Student Grading System using Machine Learning

[1]N B Kalashree, [2]D Bhulakshmi, [3]Namratha K S, [4]Narayani Khajuria, [5]Nischita Reddy P

[1,3,4,5]Student, School of C&IT, REVA University, Bengaluru-560064
[2]Professor, School of C&IT, REVA University, Bengaluru -560064

Abstract

The student's overall achievement and their pass rates shows the way of coaching given by the college system, this is vital improving the students pass rates and decrease failure rates. Researchers have used two algorithms in which are Decision tree (DT) algorithm and Support Vector Machine (SVM) algorithm in order to search the necessary student characteristics and conclude the student pass rates, but researchers do not forget about the capabilities of student dependency and coefficient of initialization. Therefore, if we look into this study, student grade prediction the use of DT and SVM. The capabilities are taken into consideration to be crucial is this study, we apply improved genetic algorithm to handle best function choice problem. Then device gaining knowledge of algorithm is applied. The results display the test can achieve higher accuracy of prediction. This study is used to help the students who are lagging behind the studies and facing some risks of graduating.

1. Introduction

In Schools, tests, examinations are conducted to assess the students based on their understanding. The total pass percentage or student pass rates depend on various factors. Out of which teaching level plays an important role. By increasing this factor, we can increase the total income. This platform provides results with these results we can identify, the average and poor students and can help their grade betterment. Based on the results, the college or an institution finalizes its overall status. This helps the institution to grow and for the common people to recognize it in a better way. Other than this in an institution, centrally and globally the platform helps predict marks and ranks for country and state level exams like NEET and CET.

The large data sets from big records units are sorted by the process of Data mining, also called as Knowledge Discovery which build the relationships through records to solve the difficulty. According to the education system, mining is found as a new idea having information about different mining strategies. For several decades, discovering doubtlessly the information which are valuable through large extent of statistics which are stored in the dataset. As the experts and students are from different research fields so it is a horrify project to discover and create models, as using this technology of mining has made the present-day condition very much satisfied.

This technology of mining shows its importance by warning the particular college about the capabilities of their students, showing their overall achievements and also by discovering the capabilities of their students who are getting effected. Predicting the student behaviour accurately is very much beneficial for the particular institute, if we are able to analyse the capabilities of the students facing difficulties in studies and higher chances of getting failed, the performances of those poor students can be improved by giving them a first-class coaching.

2. Problem Statement

There are nevertheless researches open for seeking the key features that have an impact on the pass rates of the students, and figuring ideas to enhance the best results of the pass rates prediction. Therefore, in this study, we used the SVM and DT algorithms optimized with the help of the algorithm : "Grid Search", to classify students

from college into 2 categories: Pass or fail, if you want for the identification the essential characters that affect the overall pass rates of students and are expecting the total pass rates.

Particularly, because of the life of a sure reliability between capabilities, we bring forward functions reliabilities and professional advice to calculate the similarity between student capabilities. There is a need for an improved genetic set of rules to handle characteristic choice problem also called as feature selection problem.

3. Literature Review

The algorithm used for data classification and regression is K-nearest neighbour. Set of rules that stores all the recognized situations and sorts them to new instances primarily depending on comparable capabilities. This technique has to be one of the preferences for a type observe, if there may be a few previous information regarding the statistics allocation. The K-nearest neighbour technique turned into development suggested the need of creating variance evaluation when dependable estimation of parameters of the probability density are not known.

The technique anticipates the information is located in a characteristic area. Data can be singular or multiple-dimensional vectors. All grooming information consists of a number of vectors related to one and all. The technique is supervised and its required argument is the gap among situations.

The KNN method differentiates the brand new example to segregate with the k-nearest recognised acquaintances and relying to the similar features of the brand new scenario, it'll be in the class very close to the values of its private features with the idea of uniformity heuristic.

The most important trouble of this approach is to set up the dimension of k, seeing that if it is very big, there might be a problem of creating a type depending on maximum information and no longer to the same ones and if the cost is just very small, there might be an absence of precision inside the type due for not having sufficient selected facts as contrast examples.

To alleviate this trouble, numerous tactics to the technique had been suggested: to set the price of k.

The one maximum applicable gain of the KNN technique is that it is able to modify remarkably its segregation outcomes, without altering its definition, best enhancing the calculations used to find the space. The calculation should be decided on in line with the issue that wishes to be rectified. The massive benefit for being capable of trade marks is that can acquire absolutely specific effects and the overall algorithm of the idea does not exchange.

Clustering: It is the most famous unsupervised studying techniques, in constructing a list of abstract objects in classes with same kind of objects. An appropriate clustering will give rise to companies of excessive standards with excessive similarities intra-class and low similarities among classes. It can be used as a non-dependent device to higher the statistics allocation or to serve as a p

While segregating the Twitter feeds' news, a feature selection the usage of divisive method changed into completed. To compare the powerfulness of the technique, they used classifiers with Support Vector Machine, Naive Bayes, and DTs and feature acquired higher end output over selection of other features being compared.

