EDUCATION & TRAINING

Predicting Academic Performance of Students Using a Hybrid Data Mining Approach

Bindhia K. Francis [1,2] · Suvanam Sasidhar Babu [3]

Abstract

Data mining offers strong techniques for different sectors involving education. In the education field the research is developing rapidly increasing due to huge number of student's information which can be used to invent valuable pattern pertaining learning behavior of students. The institutions of education can utilize educational data mining to examine the performance of students which can support the institution in recognizing the student's performance. In data mining classification is a familiar technique that has been implemented widely to find the performance of students. In this study a new prediction algorithm for evaluating student's performance in academia has been developed based on both classification and clustering techniques and been ested on a real time basis with student dataset of various academic disciplines of higher educational institutions in Kerala, India. The result proves that the hybrid algorithm combining clustering and classification approaches yields results that are far superior in terms of achieving accuracy in prediction of academic performance of the students.

Keywords Student academic performance · Educational data mining · Prediction accuracy · K-means clustering

Introduction

A higher education institution considers the academic performance of students as one of the most essential criterion in evaluating them. Colleges, educational institutes or school are developing on greater rate to offer scholars with better education in this competitive globe. The educational institutes concentrate on generating graduates with better academic performances as well as additional curricular activities. They are keeping track on how the students are performing in a specific sector and in what field they need much training [1].

✉ Bindhia K. Francis
dhiyajoji23@gmail.com

Suvanam Sasidhar Babu
sasidharmails@sngce.ac.in

1   Bharathiar University, Coimbatore, India

2   Department of Computer Application, St Thomas College (Autonomous), Thrissur, Kerala, India

3   Department of CSE, Sree Narayana Gurukulam College of Engineering, Kadayiruppu, Kolenchery, Ernamkulam DT, Kerala, India

According to Yassein et al. [2] the academic performance of students relies on several factors like socio economic, personal and other environmental variable. The knowledge about these factors and their impact on performance of students can assist in managing their impact. Predicting academic performance of students has become challenging due to huge number of data in the database of education. If educational institutions can predict the academic performance of students early before their final exams then additional effort must be taken to arrange proper assistance for the low grade students to develop their studies and support them to success. Analyzing students information to classify student and to create association rules or decision trees to make good decisions or to develop their performance is an essential research which focuses mainly on understanding and analyzing educational data of students that denotes their performance of education and produces particular rules and predictions assists students in their future performance of education [3].

Shahiri et al. [4] stated that presently there are several techniques used to evaluate the academic performance of students. Data mining is one of the most familiar techniques use to examine the academic performance of students. It has been used widely in education system and it is referred as educational data mining. Educational data mining is a method used to retrieve useful patterns and information from a large database of education. The

useful patterns and data can be employed in predicting the academic performance of students. As an outcome it would support teachers in offering an efficient approach of teaching. According to Ahmad et al. [5] nowadays the Higher learning institutions database comprises several data about their students and the data is kept developing by times but there is no action taken to acquire knowledge from it. Data mining is the suitable technique is handling the higher learning institutions information to discover new knowledge and data about students. Data mining comprises of machine learning, visualization and statistical techniques to extract and discover knowledge in such a way that humans can interpret easily. Agaoglu [6] focus mainly on education data mining on modeling performance of student instead of performance of instructor. One of the similar tools to estimate the performance of instructor is the questionnaire of course evaluation to evaluate based on perception of students. Acharya and Sinha [7] applied the ML algorithms for student's results prediction. They predict that the best outcomes are acquired with the algorithm decision tree. Mueen et al. [8] used educational data mining to predict the performance of students using three varied classification algorithms of data mining namely neural network using MLP with back propagation type supervised learning algorithm, naïve bayes and decision tree algorithm. Ahmad and Elaraby [9] implemented EDM to conduct the analysis of student performance. The analysis for student performance can help students to develop their performance and it can also recognize students which require special attention to reduce the failure of students and take a right decision at a right time. Similarly Kaur and Sinngh [10] recognize slow learner students and show it by a predictive model of data mining using algorithm based on classification.

Analysis and prediction of performance of students is an essential factor in the environment of education. Educational data mining could be done using several data mining techniques namely neural networks, decision tree, support vector machine, Naïve bayes classifier and K-nearest neighbor. However all of these algorithms fall under either of the two major approaches namely classification and clustering. While there are different algorithms that help in accurately predicting academic performance of students using classification approach, several other algorithms accomplish the same objective by adapting the clustering approach. This research proposes a hybrid approach with combines both clustering algorithm and classification algorithm in order to achieve greater prediction accuracy when compared with that of the existing algorithms.

Literature review

This section reviews on how researchers of the past have made use of several standard classification algorithms in predicting student academic performance.

SVM classifier in student academic performance

According to the research of Asogbon et al. [11] one way to accomplish standard quality education by higher learning institutions is to predict and evaluate the entrant student's performance properly and recommend faculty programmes for them based on the data of education. A DSS based on multi class Support vector Machine technique was constructed to find the students performance in higher learning institutions. According to the study of Pratiyush and Manu [12] with the growth in educational sector there is a raise in new technologies which outcome in huge set of data. Educational mining of data assist in facilitating the use of resources associated to performance of students predicting the results of placements and predicting new trends in education. This study considers students placement data and classification method using support vector machine is followed on training information for finding outcomes which not only supports educational institutions to develop the placement of students from retrieved knowledge a well as enhance the competitive benefit and ecision by applying the techniques of data mining. .

