Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

CLASSIFICATION OF STUDENTS BASED ON THEIR ACADEMIC PERFORMANCE

E. GOWRI, J. BHARGAVI, I. HEMANTH, Student, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh., India.
SYED SHAHEEN, Assistant Professor, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh., India.

ABSTRACT: In this study, we studies in the learning field that investigated the ways of applying machine learning techniques for various educational purposes. One of the focuses of these studies is to identify high-risk students, as well as to identify features that affect the performance of students. These features were derived from the students' grades from the Universities, an undergraduate public institution. Based on these features, we perform a study to identify different student groups of interest, while at the same time, identify their importance. We implement machine learning algorithms like SVM, Random Forest, Decision Tree, Gradient boosting, MLP classifiers for comparative analysis.

Keywords--- Feature Extraction, Machine learning, Classification, Prediction.

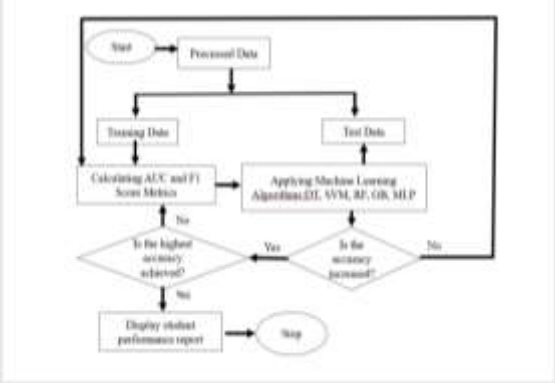
INTRODUCTION

Nowadays, as a result of increasingly competitive academic markets, most universities face difficulties in attracting potential learners. There is often a great need to be able to predict future students' behavior in order to improve curriculum design and plan interventions for academic support and guidance on the curriculum offered to the students. Predicting students' performance is one of the most important topics for learning contexts such as schools and universities, since it helps to design effective mechanisms that improve academic results and avoid dropout, among other things. At first, we need the gathered dataset from the web or from your school/university (if you're utilizing school/college information kindly do this with consent as it were).Presently we have to pre-process the information like Data cleaning, transformation, and filtering, etc. control on its own, to get where it is going.Student academic performance prediction is involved in Educational data mining (EDM) that focuses on applying data mining techniques in educational aspect. According to the target of prediction, EDM can be divided into three groups consisting of prediction learning progress, prediction dropout rate and prediction final average grade of students. On the other hand, EDM can be classified into categories such as prediction, regression, clustering, classification, relationship mining, discovery with model, outlier detection

Due to the Corona pandemic, student performance in academics has declined and drop-out numbers have increased which calls to taking an active approach for finding a solution.

The good academic performance of students in every educational system is important. Developing tools to support students and learning in a traditional or online setting is a significant task in today's educational environment. The initial steps towards enabling such technologies using machine learning techniques focused on predicting the student's performance in terms of the achieved grades.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023



PROPOSED SYSTEM

The primary goal of our work is detecting at-risk students instead of determining performance levels of students. Classifying students according to their performances in different levels (e.g. poor, average, good, excellent etc.) The best performing methods are the Decision Tree and MLP classifiers, based on AUC and F1 score metrics.

In this study, we propose the feature extraction approach using MLP and GB to process student data and determine the best accuracy of the model. The classifiers used in this model are SVM, RF, DT, GB, MLP. So comparing all the classifiers MLP and GB gives best accuracy than other classifiers

Everyone has the right to education. For Higher Educational Institutions, students serve as its best asset. The prediction of students' success in their academic performance is then vital for it will benefit both students and professors, enabling the latter to do proactive measures and find ways in helping students learn, ultimately improving their academic performance.

SVM Algorithm

Machine learning involves predicting and classifying data and to do so we employ various machine learning algorithms according to the dataset. SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple: The algorithm creates a line or a hyperplane which separates the data into classes. In machine learning, the radial basis function kernel, or RBF kernel, is a popular kernel function used in various kernelized learning algorithms. In particular, it is commonly used in support vector machine classification. As a simple example, for a classification task with only two features (like the image above), you can think of a hyperplane as a line that linearly separates and classifies a set of data.

Intuitively, the further from the hyperplane our data points lie, the more confident we are that they have been correctly classified. We therefore want our data points to be as far away from the hyperplane as possible, while still being on the correct side of it.

So when new testing data is added, whatever side of the hyperplane it lands willdecide the class that we assign to it.

RANDOM FOREST ALGORITHM

Random Forest is a popular machine learningalgorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML.

Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

DECISION TREE ALGORITHM

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems

It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.The decisions or the test are performed on the basis of features of the given dataset.

GRADIENT BOOSTING

Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting. Gradient boosting models are becoming popular because of their effectiveness at classifying complex datasets, and have recently been used to win many Kaggle data science competitions.

The Python machine learning library, Scikit-Learn, supports different implementations of gradient boosting classifiers, including XGBoost. By using multiple algorithms a single accurate train model willbe generated. In all this algorithms Gradient Boosting give better performance.

