BOHR International Journal of Smart Computing and Information Technology
2023, Vol. 4, No. 1, pp. 39–49
DOI: 10.54646/bijscit.2023.35
www.bohrpub.com

METHODS

Data mining approach to predict academic performance of students

Partha Ghosh*, Reet Roy, Souhardya Mandal, Manthan Chowdhary and Subhajit Bokshi

Department of Computer Science and Engineering, Government College of Engineering & Ceramic Technology, Beleghata, Kolkata

***Correspondence:**
Partha Ghosh,
parth_ghos@rediffmail.com

Received: 20 May 2023; **Accepted:** 12 June 2023; **Published:** 20 July 2023

Powerful data mining techniques are available in a variety of educational fields. Educational research is advancing rapidly due to the vast amount of student data that can be used to create insightful patterns related to student learning. Educational data mining is a tool that helps universities assess and identify student performance. Well-known classification techniques have been widely used to determine student success in data mining. A decisive and growing exploration area in educational data mining (EDM) is predicting student academic performance. This area uses data mining and automaton learning approaches to extract data from education repositories. According to relevant research, there are several academic performance prediction methods aimed at improving administrative and teaching staff in academic institutions. In the put-forwarded approach, the collected data set is preprocessed to ensure data quality and labeled student education data is used to apply ANN classifiers, support vector classifiers, random forests, and DT Compute and train a classifier. The achievement of the four classifications is measured by accuracy value, receiver operating curve (ROC), F1 score, and confusion matrix scored by each model. Finally, we found that the top three algorithmic models had an accuracy of 86–95%, an F1 score of 85–95%, and an average area under ROC curve of OVA of 98–99.6%.

Keywords: predictive analysis, KNN, SVM, random forest, DT classifier, students' academic performance

1. Introduction

Due to the wealth of information in educational databases, it is difficult to anticipate student success. Malaysia today does not have a mechanism to research and monitor student progress and achievement. There are two main reasons for this. First and foremost, more research needs to be done on existing prediction techniques before determining the best one for predicting academic performance in Malaysian institutions. A second reason is the lack of research on the context-specific Malaysian factors that influence student performance in different courses. Therefore, it is proposed to improve student performance by conducting a thorough literature review on data mining strategies for predicting student success.

Higher education management is very concerned with predicting a student's success. The purpose of this work is to discover the variables affecting students' performance on final exams and to create a suitable data mining method to forecast students' grades in order to provide timely and pertinent warnings to students who are at danger. In the current study, a survey-based experimental approach was used to create a database constructed using primary and secondary sources. Hypothesis test results showed that parental occupation had a significant effect on performance prediction, while school type had little effect on child performance. The results of this study will help educational institutions identify at-risk students and provide better additional education to vulnerable students.

1.1. Motivation

Analyzing educational data is not a new practice, but recent developments in educational technology, such as the development of methods to analyze the vast amounts of data generated in education, have led to new practices. This enthusiasm led to a series of EDM workshops as part of a number of international academic conferences from 2000 to 2007. The first EDM research conference, now held annually, was founded in 2008 by a group of scholars in Montreal, Quebec, Canada.

EDM researchers launched the *Journal of Educational Data Mining* in 2009 as a platform to share and publish their findings as the field grew in popularity. To connect and advance the discipline, EDM researchers founded the International Educational Data Mining Consortium in 2011.

Public education data sets have made education data mining more accessible and convenient, fueling its growth, since the launch of public education data sets in 2008, including the Pittsburgh Academic Science Center (PSLC) and the National Center for Education Statistics (NCES) Database.

Today, EDM is becoming increasingly popular to achieve the following goals:

(a) Predict future learning behavior of students – This goal can be achieved by developing student models that incorporate learner characteristics, such as specific information about learner knowledge, behavior, and motivation to learn. It also assesses learners' overall learning satisfaction and user experience.
(b) Identifying or enhancing domain models – Discovering new models and improving old models are made possible by a multitude of EDM techniques and applications. Examples of this include designing a lesson sequence that best fits a student's learning style or displaying lesson material to engage the student.
(c) Examine the results of educational support by learning systems.
(d) Create and use student models, EDM research, technology, and software to enhance learning and learners' scientific understanding. These goals serve as the basis for investing time and effort in development and are the driving force behind this work.

2. Literature survey

This section examines a number of classification techniques traditionally used by scholars to predict student academic performance. Here, we mainly discuss his two methods. The first is a hybrid data mining strategy (1) developed by Bindhia K Francis and Suvanam Sasidha rBabu, and the second is development of Hybrid Data Mining Models (HEDM) for educational purposes by V Ganesh Karthikeyan, P Tangaraj, and S Karthik (2).

The first article used a 2-layer approach for classifying: first, it fed the data to the supervised learning classifiers and then used an unsupervised learning clustering algorithm (K-means clustering plus majority voting) in order to increase the accuracy. The data set includes several features such as the Demographic, Academic, Behavioral, and Extra features. These data have been taken with regard to the interaction of the student with the online website of 2 institutions from Kerala. The researchers took a subset of different features and used 4 supervised classifying algorithms: Naive Baye's, Decision Tree (Iterative Dichotomies 3 algorithm), Support Vector Machine, and Neural Network (Multilayer Perceptron) in first layer and then the K-means clustering algorithm that groups into clusters showing high, medium, and low performance. Therefore, they used four performance measures of accuracy, recall, precision, and F1 score to compare the achievement of different classifiers on different criteria. Through various combinations of feature and machine learning models, they finally arrived at the conclusion that Academic and Behavioral features of students contributed largely to the grade received by the student and Decision Tree was the most accurate model for predicting grades with an accuracy of 75%

The model presented in the second article combines the strengths of the J48 classifier with Naive Baye's classification approach to accurately classify student performance and draw inferences. The information from two public schools in the Alentejo area of Portugal for the academic year 2005–2006 was utilized in this article. In this study, Naive Baye was used to classify the data into pass and fail lists and then a J48 decision tree classifier was employed to further classify the data into "pending list" and "good list., suggests identifying student dispositions in categories such as "average list." Precision, Recall, F1 Score, and Accuracy are used as the four performance criteria to assess the effectiveness of this technique. The final accuracy achieved by Naive Baye's + J48 Decision Tree classifier was 98.6% using the WEKA tool.

