Student Performance Analyser using Supervised Learning Algorithms

Srricharan Bolisetti, Pranith Reddy Sankepally,
Sai Rohith Reddy Kunta, Rohith Reddy Lakkireddy and
Manoj Kumar Vemula

June 7, 2021

STUDENT PERFORMANCE ANALYSER USING SUPERVISED LEARNING ALGORITHMS

Mr SRRICHARAN BOLISETTI
Vardhaman College of Engineering, Shamshabad
charan.classic@gmail.com

Mr SANKEPALLY PRANITH REDDY
Vardhaman College of Engineering, Shamshabad
sankepally256@gmail.com

Mr KUNTA SAI ROHITH REDDY
Vardhaman College of Engineering, Shamshabad
sairohithreddyk@gmail.com

Mr LAKKIREDDY ROHITH REDDY
Vardhaman College of Engineering, Shamshabad
rohith11631@gmail.com

Mr. MANOJ KUMAR VEMULA
Assistant Professor
Vardhaman College of Engineering, Shamshabad
manojkumarvemula@vardhaman.org

ABSTRACT

In today' academic environment, it's essential to make tools that facilitate students learn during a casual or online environment. the primary step in victimization machine learning technology to boost these advances focuses on predicting student performance supported the results achieved. one in each of these ways is that they are doing not provide competent leads to expecting underperforming students. Our work aims to double overlap. To beat this limitation, we have a tendency first to check whether or not it is doable to predict underperforming students a lot accurately. Second, we developed numerous human explainable characteristics to live these factors to determine that factors lead to poor tutorial performance. These factors are supported student ratings at the University of Minnesota. Considering these factors, you ought to analyze to spot numerous student stakeholders and perceive their importance.

Keywords— *Decision Tree Algorithm, Random Forest Algorithm, Feature Extraction, Confusion Matrix, Precision, Accuracy.*

I. INTRODUCTION

Higher instructive foundations continually attempt to improve the maintenance and achievement of their selected understudies. As per the US National Center for Education Statistics [8], 60% of college understudies on four-year degrees won't graduate at a similar foundation where they began inside the initial six years. Simultaneously, 30% of school green beans drop out after their first year of school. Therefore, schools search for approaches to serve understudies all the more productively and viably. This is the place where information mining is acquainted with give a few answers for these issues. Instructive information mining and learning examination have been created to provide devices for supporting the learning interaction, similar to screen and gauge understudy progress, yet in addition, anticipate achievement or Most of the current methodologies centre around distinguishing understudies in danger who could profit by additional help with request to finish a course or action effectively. A primary assignment in this interaction is to anticipate the understudy's presentation as far as evaluations. While sensible forecast exactness has been accomplished [14, 10], there is a massive shortcoming of the models proposed to distinguish the poor-performing understudies [18]. Typically, these models will, in general, be over-hopeful for the presentation of understudies, as most of the understudies' progress nicely or have palatable enough execution.

II. RELATED STUDY

In any case, achievement and disappointment can be relative or not. For instance, a B− evaluation may be viewed as an awful evaluation for a phenomenal understudy while being a passing mark for a highly powerless understudy. We examined various approaches to characterize gatherings of understudies taking a course: bombing understudies, understudies dropping the class, understudies performing more awful than anticipated, and understudies performing more terrible than anticipated while contemplating the trouble an approach. To understand the learning cycle and its most significant attributes, we have made highlights that catch potential factors that impact the evaluations toward the finish of the semester. Utilizing these highlights, we present a far-reaching study to address the accompanying inquiries: which highlights are good pointers of an understudy's presentation? Which highlights are the most significant? The disclosures are interesting, as different features are the most critical for various arrangement errands.

III. PROPOSED MODEL

We will propose the system by using which the user can give a test on specific educational or subject categories. When the student completes the test, the system will calculate the user's performance by using the algorithm decision tree. The system will suggest to the teacher which topics the user is weak or need to study again. To take care of the issues confronted with manual assessment composing, there is a need for a modernized framework to deal with every one of the works. We propose an application that will give a workplace that will be adaptable and will give simplicity of work and reduce the time for report generation and other paperwork. Today many organizations are conducting online examinations worldwide successfully and issue results online. Still, they do not measure the performance of the student and the teacher not know about the weak points of the students, and we are focusing on this issue. The main advantage is that the evaluation of answers can be fully automated for all questions. Other essay-type questions can be evaluated manually or through an automated system, depending on the nature of the questions and the requirements. This record is also seen by a teacher for analyzing the student's performance, and they can take suitable action to improve the student performance before the student in the critical area. This technique will

monitor and evaluate the student academic performance at different year levels before the final test to forecast the students' weaknesses. The teacher can play the admin role, who has the authority to add subjects, topics, and questions for test purposes.