There has existed an upgrading hobby in the use of metamorphosising computation strategies to find solution to the mission of characteristic selection. The authors put forward a feature choice way based totally on Particle swarm optimization. Two fitted capabilities had been described which includes defect rate of the classifier and functions sizes. Correction outcomes confirmed that the present operators made the set of instruction go through the solution area greater powerfully.

At the utilized ACO to discover the answer area of all parts of sets of capabilities. The selected feature subset is calculated through a correction function with the ascending decreased function set and the class output. The result of the characteristic selection set of rules is used as the implementation set in schooling the sorter. The output of experiment conducted indicates that the current method performed filter-based characteristic choosing approach which include IG and CHI square. Between a different roulette wheel allocation process, their current improvement is trying to break out local minimum value by neglecting mating of precise number of members and using more different features to Genetic Algorithm members. Experimental end output with different datasets confirmed that the current GA improvement performed different GA variants with same kind of computational cost.

4. Objectives of Study

The goal of this evaluation is to work with machine learning to build a few prediction models. Some comparisons will be made among models' performance and we are able to have the ability to pick out the satisfactory one. We will look at here the alcohol consumption in relation with the grades of students. The goal is to predict the third grade of the year for a student. The anticipated grade is meant to be among zero and 20. Here, we can be aware that we can also use paradigms to deal with this problem. We can recall that the grade we will predict is a non-stop value; here, we can use SVM and DT algorithm for prediction. Here, we also taken into consideration genetic set of rules-based feature selection version after which apply machine learning for efficient prediction.

5. Existing System

Two people named by Quadri and Kalyankar implemented the strategies using Decision Tree Algorithm to dig up the failures functions for the scholastic achievements. They have used a measure in their research to analyse the student's scholastic performance through cumulative grade factor average (CGPA).

Disadvantages

When there are too many features to handle, the algorithm is in-efficient.

To expect behaviour of college students, two people Yadav and Pal have used a type approach in the technology of mining .Reason behind this observation is to analyse the college students those may additionally pass or fail or locate main elements which have an effect on student behaviour to assist enhanced deprived college students or college students with getting to know disabilities.

Disadvantages

This looks at no longer considered factors affecting the college students performance, while in our work, we considered college students outgoing time, alcohol consumption etc.

Hamoud applied a lot of DT algorithms to categorise also to analyse the student activity. Handling the students is turned out as various factors and examining the students, the fine set of rules need to be located for the students information and calculate it.

Disadvantages

Feature selection is done, accuracy of the look at is less.
The statistic mining technique (example: J48 selection tree technique) is used by Kostopoulo et al, to examine the student's behaviour which are affecting in their studies. The result of the analysis display that the second student's grade appears as an essential characteristic of the student which is affecting their conduct.

Disadvantages

Lets have a look at considered the second function written undertaking as important, whereas more than one characteristic must be used to get greater accuracy.

Two datasets were selected by Fonseca et al which were on-campus datasets and distance schooling. There were several functions covered in student dataset based on their daily work. This analysis uses academic records mining technique (SVM, DT) expect those students that may additionally decline the student's bypass failure.

Disadvantages

The dataset has now not taken into consideration any distraction factors of students.

6. Proposed System

There are still researches open on looking for the key functions that have an effect on the student pass rates, attempting to revise the efficiency of students pass rates

analysis. Proposed device used DT and SVM algorithms to classify college students' grades and analyse the student pass rates. So as to achieve the crucial capabilities, progressed Genetic set of rules are used on student datasets. As the student dataset has many features, there is a need for a stepped forward genetic algorithm to handle characteristic selection problem. In proposed model, we goal to achieve better accuracy on predictions.
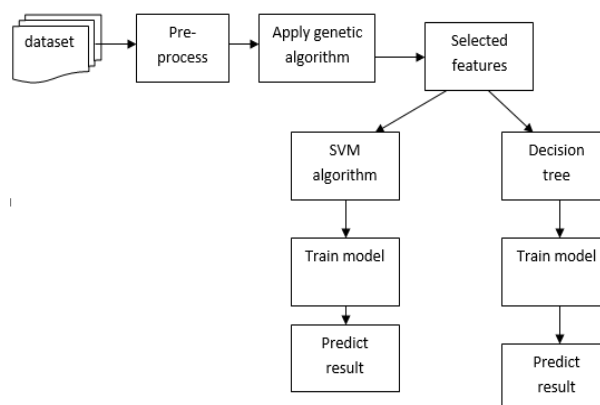
7. System Architecture

Figure 1: Overall Architecture

The above figure represents overall architecture of the system. The genetic algorithm is applied for feature selection process. The machine learning such as SVM and decision tree classifiers are applied. The student grades are predicted through this process and compared to arrive the accuracy.