3. The proposed work

3.1. Objective and goals

The purpose of this project is to take various features of students as input features and predict the grade of the student based on this factor using various ML classification techniques to classify students and help the institute to take necessary actions and effort to improve the performance of students of each category and support them to success.

– The core objective of the suggested study is to put several ML classifiers into practice for estimating learner progress.

- We want to analyze our model's performance using various performance metrics.
- In-depth analysis considers the impact of all classifiers to determine the best classifier to predict student performance.

3.2. Data set

- The information utilized in this lesson was gathered from two community schools in Portugal's Alentejo area in the academic year 2005–2006.
- The database was built from following sources:
  - Paper-based reports for school
  - Characteristics (the absence and three-period grades)
  - Questionnaires, which are used to supplement the earlier data
- Characteristic Information:

Both the student-mat.csv (for a math course) and student-por.csv (for a Portuguese language course) databases include the same attributes:

1. School (binary: "GP" for Gabriel Pereira or "MS" for Mousinho da Silveira)
2. Student's gender (binary: "F" – female or "M" – male)
3. Age of the student, 3 (numerical: 15–22)
4. Address – the address of the student residence (in the city or in the countryside, in binary format)
5. Famsize – family size (binary: "GT3" – higher than 3 or "LE3" – fewer than 3)
6 Pstatus: the parents' status of cohabitation (binary: "T" for "living together" or "A" for "apartment")
7. Medu – maternity education (number: 0 – none, 1 – primary school (grade), 2 – "grades 5–9", 3 – "secondary education" or "further education")
8 Fedu – father's education (count: 0 – none, 1 – primary education (grade 4), 2 – grades 5–9, 3 – secondary, or 4 –continuing education)
9. Mjob stands for mother's job and means "education," "medical care," "public service" (police, government, etc.), "home," or "other"
10. Fjob – father's occupation (nominative: "education," "medical," "public service" (government, police, etc.), "household," or "other")
11. Proximity to this school choice justification (nominally: "home," school "call," desired course, or "other")
12. Parent – the person responsible for the student (nominative: "mother,, "father," or "other")
13. Travel time from house to school (in minutes: 1: 15 min, 2: 15–30 min, 3: 1 h, or 4: more than an hour)
14. Studytime – total amount of time spent studying each week (in hours: 0: 1, 1: 2, 2: 5, 3: 5, or 4: > 10)
15. Failures – this is the total number of failures in the previous class (number: n if 1 = n3, otherwise 4)
16. schoolsup – continuing education support (binary: yes or no)
17. Family Education Support Program (famsup) (binary: yes or no)
18. Paid – additional paid tuition (binary: yes or no) for course subjects (mathematics or Portuguese)
19. After-school activities (binary: yes or no)
20. Nursery – attended a preschool (binary: yes or no)
21. Higher – intends to further their study (binary: yes or no)
22. Access to the Internet at home (number 22) (yes or no in binary)
23. Romantic – engaged in a romantic partnership (binary: yes or no)
24. famrel – strength of kinship bonds (numeric: from 1 – very bad to 5 – excellent)
25. Especially after school, free time (count: 1 – very low to 5 – very high)
26. Fun with friends on the go (number: 1 – very low to 5 – very high)
27. Darc – alcohol consumption at work (number: 1 – very little to 5 – very high)
28. Weekend alcohol consumption statistics (number: 1 – very low to 5 – very high)
29. Health – current state of health (number: 1 – very bad to 5 – very good)
30. Absences – total number of days absent from school (number: 0–93). # These numbers relate to math or Portuguese classes.
31. G1 – first-grade point (numeric: 0–20)
32. G2 is the second-class (numeric: 0–20)
33. G3 – final grade (numeric: 0–20, output target)

3.3. Methodology

The proposed model (**Figure 1**) is structured for the analysis and evaluation of PIDD. In our model, we first import the specified data set. Then, we use different data visualization techniques:

- Histogram (to check count of student receiving final grade for each age)
- Count plot (to compare the count of students with different attributes)
- Box plot (to check if any outliers are present in the data)

Based on the nature of data, we perform data preprocessing by removing the outliers in the data. Next, we divide the data set into test and training data sets. Then we train the data set individually on 4 different classification algorithms:

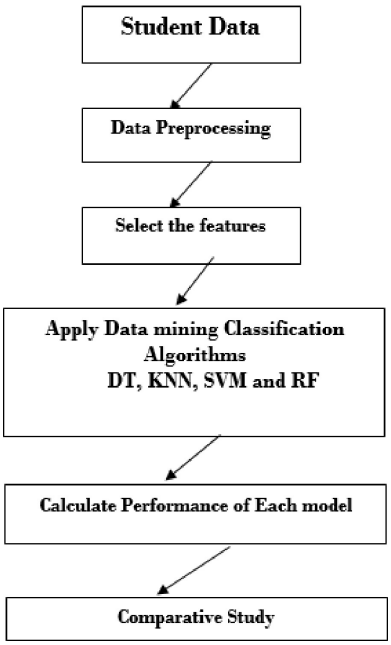

FIGURE 1 | Data flow diagram of our model.

