to deliver online courses and video content to students. This data is given as input to classifiers like random forest, naive Bayes, decision tree, support vector machine (SVM), and XGBoost and multiclass classification is performed to classify students into 4 categories: Excellent, Good, Average and poor.

Sokkhey Phauk,Takeo Okazaki[6] use domestic, student/individual and school factors to predict students' performance .The attributes used are nominal or ordinal .These attributes are given as an input and the prediction of students is done into 4 classes :Excellent learner, good learner, average learner and slow learner. The classifiers used for classification are Support Vector Machines, Naïve Bayes, C5.0 and Random Forest (RF).The accuracy for each classifier model is evaluated

Hanan Abdullah Mengash[7] suggest predicting 1st year CGPA of student based on various test like high school grade average(HSGA), Scholastic Achievement Admission Test(SAAT) score, and General Aptitude Test score(GAT).The GAT assesses mathematical and verbal skills to measure students' comprehension, logical reasoning, problem solving, and inductive/deductive skills, the SAAT test is based on five subjects biology, chemistry, physics, mathematics, and English. The HSGA is high school test conducted by board. These test are used for taking undergraduate admission in Kingdom of Saudi Arabia. The marks of these tests are given as input to classifiers like Artificial Neural Network(ANN),Decision tress and Support Vector Machine(SVM) predict the final semester result of students

In order to predict academic success R. D. Ibrahim Z[8] compares the effectiveness of Linear Regression (LR) with Artificial Neural Networks (ANN). The academic performance at semester eight was evaluated using the cumulative grade point average (CGPA).The study was carried out at Malaysia's Universiti Teknologi MARA (UiTM) Faculty of Electrical Engineering. The CGPA in the final semester, which is at semester 8, is utilised as the output or the dependent variable, whilst the students' essential subject scores from the first semester were employed as independent variables or input predictor variables.

Vora, D. R., & Rajamani, K[9] suggest techniques for handling large data of student databases. It consists of two modules .The first module handles massive data using the Map Reduce framework (which contains the notion of principal component analysis). Whereas the second module is an intelligent module that forecasts the performance of the students utilizing data processing stages, For this, the deep belief network and support vector machine are combined to create a novel hybrid classifier, which is presented in this paper. The suggested classifier's output provides a reliable forecast of student performance.

Jayaprakash, S. [10] suggests academic achievement is based on several demographic factors .This study also suggests a method known as improved random forest classifier for use as an interim or early warning system for risk based on qualities and behaviours that have been shown to be significant in determining of students' performance. The study reveals factors as gender and family Size, parenting status, mother and father education, mother and Father Job are one of the influencing factors that would be affect student achievement negatively

III. DATA COLLECTION AND GATHERING

i. Academic data:

Real time data was collected for 1st year Engineering students of K K Wagh Institute of Engineering and Research, Nashik. The target variable is classification of students' academic performance based on final result of particular semester/year. The features used for classification are 12th/HSC PCM Percentage, CET/JEE Percentile, Mid Semester Exam Marks, Psychometric test marks, practical/term work marks, LearniCo Test Marks

Features	Description
12th/HSC Marks	The PCM percentage of 12th marks
CET/JEE Percentile	The percentile marks of Common Entrance Test(CET)/JEE(Joint Entrance Examination) for taking admission
Mid Semester Exam Marks	The marks obtained by student in In sem exam of University of 30 marks

Psychometric test marks	The general aptitude test, technical tests based on numerical reasoning measuring the basic engineering aptitude for engineers. The test is of 30 marks
Term Work marks	The marks assessing the practical's/assignments performed by students throughout the semester
LearniCo Tests	MCQ (Multiple Choice Questions) based on the topic being taught. The test is taken after every topic being taught.