IV. EXPERIMENTAL SETUP

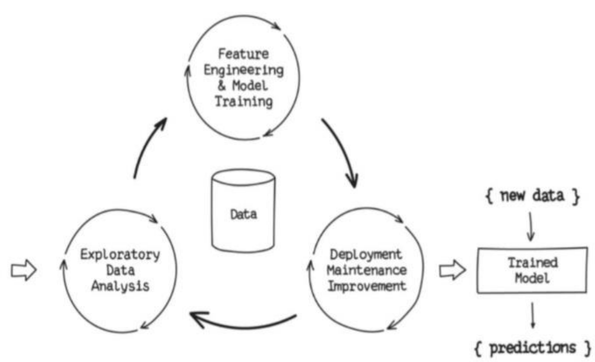

Fig-1 : Process Model

Exploratory Data Analysis :

In data mining, Exploratory Data Analysis (EDA) is a way to deal with dissecting datasets, to sum up, their principle qualities regularly with visual strategies. EDA is utilized for seeing what the information can advise us before the demonstrating task.

Feature Engineering and Model Training :

Feature Engineering uses domain information on information to make features that make machine learning algorithms work.

Preparing a model essentially implies picking up (deciding) great qualities for all the weights and the bias from labelled examples. In supervised learning, machine learning algorithms assemble a model by analyzing numerous examples and endeavouring to find a model that limits loss; this cycle is called empirical risk minimization.

Deployment Maintainance and Improvement:

Deployment of an ML model basically implies the combination of the model into an existing production environment which can take in an input and return an output that can be utilized in making beneficial business choices.

Another approach to staying up with the latest model is to have a computerized framework to assess and retrain your models constantly. This framework is frequently alluded to as continuous learning and may look something like this: Save new training data as you get it.

V. ALGORITHMS:

SVM Algorithm: Machine learning includes anticipating and grouping information, and to do so, we utilize different Machine learning algorithms as per the dataset. Support Vector Machine is a linear model for grouping and regression issues. It can take care of linear and non-linear issues and function admirably for some practical problems. The possibility of SVM is straightforward: The algorithm makes a line or a hyperplane that isolates the data into classes. In machine learning, the radial basis function Kernal, or RBF portion, is the main piece of Kernal function utilized in different kernelized learning algorithms. In particular, it is generally used in support vector machine classification. As a basic example, for a classifying task with just two features (like the picture above), you can think about a hyperplane as a line that linearly isolates and characterizes a set of data.

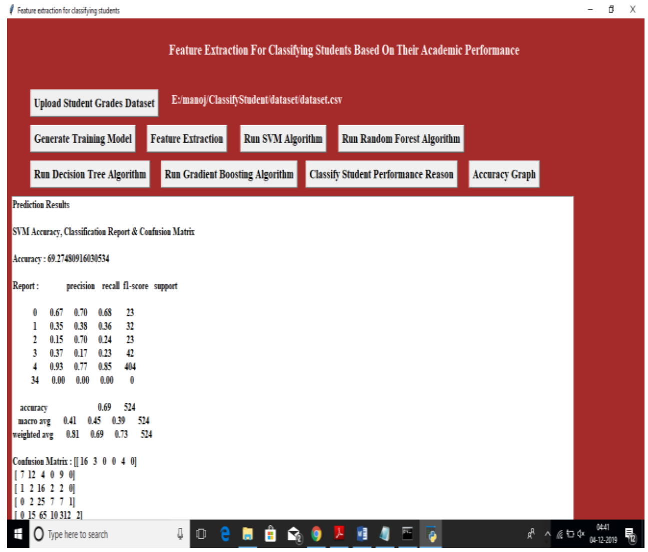

Fig-2 : SVM output-screen

In the above screen, SVM accuracy is 69%, and we can see the score value also. Now click on 'Run Random Forest Algorithm.'

Random Forest Algorithm: It's an ensemble algorithm which means internally, it will use multiple classifier algorithms to build an accurate classifier model. Internally this algorithm will use a decision tree algorithm to generate its train model for classification.