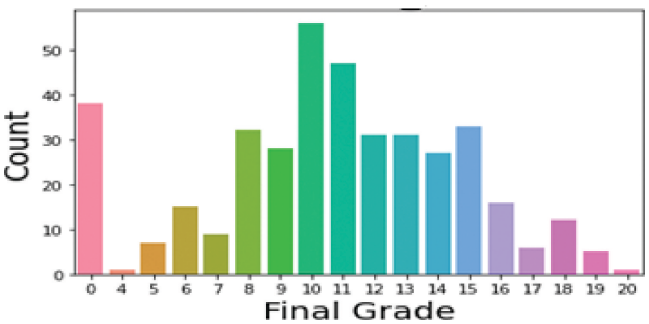

FIGURE 2 | Count plot showing Students final grade.

- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Decision Tree Classifier (DT)
- Random Forest Classifier (RF)

We evaluate the performance of each of the algorithms. The performance metrics used are as follows:

- Accuracy measure
- ROC score
- F1 score

Finally, we compare the analysis based on accuracy and obtain the final results.

3.4. Data preprocessing

3.4.1. An overview of the data set

Figure 2 represents the final grade (G3) of data set vs. count of students in each age group for whom the data were collected. As can be seen in the figure, there is a large group of students whose age is evaluated as 0. Since these students have no age record, they were replaced as 0 in order to eliminate null.

Figure 3 shows whether or not any attribute of the data set contains any null value. The uniform cyan color confirms that none of the attributes has any null values.

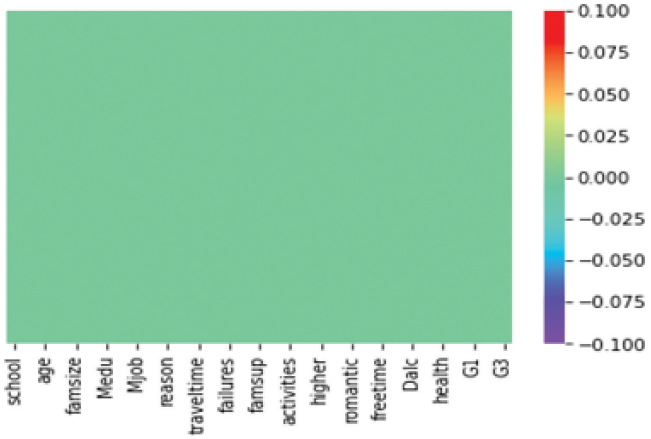

FIGURE 3 | Graph to visualize null values.

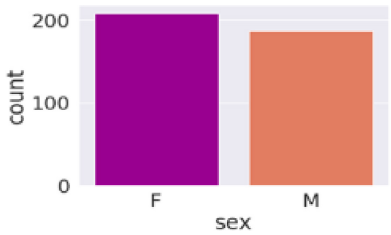

FIGURE 4 | Count plot showing male/female count.

FIGURE 5 | Count plot showing rural and urban students.

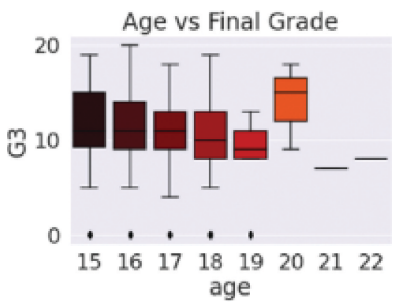

FIGURE 6 | Boxplot showing final grade (G3) vs. age.

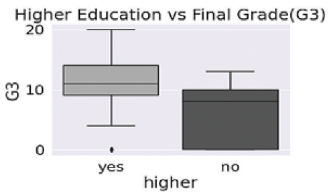

FIGURE 7 | Boxplot showing final grade (G3) vs. higher education.

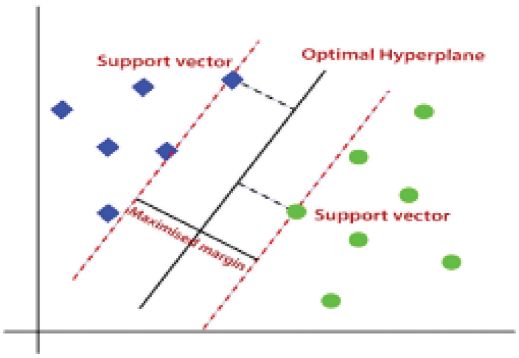

FIGURE 8 | Working of SVM (4).

Figure 4 compares the count of female students and male on whom the data set was made.

The number of female candidate is 208 and that of male candidate is 187.

Figure 5 compares the residential status (rural or urban origin) and the count of students.

The plot shows the number of urban students as 307 and that of rural students as 88.

Figure 6 is a boxplot showing the range of final grade (G3) received by students of a particular age and compared with other age groups.

Figure 7 is a boxplot showing the range of final grade (G3) received by students who opted for higher education vs. those who did not opt for higher education.

Once the data set is collected, preprocessing techniques are employed to enhance the data set's quality. Data cleansing, feature selection, data transformation, and data reduction are all components of data preprocessing and are considered important phases of the knowledge discovery process. The process of transforming real-world information into a form that can be used by a particular data mining technique is called data preprocessing, and this is performed before algorithms are used.

The data set used by us is a rather clean data set that has no null values, as indicated by **Figure 3**. Therefore, there is no need to filter null values and either delete or replace them by 0. However, the data set contains various text values for attributes, such as Parent's Cohabitation status, Family size, and School. Most of these attributes have a yes or no binary answer, which we mapped to 1 and 0, respectively. However, for attributes like Family size, we split the attribute to the corresponding number of categories and populate the correct category with 1 and others with 0.

3.5. Algorithms

In this work, tests were performed using classifiers such as SVMs, ANNs, decision trees, and random forests. These classifiers evaluate the metrics of the data set. Pattern recognition and nonlinear function estimation problems are solved with SVM. Training data are nonlinearly represented in a high-dimensional feature space using support vector machines. This helps to build a separating hyper-plane with the widest possible margins, resulting in a nonlinear decision boundary in the input space. A quadratic programming problem with a global solution provides a support vector machine solution.