Fig 1.Dataset Description

ii. Learning Style tests

Learning styles refer to various strategies of learning or understanding new knowledge, as well as the manner in which a person absorbs, comprehends, expresses, and recalls information. By understanding the learning style of students' teachers can understand the type of learner and students that allow them develop a better grasp of how to incorporate various learning types into their lesson planning and study practices. The Richard Felder learning style test was conducted in different classrooms and students were asked to fill the scores of the results of tests in Google forms

The Richard Felder Learning Style Test [11] is an instrument to access the preferred way of learning of students based on 4 types:

1. Active/reflective learners

a) Active Learners- Active learners' students are encouraged to participate in their learning by thinking, talking, discovering, and creating. Active learners learn best when they discuss what they've learned, experiment with it, or put it to the test. This learning process requires timely feedback from either the instructor or fellow pupils.

b) Reflective Learners-Reflective learners ask questions and think critically about their own previously formed views. They prefer to study on their own rather than participate in group discussions.

2. Sensing/Intuitive learners

a) Sensing Learners- Sensing learners recall and understand material better when they can see how it applies in real life. When the teacher teaches the theatrical concept information, the sensing learner tries to relate the contents to the real world as much as possible, and understand how concepts apply in practice. These students learn best through real-world examples.

b) Intuitive Learners- Intuitive learners will prefer to focus on possibilities and 'what could be'. These types of learners dislike repetitive tasks since they like learning new things. They can rapidly learn new material and will typically hunt for fresh approaches to achieve answers rather than employing the same one.

3. Visual/Verbal Learners

a) Visual Learners- Visual learning is a learning method in which pupils prefer to express ideas and concepts through visuals, graphics, colors, and maps. Visual learners want information to be seen in order to learn it. These students are likely to have a photographic memory and may retain information using color, tone, and brightness. Visual learners will benefit from seeing diagrams drawn out in class, such as on a chalkboard or in slideshows. Color-coding notes, generating to-do lists, and, other visual learning techniques help visual learners learn best.

b) Verbal Learners-Verbal learners are fluent in both written and spoken language. They learn best when reading, speaking or listening to information

4. Sequential/Global Learners

a) Sequential Learners-Sequential learners often develop comprehension in linear phases, with each step logically following the previous one.

b) Global Learners-Global learners lean in enormous leaps, taking in material and knowledge at random without establishing connections and then suddenly comprehending it.

IV. RESEARCH METHODOLOGY

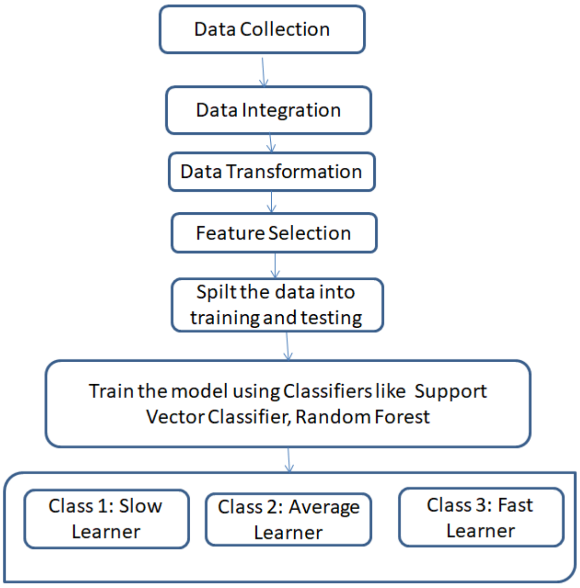

Fig 2.Steps for ML Model Preparation

1. Dataset Collection: The data for 1st year Engineering students was collected [III]**.**

2. Data Integration: The data is integrated from different sources and finally converted into a Comma Separated Value (CSV) file. Data integration is used to handle data heterogeneity as teachers use different formats to store the data

3. Data Transformation: Data transformation is a technique used to convert the raw data into a suitable format that efficiently eases data mining and retrieves strategic information.

4. Feature Selection: Feature selection is a method of reducing the redundant, irrelevant, or noisy features from a feature set in order to choose the subset of the most relevant characteristics.

5. Multiclass Classification: Multiclass classification or multinomial classification is the task of categorizing samples into one of three or more classes in machine learning. Here every student is classified into one of 3 classes(slow, average or fast learner).