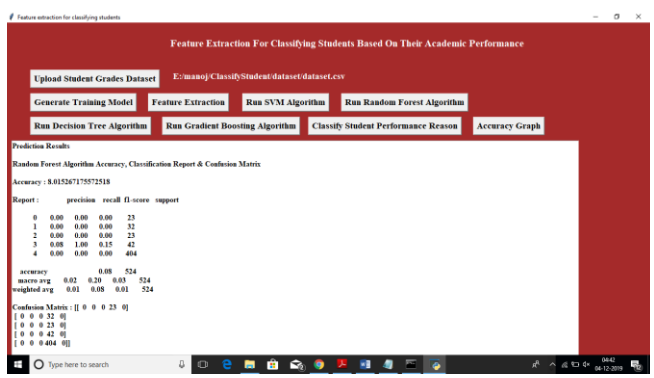

Fig-3 : Random Forest output-screen

In the above screen, the random forest got only 8% of accuracy. Now click on the 'Run Decision Tree Algorithm' button to build a tree model.

Decision Tree Algorithm:

This algorithm will create a building model by arranging all comparative records in a similar part of a tree and proceed till all documents organized in the whole tree. The total tree will have alluded as a classification train model.

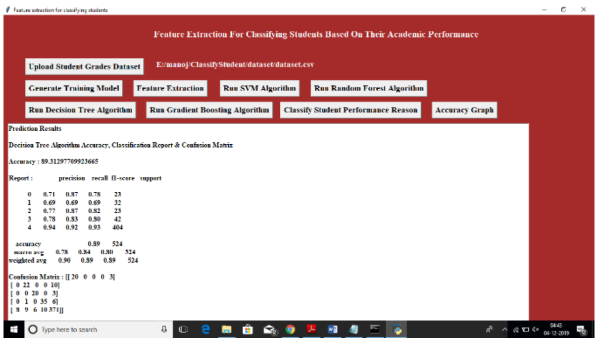

Fig-3: Decision Tree output-screen

The above screen decision tree got 89% accuracy; now click on 'Run Gradient Boosting Algorithm'.

Gradient Boosting Algorithm:

Gradient Boosting classifiers are a series of machine learning algorithms that mix weak learning models to make reliable prognosticative models. Classification of complicated information sets has recently been accustomed win many Kaggle data science competitions.

The Python Machine Learning library, Scikit-Learn, upholds various implementing of gradient boosting classifiers, including XGBoost. By using multiple algorithms, a single accurate train model will be generated. In all these algorithms, Gradient Boosting gives better performance.

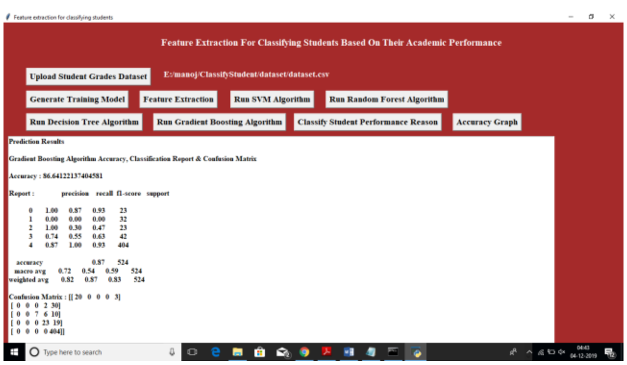

Fig-4: Gradient Boosting output-screen

In the above screen, gradient boosting got 87% accuracy, and the decision tree got high accuracy, but the score is less compare to gradient boosting. Now click on the 'Accuracy Graph' button to get below the accuracy graph.

VI. RESULT ANALYSIS

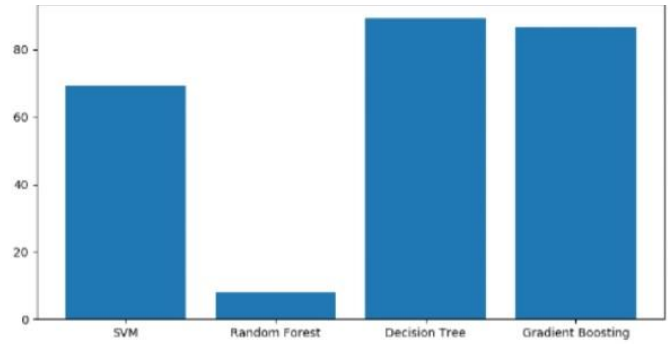

Fig-5 : Graph comparison of different models

In the above screen, the x-axis represents the algorithm name, and the y-axis represents the accuracy of that algorithm. Now we can test new student records on this training model to predict or classify recent student performance. To check new students, We need to upload the 'text.txt' test dataset from the dataset folder, and this dataset contains the below data.