3.5.1. Decision tree algorithm

By inferring straightforward decision rules from the data, decision trees build a model or tree that anticipates the value of a target variable (in this case, passed). For instance, based on our data set, children with a single guardian are more likely to drop out of school. If one of the guardian variables is zero in this situation, a straightforward decision tree with one node will designate the students as dropouts (3).

3.5.1.1. Advantages and disadvantages of decision tree. It is easily imaginable. With many other algorithms, this is not achievable. For instance, with SVM, it is practically hard for us to imagine data in a 10-dimensional space. It is fast since they have logarithmic run times. Although we are already preprocessing the data in our instances, another benefit of decision trees is that they can employ categorical data even without formatting.

Decision tree suffers from overfitting, and the F1 scores for both training and test data sets differ significantly. It does not perform well if one class dominates; however in such situation, I believe it to be 30% vs. 70%, making them fairly equal.

In light of this discussion, we attempted using decision trees for this specific issue because they are frequently utilized for classification, simple to comprehend and depict, and appropriate for our data, where the majority of the variables are categorical.

3.5.2. SVM

SVMs are a group of classifiers that look for the best linear separator between two classes. Although in fact most data would not be linearly separable, SVM uses a kernel method to add additional features to the data by combining existing features in different ways. For instance, a group of data points with x and y attributes might not be linearly separable in two-dimensional space, but they might be in three-dimensional

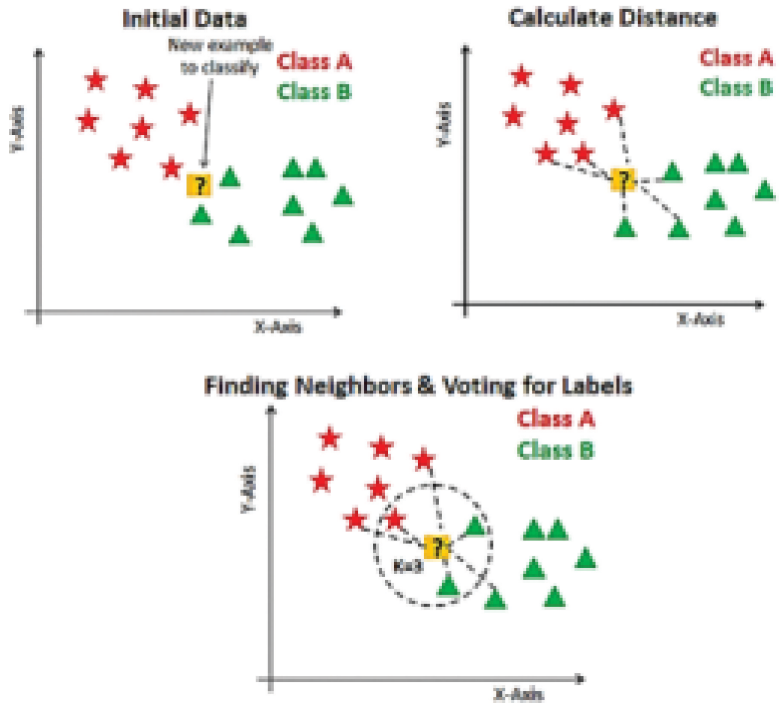

FIGURE 9 | Working of KNN (5).

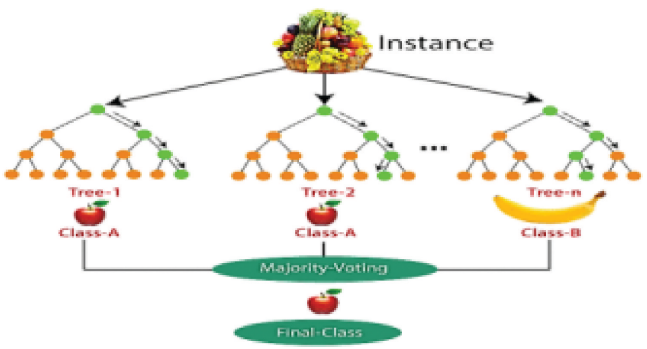

FIGURE 10 | Working of random forest classifier (6).

space where $x^2 + y^2$ is the third variable. The working principle of SVM is shown in **Figure 8**.

3.5.2.1. SVM: Benefits and drawbacks. It offers two benefits. One reason they are known as support vector is that they only employ a small subset of data to train the model. When we have a lot of data, this strategy is more memory-efficient. Another benefit is its robustness in situations where the number of data points is on the order of the number of features. It experiences the dimensionality curse less severely.

It has disadvantages too. It may be difficult to explain how they work. The concept of high dimensionality may be difficult to grasp. SVM also does not work when the number of features is much bigger than data, but this is rarely the case. I believe it is suitable in this case because it is a widely used classification algorithm that performs best right out of the box.

3.5.3. K-Nearest neighbors algorithm

Classification based on nearest neighbors' algorithm is one where no explicit model for the data is constructed. Classifier simply stores the data and, whenever a new point comes, classifier assigns the new point to a class based on its closest k neighbors. The working of KNN is shown in **Figure 9**.

3.5.3.1. Advantages and disadvantages of K-Nearest neighbors algorithm. Here training times are short. There is essentially no training and no model. Therefore, it is easy to update KNN as the new data come. It will be very easy to explaining this model in Laymen's terms.

Predicting each new data point takes a long time. For each data point, we have to calculate its nearest neighbors. Another disadvantage is the curse of dimensionality. As the dimensions of the feature space increase, each data point is going to represent a bigger volume of space for which it may not be representative. We wanted to try KNN, as it is a very simple algorithm that makes no assumption about the data.

3.5.4. Random forest algorithm

To generate a random forest, combine N decision trees. Then, in the subsequent stage, we make guesses for each tree in the first stage. The working of random forest classifier is shown in **Figure 10**.