6. Classifiers
(a) Support Vector Classifier(SVC):
A Support Vector Classifier classifies the data by:
Hyperplane: Several lines/decision boundaries can be used to separate classes in n-dimensional space, but SVM selects the optimum decision boundary to categorize the data points. A hyperplane is the optimal choice boundary.

Support Vectors: Support Vectors are the data points or vectors that are closest to the hyperplane and affect the position of hyperplane The features are given as input to SVC classifier along with target variable for training and testing of dataset.

Here One vs. All SVM (Support Vector Machine) Strategy is used:
In One vs. All classification, each class has one binary SVM to distinguish its members from members of other classes.
One vs All Classification:
One-vs-All is a heuristic technique for using binary classification algorithms for multi-class classification.
Among several classes, the one vs. all technique evaluates only one class as positive and all other classes as negative. The one-versus-all tactic divides an N-class problem to n-binary problems, which is created scenario as one class to all other classes.
To determine the output of one vs. all, we examine the output of each class and seek for positive output classes that imply instances belong to that class. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary

In One Vs All SVM, if we have N class problem, then we learn N SVMs:
SVM number -1 learns “class output = slow learner” vs. “class output ≠ slow learner″
SVM number -2 learns “class output = average learner” vs. “class output ≠ average learner″
SVM number -3 learns “class output = fast learner” vs. “class output ≠ fast learner”
Then to predict the output for new input, SVM finds which one puts the prediction the farthest into the positive region (behaves as a confidence criterion for a particular SVM)

(b) Decision Tree Classifier:
Decision Tree Classifier is capable of performing multi-class classification on a dataset. Decision Trees consist of nodes and branches The nodes can also be divided into a root node (the tree's starting node), decision nodes (sub-nodes that branch out based on circumstances), and leaf nodes (nodes that cannot branch further).By segmenting a dataset into smaller sub segments depending on certain criteria at each level, a decision tree is created. Every time a partition or division is established, it is intended that related data samples be grouped together. The decision of how to split depends on entropy and information gained.

(c) Random Forest Classifier
Random forest is a Supervised Machine Learning Algorithm used for classification
Random Forest Algorithm follows following steps:
Step 1: Select random samples from the dataset
Step 2: Construct a decision tree for every sample. A decision tree is a graphical structure that shows every possible outcome or consequence of a choice using a branching mechanism. The structure contains the root node, which represents the attribute with the most significant information gain, branches, which indicate the result of a test, the internal node, which represents the test on an attribute or feature, and leaf nodes, which store a class label.
Step 3: For a new sample result is predicted from every decision tree.
Step 4: Voting will be performed for every predicted result.
Step 5: The most voted prediction result is selected as the final prediction result.

(d) K Nearest Neighbors
The K-nearest neighbors (KNN) method predicts the values of new data points based on "feature similarity," which further indicates that the new data point will be given a value depending on how closely it resembles the points in the training set. The algorithm works in following steps:

Step 1: Select value K, or the nearest data points, as the value. Any integer can be K.
Step 2: For a new data point calculate the distance using either the Euclidean, Manhattan, or Hamming between the new data point and each data point in the dataset.
Step 3: Sort the data points in ascending order depending on the distance value.
Step 4: The top K rows from the sorted array will then be selected.
Step 5: The test point will be assigned a class based on the most frequent class of these rows.

V.MODEL EVALUATION METRICS

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

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

1. Accuracy: Accuracy score is the ratio of number of correct predictions made by a model in relation to the total number of predictions made

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

True Positive (TP) is an outcome where the model correctly predicts the positive class.
True Negative (TN) is an outcome where the model correctly predicts the negative class.
False Positive (FP) is an outcome where the model incorrectly predicts the positive class.
False Negative (FN) is an outcome where the model incorrectly predicts the negative class.