8. Implementation Methodology

Dataset Details

The dataset downloaded from kaggle.com with following attributes as features for student grade prediction.

The attribute Dalc represents the workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

The attribute Fedu represents the father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)

The attribute G1 represents the first period grade (numeric: from 0 to 20)

The attribute G2 represents the second period grade (numeric: from 0 to 20)

The attribute G3 represents the final grade (numeric: from 0 to 20, output target)

The attribute Medu represents the mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)

The attribute Walc represents weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

The attribute absences represent the number of school absences (numeric: from 0 to 93)

The attribute age represents the student's age (numeric: from 15 to 22)

The attribute failures represent the number of past class failures (numeric: n if 1<=n<3, else 4)

The attributes famrel represents the quality of family relationships (numeric: from 1 - very bad to 5 - excellent) freetime - free time after school (numeric: from 1 - very low to 5 - very high)

The attribute goout represents the student going out with friends (numeric: from 1 - very low to 5 - very high) health - current health status (numeric: from 1 - very bad to 5 - very good)

The attribute studytime represents the weekly study time of a student(numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)

The attribute traveltime represents the home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)

Dataset Pre-Processing

We have taken multiple attribute in our case study, dataset 16 features/ attributes are taken for study. Pre-processing of dataset is done for converting the string attributes to numerals and missing data records are dropped. The pre-processed data is stored in "dataset.csv" file, which is given as input for machine learning models.

Genetic algorithm

In our case, every character in the populace represents a predictive version. The range of genes is the total quantity of functions in the information set. Genes right here are binary values, and represent the inclusion or now not of particular functions inside the version. The number of individuals, or populace size, should be chosen for each application. Usually, that is set to be 10N, being N the wide variety of capabilities.

Step 1: Initialization
The fist step is to create and initialize the individuals in the populace.

Step 2: Fitness project
Once we've got generated and initialized the populace, we want to assign the fitness to each individual. To evaluate the fitness, we want to teach the predictive version with the training facts, and then examine its selection blunders with the choice records.

Step 3: Selection
After fitness challenge has been performed, the selection operator chooses the people with the intention to recombine for the next generation. choice operator selects

the people consistent with their health level. The range of selected people is N/2, being N the population size.

Step 4: Crossover
Once the choice operator has selected half of the population, the crossover operator recombines the selected individuals to generate a new population.

Step 5: Mutation
The crossover operator can generate off springs which are very similar to the parents. This might motive a new era with low diversity.
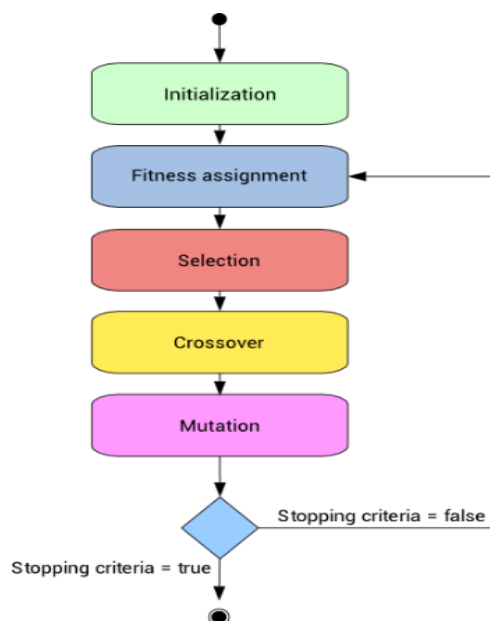

Figure 2: Genetic Algorithm Steps

Decision Tree

The Decision tree algorithm is a type of supervised mastering algorithm that is by and large used in classification problems. It works for both unconditional and continuous input and output variables. In this algorithm, we divide the sample into two or more similar sets (or sub-populations) based on most important divider / differentiator in input variables. In decision tree algorithm the internal node represents a test on the attribute, branch represents the outcome and leaf represents the decision made after calculating the attribute.

The Decision Tree algorithm works in following manner:

1) Place the best attribute of the dataset at the root of the tree.

2) Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.

3) Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

In decision trees, for predicting a class label for a record we start from the root of the tree. Then compare the values of the root attribute with record's attribute. On

the basis of comparison, follow the branch corresponding to that value and jump to the next node.
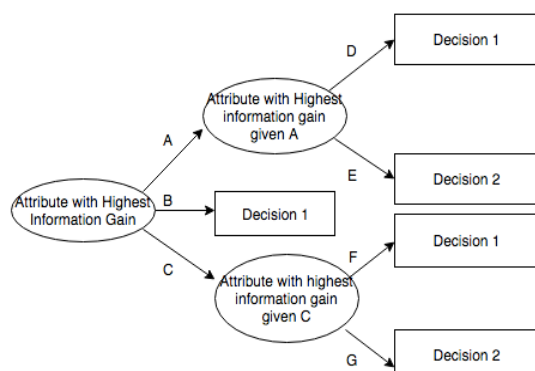
Support Vector Machine Algorithm

The algorithm named Support Vector Machine is a supervised gaining knowledge algorithm. This algorithm models the facts into k categories, by performing type and making an N-dimensional hyper plane. These representations are very much like neural networks. Consider a dataset of N dimensions. The SVM marks the instruction data into an N dimensioned space. The instruction facts factors are then divided into k different regions depending on their labels by hyper-planes of n different dimensions. After the testing phase is complete, the test points are marked in the equal N dimensioned plane. Depending on which areas the points are located, they are correctly categorized on that region.
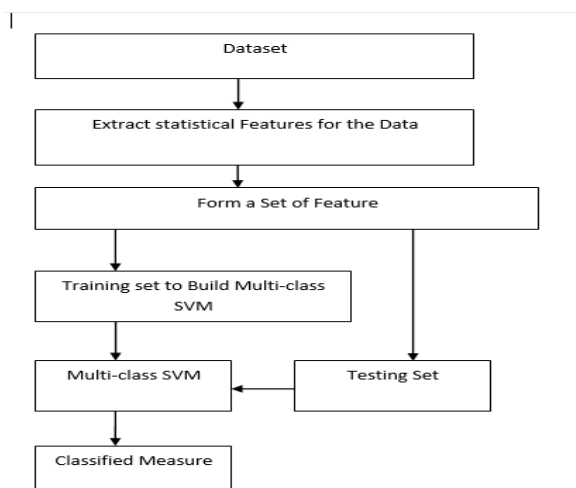
9. Results

The motive of this study is to provide help to students who have greater difficulties in their studies and to predict students who can graduate through data mining techniques. This prediction system enables in retaining a high common of the general students' grades. This is achieved because the bad students are diagnosed and are helped to improve.

10. Conclusions

Educational data mining (EDM) is one among the maximum essential regions of research today. Considering almost 46 papers, we observed that, predicting student's overall performance is a useful subject matter of research which helps educator, academician, coverage makers and management for improving the teaching and studying system in the course. The student's attributes taken by using unique researchers are categories into academic, psychological, social behavior, own family attributes, and personal attributes. In maximum of the reviewed papers, researchers used CGPA and inner and external evaluation marks for their prediction. While reading prediction algorithms we identified that Classification, Clustering, Linear regression, Association Rule mining are used for prediction. But in most of the cases, Classification algorithms are frequently used for making student's educational prediction. Under the Classification algorithms, DT, NN, NB, K-NN and SVM are frequently used by many researchers. In the end, we approve to accomplish our studies on student's academic prediction in our educational gadget which facilitates students, educators, control and policymaker to enhance the educational system by using continues monitoring of students.

References

[1] Mukesh Kumar, Yass Khudheir Salal, "Systematic Review of Predicting Student's Performance in Academics". – March 2019.

[2] Bashir Khan, Malik Sikandar Hayat Khiyal, Muhammad Daud Khattak, "Final Grade Prediction of Secondary School Student using Decision Tree". – April 2015

[3] D. M. S. Anupama Kumar, "Appraising the importance of self-regulated studying in better training the use of neural networks", International Journal of Engineering Research and Development Volume 1 (Issue 1) – September 2012.

[4] K. Bunkar, U. K. Singh, B. Pandya, R. Bunkar, Data mining: Prediction for performance improvement of graduate students the usage of classification, in Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, IEEE – September 2012

[5] Ricardo Mendes And Joao P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications", IEEE, 2017.

[6] Edin Osmanbegović and Mirza Suljic, Data Mining Approach For Predicting Student Performance, Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

[7]   V. Ramesh, P. Parkavi, K. Ramar, Predicting student performance: a statistical and data mining approach, International Journal of Computer Applications 63 (8) (2013).

[8]   Julia Rudnitckaia. Process Mining: Data Science in Action. Berlin, Germany: Springer; 2016.

[9]   A. Boarding, C. Romero, R. Cerezo, M. S´anchez-Santill´an Clustering for improving educational process mining, in Proceedings of the Fourth International Conference on Learning Analytics And Knowledge, ACM, 2014.

[10]   A Kadambande, S Thakur et. al., 'Predict Student Performance by Utilizing Data Mining Technique and Support Vector Machine", International Research Journal of Engineering and Technology, e-ISSN: 2395 -0056, 2017.

[11]   K. B. Eashwar, R. Venkatesan, "Student Performance Prediction Using SVM", International Journal of Mechanical Engineering and Technology (IJMET), Volume 8, Issue 11, November 2017.