3.5.4.1. Advantages and disadvantages of random forest algorithm. Classification and regression tasks can be performed on random forests. It can handle large data sets

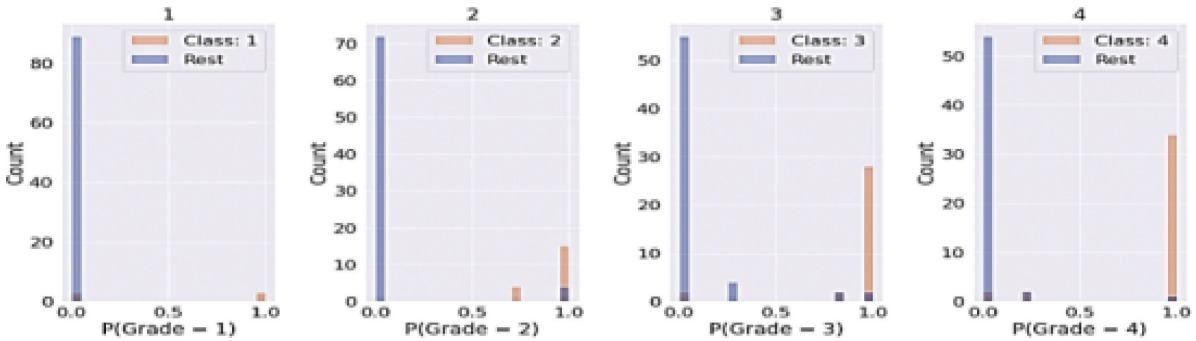

FIGURE 11 | Count plot for comparison of probability of outcome as grade X vs. All for decision tree.

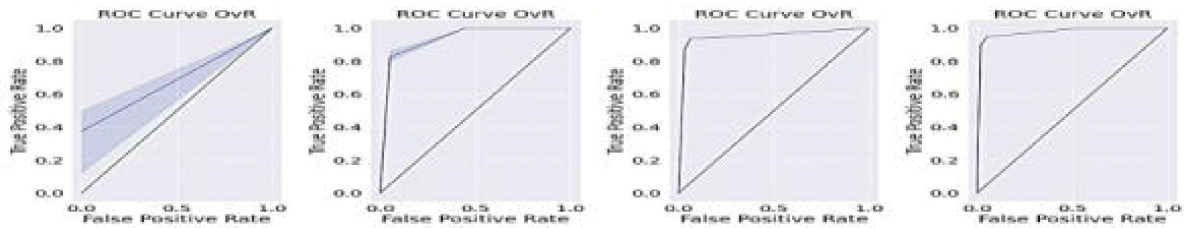

FIGURE 12 | ROC OVR of the 4 grades using decision tree.

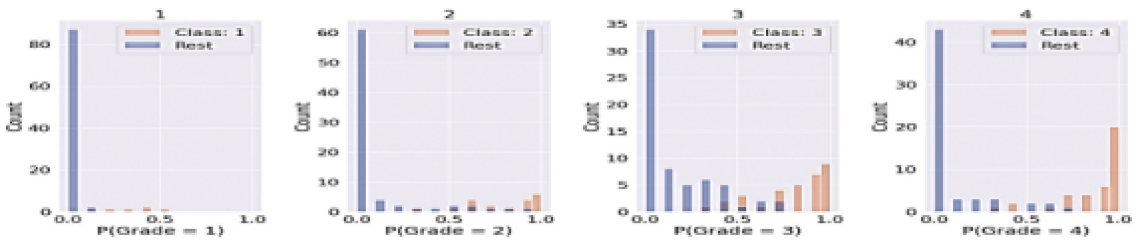

FIGURE 13 | Count plot for comparison of probability of outcome as grade X vs. all for KNN.

with high dimensions. It avoids overfitting problems and improves model accuracy.

Regression and classification problems may be both accomplished using random forests, but they are not well suited for applications that require regression.

3.6. Performance metrics used

This is an illustration of accuracy metrics including accuracy, F-measure, and ROC (receiver operating curve) examined for several models.

Accuracy (A) – It is the proportion of accurate predictions to all input samples.

Accuracy (A): $(TN + TP)/(TN + TP + FP + FN)$ (i)

F1 measure between recall and precision – It is the harmonic mean. Better model performance is indicated by a higher F1 score.

F1 measure (F): $(2 * P * R) / (P + R)$ (ii)

Plotting the ratio of true positives (TP) to false positives (FP) is known as ROC. Measurement of test utility is helpful.

4. Experimental results and discussion

We have applied all the above-mentioned classifiers on the data set one by one and have calculated the performance metrics.

i) Decision Tree

Confusion matrix for test and train data set

```
{'max_depth': 4, 'min_samples_split': 8}
[1 2 3 4]
0.9016541353383457
0.9052631578947369
Predicting labels using GridSearchCV...
Done!
Prediction time (secs): 0.002453804016113
[[  7   0   0   0]
 [  0  59   2   0]
 [  0   3 111   3]
 [  0   0   3 112]]
Train: (0.9633224001767915, 0.9633333333333334)
Predicting labels using GridSearchCV...
Done!
Prediction time (secs): 0.002058744430542
[[ 3  3  0  0]
 [ 0 19  0  0]
 [ 0  1 29  2]
 [ 0  0  3 35]]
Test: (0.9016541353383457, 0.9052631578947369)
```

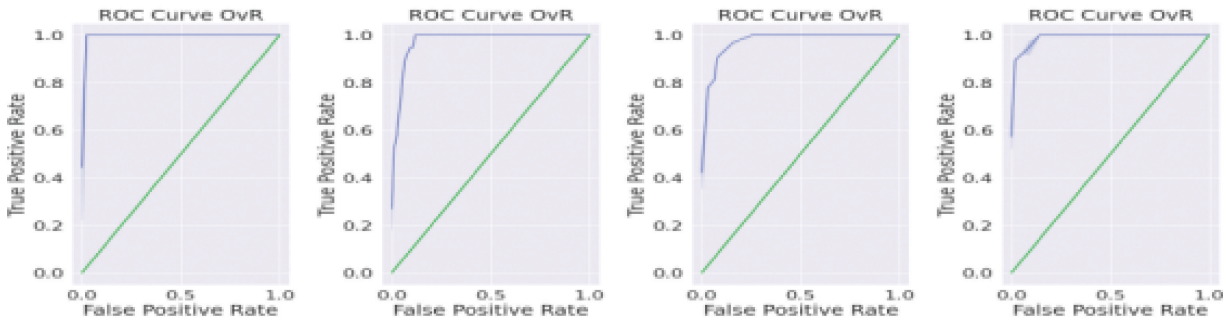

FIGURE 14 | ROC OVR of the 4 grades using KNN.