2. Precision, Recall and F1-Score
a) Precision: Precision is the ratio of correct positive prediction made by model in relation to actual positive samples present
b) Recall: The recall is calculated as the ratio between the numbers of Positive samples correctly classified as Positive to the total number of Positive samples.
c) F1 Score: The F1 score is defined as the harmonic mean of precision and recall.

$$MacroAveragePrecision = \frac{\sum_{k=1}^{K} Precision_k}{K}$$

$$MacroAverageRecall = \frac{\sum_{k=1}^{K} Recall_k}{K}$$

$$F1\ score = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

3. Macro-average precision, recall and F1-score:

In multiclass classification, the macro-average precision and macro average recall score are calculated as the arithmetic mean of the accuracy and recall scores of individual classes.

The performance metrics for the different models are listed in table

Classifier	Accuracy	Precision	Recall	F1 Score
Random Forest	100%	100%	100%	100%
Support Vector Machine(SVM)	54%	42%	54%	43%
Decision Tree	79%	79%	79%	79%
K Nearest Neighbors(KNN)	64%	58%	64%	61%

Fig 3: COMPARISON OF THE RESULTS OF DIFFERENT CLASSIFIERS

VI.CONCLUSION

Early prediction of student performance can assist institutions in taking timely decisions, such as preparing for suitable training to promote student achievements .The dataset used for student performance prediction consists of previous academic records (12^{th} marks, CET percentage) and current academic records(Psychometric Tests, Insem marks ,Practical/Term work marks) .This data is given an input to classifiers to predict the learners into 3 categories(slow learners, average learners and fast learners) .Different classifiers were used to predict the output but Random Forest Classifier gave the highest accuracy of 100%.Identification of slow learners at an early stage could serve as a warning for the students, and also help the teachers to keep track of progress of students. The data was gathered only for one institute but the model can be implemented or other institutes also

REFERENCES

[1] Swati Verma, Rakesh Kumar Yadav, and Kuldeep Kholiya, "Prediction of Academic Performance of Engineering Students by Using Data Mining Techniques," International Journal of Information and Education Technology vol. 12, no. 11, pp. 1164-1171, 2022.

[2] Feldman, Juan & Monteserin, Ariel & Amandi, Analía. (2014). Automatic detection of learning styles: state of the art. Artificial Intelligence Review. 44. 10.1007/s10462-014-9422-6.

[3] Alyahyan, E., Düştegör, D. Predicting academic success in higher education: literature review and best practices. Int J Educ Technol High Educ 17, 3 (2020). https://doi.org/10.1186/s41239-020-0177-7

[4] R. R. Halde, "Application of Machine Learning algorithms for betterment in education system," 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 2016, pp. 1110-1114, doi: 10.1109/ICACDOT.2016.7877759.

[5] Y. Joshi, K. Mallibhat and V. M., "Students' Performance Prediction Using Multimodal Machine Learning," 2022 IEEE IFEES World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC), Cape Town, South Africa, 2022, pp. 1-5, doi: 10.1109/WEEF-GEDC54384.2022.9996212.

[6]Phauk, Sokkhey & Okazaki, Takeo. (2020). Hybrid Machine Learning Algorithms for Predicting Academic Performance. International Journal of Advanced Computer Science and Applications.11.10.14569/IJACSA.2020.0110104

[7] H. A. Mengash, "Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems," in IEEE Access, vol. 8, pp. 55462-55470, 2020, doi: 10.1109/ACCESS.2020.2981905..

[8] R. D. Ibrahim Z, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression, 2007," presented in 21st Annual SAS Malaysia Forum Shangri-La Hotel, Kuala Lumpur, 2007,5th September 2007

[9] Vora, D. R., & Rajamani, K. (2019). A hybrid classification model for prediction of academic performance of students: A big data application. Evolutionary Intelligence.https://doi.org/10.1007/s12065-019-00303-9

[10] S. Jayaprakash, S. Krishnan and V. Jaiganesh, "Predicting Students Academic Performance using an Improved Random Forest Classifier," 2020 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2020, pp. 238-243, doi: 10.1109/ESCI48226.2020.9167547.

[11] Soloman, Barbara & Felder, Richard. (1999). Index of Learning Styles Questionnaire. Learning.