VII. CONCLUSION AND FUTURE SCOPE

The motivation behind this paper is to recognize understudies that are in danger precisely. These understudies may bomb the class, drop it, or perform most noticeably awful than they typically do. We removed highlights from chronicled evaluating information to test distinctive basic and complex order strategies dependent on enormous information draws near. The Gradient Boosting and Random Forest classifiers' best performing strategies are in light of AUC and F1 score measurements. We likewise got intriguing discoveries that can clarify understudy execution.

VIII. FUTURE SCOPE

One of our objectives is to contemplate which components are significant pointers of an understudy's exhibition. We can get numerous bits of knowledge on the elements that influence understudy execution. For instance, the highlights identified with the understudies 'grades (bunch 1) have a generally excellent prescient capacity in practically every one of the errands, aside from the undertaking of foreseeing the W grades. In this undertaking, highlights identified with the course's trouble and fame (bunch 4) just as highlights that are course-explicit (bunch 8) figure out how to accomplish a similar precision when utilizing every one of the highlights. This shows that the explanations for substitution dropout are a lot of systematics than particular substitution. Future best result's an element of the choice student syllabus for this semester. RelCF: However, we have a tendency to found that for RelCF issues, the performance of the component teams coupled to the highlights of the key courses is better . In contrast, the explicit understudy gatherings have insignificantly most observably horrendous performance, diverged from the errand of RelCF. This is going because, for RelCF, we think about how different understudies typically perform on the objective course. Every gathering has sufficient data for the RF to accomplish execution, which is as great as 75% of the best case, i.e. when utilizing every one of the highlights.

IX. REFERENCES

[1] L. Breiman. Arbitrary timberlands. AI, 45(1):5–32, 2001.

[2] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. Arrangement and relapse trees.

CRC Press, 1984.

[3] C. Cortes and V. Vapnik. Backing vector organizations. AI, 20(3):273–297, 1995.

[4] J. H. Friedman. Eager capacity guesses: an angle boosting machine. Records of measurements, pages 1189–1232, 2001.

[5] Z. Iqbal, J. Qadir, A. N. Mian, and F. Kamiran. AI based understudy grade expectation: A contextual investigation. arXiv preprint arXiv:1708.08744, 2017.

[6] J. E. Knowles. Of needles and piles: Building a precise statewide dropout early cautioning framework in Wisconsin. Diary of Educational Data Mining, 7(3):18–67, 2015.

[7] Hussain, M.A., Bhuiyan, A., Luu, C.D., Smith, R.T., Guymer, R.H., Ishikawa, H.Ramamohanarao, K.: Classification of sound and sick retina utilizing SD-OCT imaging and Random Forest calculation. PLoS ONE 13(6), 19–28 (2018).

[8] Subasi, A., Alharbi, L., Madani, R., Qaisar, S.M.: Surface EMG based classification of basic hand developments utilizing turn timberland. In: 2018 Advances in Science and Engineering Technology International Conferences (ASET), pp. 1–5 (2018).

[9] Thanh Noi, P., Kappas, M.: Comparison of arbitrary timberland, k-closest neighbor, and support vector machine classifiers for land cover classification utilizing Sentinel-2 symbolism. Sensors 18(1), 12–18 (2018)

[10] Das, S., Behera, R.K., Rath, S.K.: Real-time estimation examination of Twitter streaming information for stock expectation. Procedia Comput. Sci. 132, 956–964 (2018).

[11]Zuo, Y., Avraham, G., Drummond, T.: Generative Adversarial Forests for BetterConditioned Adversarial Learning. arXiv preprint arXiv:1805.05185, pp. 1–17 (2018)

[12] Houborg, R., McCabe, M.F.: A half and half preparing approach for leaf region record assessment via cubism and arbitrary timberlands AI. ISPRS J. Photogram. Far off Sens. 135, 173–188 (2018)

[13] Teluguntla, P., Thenkabail, P., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Huete, A.: A 30-m Landsat-determined cropland degree result of Australia and China using random woodland AI calculation on Google Earth Engine cloud computing platform. ISPRS J. Photogram. Far off Sens. 144, 325–340 (2018)

[14] Wang, S., Azzari, G., Lobell, D.B.: Crop type planning without field-level marks: Randomforest move and unaided bunching procedures. Far off Sens. Environ. 222, 303–317(2019)

[15] Khatami, R., Mountrakis, G., Stehman, S.V.: A

meta-investigation of far off detecting research supervised pixel-based land-cover picture classification measures: general rules for practitioners and future exploration. Distant Sens. Environ. 177,89–100 (2016)

[16] Gómez, C., White, J.C., Wulder, M.A.: Optical distantly detected time arrangement information for landcover classification: a survey. ISPRS J. Photogram. Distant Sens. 116,55–72 (2016)