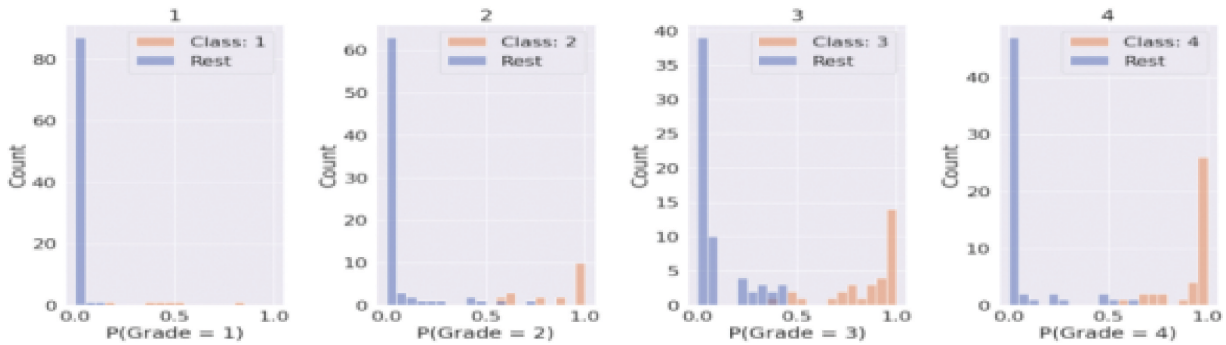

FIGURE 15 | Count plot for comparison of probability of outcome as grade X vs. all for SVC.

```
1 ROC AUC OvR: 0.7500
2 ROC AUC OvR: 0.9681
3 ROC AUC OvR: 0.9479
4 ROC AUC OvR: 0.9746
average ROC AUC OvR: 0.9102
```

4.1. Conclusion from the train data's confusion matrix

Seven children in group 1 were correctly classified as such.

Of the 61 children, 59 were correctly assigned to Group 2, while the other 2 were incorrectly assigned to Group 3.

Of the 117 third graders, 29 were correctly identified, 1 was misclassified as a second grader, and the other 2 were misclassified as fourth graders.

Of the 115 grade 4 students, 112 were correctly identified as grade 4 and 3 were incorrectly identified as grade 3.

4.2. Conclusion from the test data's confusion matrix

Three of the six grade 1 kids in the test group were accurately categorized as being in grade 1, while the other three were categorized as being in grade 2.

All 19 of the grade 2 kids were appropriately categorized as being in grade 2.

The 30 third-grade pupils were divided into 29 who were accurately identified as such, one who was misclassified as a second grader, and three who were misclassified as fourth graders.

A total of 38 children were in grade 4, of whom 35 were accurately identified as such, while 3 were misidentified as being in grade 3. Count plot for comparison of probability of outcome as grade X vs. all for decision is shown in **Figure 11**. And ROC OVR (One vs. Rest) of the 4 grades using decision tree is shown in **Figure 12**.

ii) KNN

```
{'n_neighbors': 10}
0.8486448679839133
0.8631578947368421
Predicting labels using GridSearchCV...
Done!
Prediction time (secs): 0.041543483734131
[[  4   3   0   0]
 [  0  54   7   0]
 [  0   2 107   8]
 [  0   0   9 106]]
Train: (0.9025959706232634, 0.9033333333333333)
Predicting labels using GridSearchCV...
Done!
Prediction time (secs): 0.012665748596191
[[ 1  5  0  0]
 [ 0 18  1  0]
 [ 0  1 28  3]
 [ 0  0  3 35]]
Test: (0.8486448679839133, 0.8631578947368421)
```

```
1 ROC AUC OvR: 0.9981
2 ROC AUC OvR: 0.9730
3 ROC AUC OvR: 0.9759
4 ROC AUC OvR: 0.9905
average ROC AUC OvR: 0.9844
```

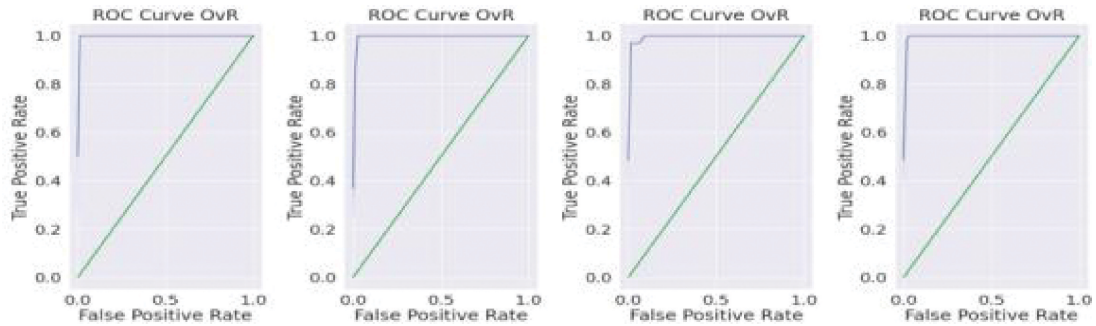

FIGURE 16 | ROC OVR of the 4 grades using SVC.