MULTI LAYER PERCEPTRON

The Multilayer Perceptron (MLP) algorithm is a type of artificial neural network that has widespread use in classification and regression tasks. It consists of an input layer, one or more hidden layers, and an output layer, with each layer's nodes connected to the next. These connections have weights, which the MLP algorithm adjusts iteratively to minimize the difference between the predicted and actual outputs. It uses backpropagation to calculate the output layer's error and propagate it backward through the network to modify the weights. Due to its ability to model intricate non-linear relationships, the MLP algorithm finds extensive application in fields like finance, healthcare, and image and speech recognition.

LITERATURE SURVEY

In 2022, Neeta Sharma et. el. examines the value of analysing academic performance in the year 2022. Additionally using extracurricular activities as a predictor, the authors' study. On their selected predictor, they use Logistic Regression (LR) and the K-NN model. In comparison to KNN, the accuracy and F1 Score of logistic regression are the highest. As part of a future project, the authors advise implementing SVM, MLP, Decision Trees.

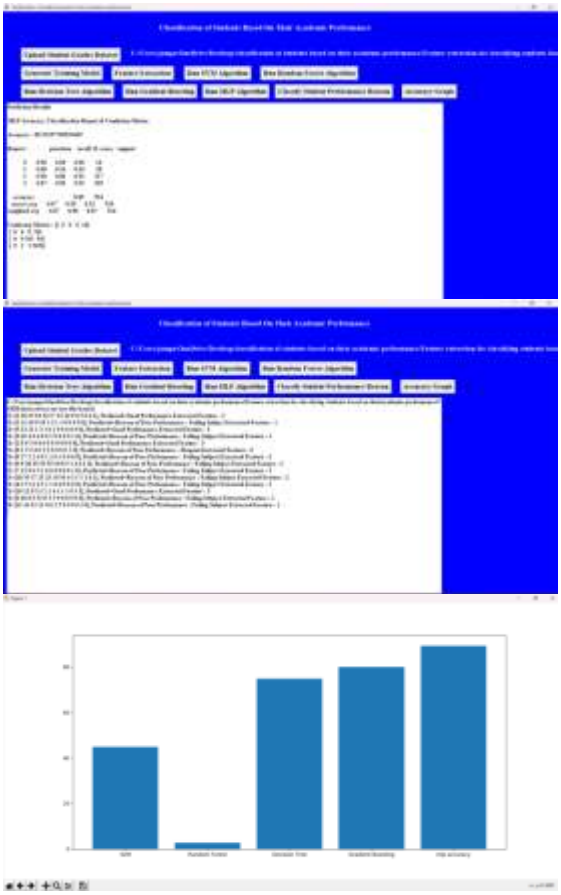
In 2022, Qiu et. el. highlighted the use of information technology in the eLearning process in terms of modern education to improve quality of educational that may be attained through real-time monitoring and feedback. As a result of his research, the author developed the BCEP (Behaviour Classification based eLearning Performance) prediction Framework that found to be moreeffective than conventional classification techniques.

In 2021, LoniaMasangu et al. pointed out that most students receive some instruction on how to use online learning environments so they may comprehend the method of precise outcome prediction. According to their methodology, students those do not turn-in an valuation or do not attend a requisite forum will receive a grade of 0% for that particular assessment because all students are expected to receive the same amount of course information.

SAMPLE RESULTS

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023



CONCLUSION

In this project to accurate identify students that are at risk. These students might fail the class, drop it, or perform worst than they usually do. We extracted features from historical grading data, in order to test different simple and sophisticated classification methods based on big data approaches. The best performing methods are the MLP and GB classifiers, based on AUC and F1 score metrics. We also got interesting findings.

FUTURE SCOPE

Future study on this topic is advised to use more machine learning (ML) prediction models, namely, MLP, DT, GB among others, for comparably better outcomes and to develop the best algorithm for predicting students' academic success.

REFERENCES

L. Breiman. Random forests. Machine learning, 45(1):5{32, 2001.
L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. Classification and regression trees. CRC press, 1984.
C. Cortes and V. Vapnik. Support-vector networks. Machine learning, 20(3):273{297, 1995.
J. H. Friedman. Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189{1232, 2001.
Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran. Machine learning based student grade prediction: A case study. arXiv preprint arXiv:1708.08744, 2017.
J. E. Knowles. Of needles and haystacks: Building an accurate statewide dropout early warning system in wisconsin. Journal of Educational Data Mining, 7(3):18{67, 2015.

S. Kotsiantis, C. Pierrakeas, and P. Pintelas. Predicting students'performance in distance learning using machine learning techniques. Applied Articial Intelligence, 18(5):411{426, 2004.
J. McFarland, B. Hussar, C. de Brey, T. Snyder, X. Wang, S. Wilkinson-Flicker, S. Gebrekristos, J. Zhang, A. Rathbun, A. Barmer, et al. Undergraduate retention and graduation rates. In The Condition of Education 2017. NCES 2017-144. ERIC, 2017.