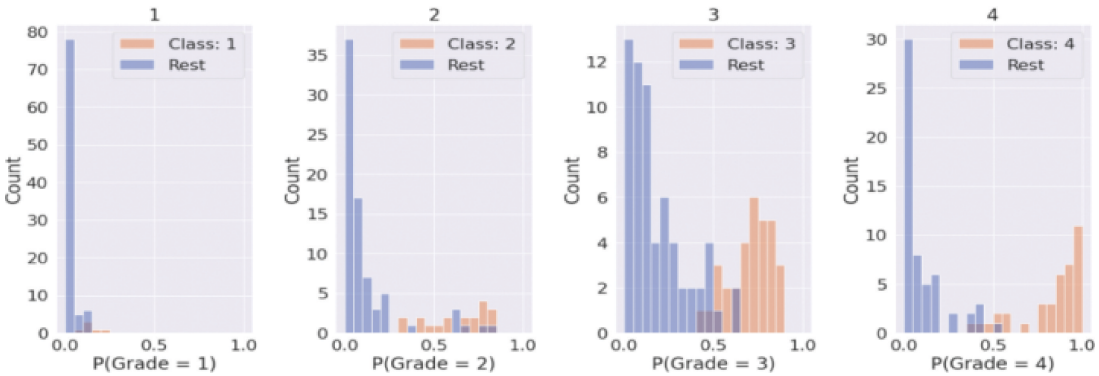

FIGURE 17 | Count plot for comparison of probability of outcome as grade X vs. all for random forest.

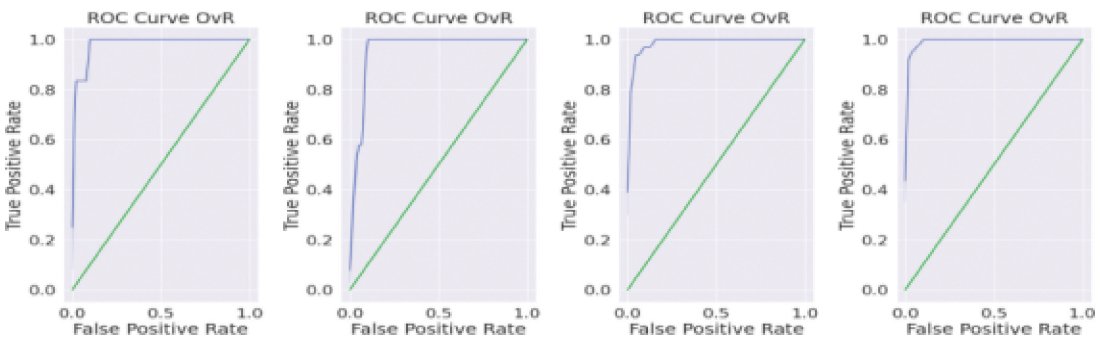

FIGURE 18 | ROC OVR of the 4 grades using random forest.

4.3. Conclusion from the train data's confusion matrix

Four of the seven grade 1 kids in the test group were accurately categorized as being in grade 1, while three were appropriately categorized as being in grade 2.

Of the 61 grade 2 pupils, 54 were accurately categorized as being in that grade, while 7 were misclassified as being in that grade.

Of the 117 third-grade pupils, 107 were accurately identified as such, while 2 were misclassified as second graders and another 8 as fourth graders.

Of the 115 grade 4 pupils, 106 were categorized properly as being in grade 4 and 9 were misclassified as being in grade 3.

TABLE 1 | Comparison between performances of different classifier.

Classifier Name	F1 score	Accuracy	Area under ROC curve
Decision tree	0.90	0.90	0.91
K-Nearest neighbor	0.85	0.86	0.98
SVM	0.95	0.95	0.996
Random forest	0.80	0.83	0.98

4.4. Conclusion from the test data's confusion matrix

One of the six grade 1 pupils in the test group was accurately categorized as in grade 1, while the other five were categorized as in grade 2.

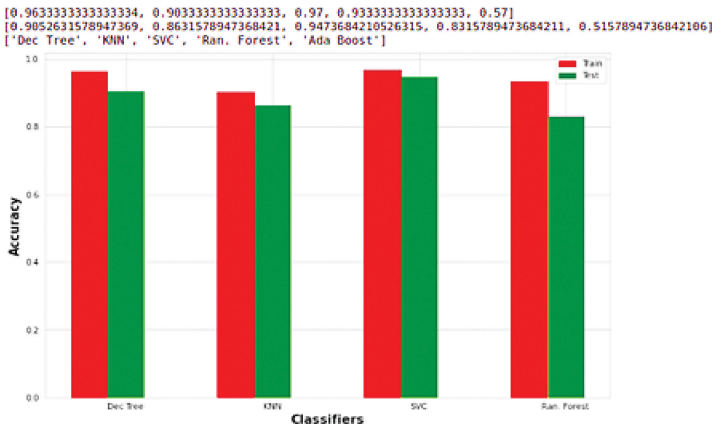

FIGURE 19 | Comparison of performance of different classifiers.

Of the 19 kids in grade 2, 18 were appropriately identified as such, while 1 was mistakenly placed in grade 3.

Of the 30 third-grade pupils, 28 were accurately identified as such, while 1 was misclassified as a second grader and the remaining 3 as a fourth grader.

Of the 38 fourth-grade pupils, 35 were accurately identified as such, while 3 others were misidentified as such. Count plot for comparison of probability of outcome as grade X vs. all for KNN is shown in **Figure 13**. And ROC OVR of the 4 grades using KNN is shown in **Figure 14**.

iii) SVM

Confusion matrix for test and train data set

```
{'C': 0.4, 'gamma': 0.01, 'kernel': 'linear'}
0.9456125424580334
0.9473684210526315
Predicting labels using GridSearchCV...
Done!
Prediction time (secs): 0.003804445266724
[[  7   0   0   0]
 [  0  60   1   0]
 [  0   2 111   4]
 [  0   0   2 113]]
Train: (0.9699320521082506, 0.97)
Predicting labels using GridSearchCV...
Done!
Prediction time (secs): 0.001770257949829
[[ 4  2  0  0]
 [ 0 19  0  0]
 [ 0  0 29  3]
 [ 0  0  0 38]]
Test: (0.9456125424580334, 0.9473684210526315)

1 ROC AUC OvR:  1.0000
2 ROC AUC OvR:  0.9965
3 ROC AUC OvR:  0.9975
4 ROC AUC OvR:  0.9995
average ROC AUC OVR: 0.9984
```

4.5. Conclusion from the train data's confusion matrix

All seven grade 1 pupils in the test group were accurately identified as such.

Of the 61 grade 2 kids, 60 were accurately identified as such, while 1 was misidentified as a grade 3 student.

Of the 117 third-grade pupils, 111 were accurately identified as such, while 2 were misclassified as second graders and another 4 as fourth graders.

Of the 115 grade 4 pupils, 113 were accurately categorized as being in grade 4, while 2 were mistakenly categorized as being in grade 3.

4.6. Conclusion from the test data's confusion matrix

Of the six grade 1 pupils tested, four were accurately categorized as in grade 1, while two were categorized as in grade 2.

All 19 of the grade 2 children were indeed placed in that grade.

Notably, 32 pupils in grade 3 were evaluated, of whom 29 were accurately identified as such, while 3 were misclassified as such.

Out of the 38 grade 4 students, all were correctly classified as grade 4. Count plot for comparison of probability of outcome as grade X vs. all for SVC is shown in **Figure 15**. And ROC OVR of the 4 grades using SVC is shown in **Figure 16**.

iv) Random Forest

```
Random Forest Model Score : 0.9333333333333333 ,
Cross Validation Score : 0.8315789473684211
Predicting labels using RandomForestClassifier...
Done!
Prediction time (secs): 0.007281541824341
[[  0   7   0   0]
 [  0  57   4   0]
 [  0   2 109   6]
 [  0   0   1 114]]
Train: (0.9224865267083394, 0.9333333333333333)
Predicting labels using RandomForestClassifier...
Done!
Prediction time (secs): 0.008446455001831
[[ 0  6  0  0]
 [ 0 15  4  0]
 [ 0  0 28  4]
 [ 0  0  2 36]]
Test: (0.8050361918782972, 0.8315789473684211)
```

```
1 ROC AUC OvR: 0.9813
2 ROC AUC OvR: 0.9557
3 ROC AUC OvR: 0.9881
4 ROC AUC OvR: 0.9952
average ROC AUC OvR: 0.9800
```

4.7. Conclusion from the train data's confusion matrix

All seven grade 1 children in the test data were misclassified as being in grade 2.

Of the 61 grade 2 kids, 57 were accurately identified as such, while 4 were misclassified as grade 3.

Of the 117 third-grade pupils, 109 were accurately identified as such, while 2 were misclassified as second graders and another 6 as fourth graders.

Of the 115 grade 4 pupils, 114 were accurately identified as being in that grade, while 1 kid was misclassified as being in grade three.

4.8. Conclusion from the test data's confusion matrix

All six grade 1 pupils in the test sample were appropriately categorized as grade 2 students.

Of the 19 pupils in grade 2, 15 were accurately identified as such, while 4 were identified as grade 3.

Of the 32 third-grade pupils, 28 were accurately identified as such, while 4 were misclassified as fourth-grade students.

Out of the 38 grade 4 students, 36 were correctly classified as grade 4 while 2 were classified as grade 3. Count plot for comparison of probability of outcome as grade X vs. all for random forest is shown in **Figure 17**. And ROC OVR of the 4 grades using random forest is shown in **Figure 18**.

Performance of each model on test data is mentioned in **Table 1**.

The training set and test set accuracy for each model is presented side by side in **Figure 19**.

In this work, we applied various classification techniques on the student data set to determine the student's final result (i.e., Pass or Fail) and we have found that the SVC can be a reliable model to predict student performance accurately because it shows consistent accuracy, F1 score, and AU-ROC curve performance.

5. Conclusion and future scope

This study suggests using data mining techniques to forecast students' final grades based on past performance. The accuracy rates of four classification methods (KNN Classifier, Support Vector Classifier, Random Forest, and DT classifier) were compared.

Using the accuracy score, ROC, F1 score, and Confusion Matrix calculated from each model, the performance of all four classifiers is compared. Finally, we found that the average area under the ROC curve for OVA for our top 3 algorithm models is between 98 and 99.6%, with an accuracy range of 86–95% and F1 score of 85–95%.

Other factors can be used as input variables and other machine learning algorithms added to the modeling process to perform future research. In addition, it is important to make use of data mining techniques' effectiveness to examine students' academic behaviors, deal with their issues, improve the learning environment, and enable data-driven decision-making.

Different feature selection techniques may be utilized in the future. The data sets can also be used with a variety of classification algorithms.

References

1. Francis BK, Babu SS. Predicting academic performance of students using a hybrid data mining approach. *J Med Syst.* (2019) 43:162. doi: 10.1007/s10916-019-1295-4
2. Karthikeyan VG, Thangaraj P, Karthik S. Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation. *Soft Comput.* (2020) 24:18477–87. doi: 10.1007/s00500-020-05075-4
3. AfrozChakure. *Decision Tree Classification*. Medium (2019). Available online at: https://medium.com/swlh/decision-tree-classification-de64fc4d5aac (accessed July 6, 2019).
4. Javatpoint. *Support Vector Machine Algorithm*. Available online at: https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
5. AfrozChakure. *K-Nearest Neighbors (KNN) Algorithm*. Medium (2019). Available online at: https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn (accessed July 6, 2019).
6. Javatpoint. *Random Forest Algorithm*. Available online at: https://www.javatpoint.com/machine-learning-random-forest-algorithm