As the sector of education is very essential from the modern era Kadambande et al. [13] is developing application for finding the performance of students by using the techniques of data mining as it is now used widely in the field of education. SVM algorithm is used in this research which is a technique of supervised learning. Through SVM algorithm the prediction is performed and data is analyzed using regression and classification. The Support Vector Machine will assist students to know how much they have to develop themselves so that they can be fit for placements. Oloruntoba and Akinode [14] investigates students academic performance prediction using support vector machine. This study examines the association between preadmission academic profile of students and final performance of academics. The support vector machine outperforms other machine learning algorithms. The support vector machine algorithm parameters was tuned to develop its accuracy and the outcome acquired reveals that radial basis function kernel with penalty outperforms best. Similarly Raihana and Farah Nabilah [15] proposed classification of students based on quality of life and academic performing using SVM. This research classifies students based on academic performance and quality of life. The outcomes for every domain of quality of life revealed that students with both high and low academic performance were categorized into greater class of academic performance.

The below table shows the review of SVM classifier in student academic performance (Table 1):

Naive bayes classifier in student academic performance

Shaziya et al. [16] provides a method to predict students performance in semester exams. This method is based on Naïve

Table 1 Reviews of SVM classifier in student academic performance

S. no	Author	Year	Technique used	Findings of the study
1	Asogbon et al.	2016	Multi Class Support Vector Machine	Helps to predict student's academic performance and provide the management of institution with data that could be used to place students into different programmes of faculty.
2	Pratiyush and Manu	2016	SVM classifier	Helps to improve students placements from extracted knowledge as well as develops the decision making and competitive benefit by using techniques of data mining
3	Kadambande et al	2017	Support Vector Machine algorithm	Optimize educational procedures quality, inspire students by finding the performance in academics and support them to know how much they can develop themselves for getting placements applicable
4	Oloruntobe and Akinode	2017	Support Vector Machine	Accuracy of support vector machine for predicting performance of students
5	Raihana and Nabilah	2018	Support vector machine	Provides insight to both educators and students by unveiling the significance and the role of life quality towards academic performance.

bayes classifier and the objective is to know what students may acquire in their end results of semester. They can be benefitted from prediction of results of students in several ways. Teachers and students take essential steps to develop the outcomes of those students whose prediction result is not fulfilled and a training set of students data is taken to construct the model of naïve bayes and then it is applied on test data to find the results of students end semester. Makhtar et al. [17] examines student's performance using naïve bayes classifier which is one of the methods of classification in data mining to recognize the hidden data between subjects that influenced students performance in Sijil Pelajaran Malaysia. The naïve bayes algorithm can be employed for classification of performance of students in early stage of 2nd semester with 74% accuracy.

According to the study of Patil et al. [18] students choosing engineering as their discipline is developing rapidly but due to different factors and improper education in India the rates of dropout are greater. Students are not capable to shine in the subjects of engineering which are mathematical and complex hence mostly keep term or get drop out in that subject. With the use of data mining techniques the students performance can be predicted in terms of drop out and grade for a subject. Naives bayes algorithm is used in this research and based on the rules acquired from the developed method the system can derive the major factors impacting the performance of students. Razaque et al. [19] described the method of classification which was based on the algorithm of naïve bayes and use for mining academic data. It was used for students along with teachers for academic performance evaluation. It was cautionary approach for students to develop their study performance. This research was an effort to recognize students who need special attention in reducing the failure and take appropriate steps for upcoming semester exams. Divyabharathi and Someswari [20] constructed a predictive model for academic performance of students. As there are several classification

methods available this research used naïve bayes classification technique. By using this model timely decisions can be taken to avoid student's academic risk. The instructor can know how poorly or how well students in class will perform. This study concentrated on validating and developing mathematical models that can be used to predict the academic performance of students in educational institutions.

The below table shows the reviews of Naïve bayes classifier in student academic performance (Table 2):

Decision tree classifier in student academic performance

Kolo et al. [21] proposed a decision tree approach for predicting academic performance of students. To develop on education quality there is a requirement to be capable to predict students academic performance. The factors such as students financial status, gender and motivation to study were discovered to influence the students performance. Several number of students were probable to pass and there is a higher desire of male students to pass than female students. Hamsa et al. [22] develops academic performance prediction model of students for the master and bachelors degree students in electronic and communication and computer science streams using two methods of classification genetic algorithm and decision tree. The resultant model of prediction can be employed to recognize performance of students for every subject. Thereby teachers can categorize students and take early strategies to develop their performance with time. Due to early solutions and predictions are done good outcomes can be anticipated in final exams.

Raut and Nichat [23] research measures performance of students using decision tree classification technique. This research has concentrated much on the techniques of classification which are employed to examine performance by knowledge scope. Providing the data about the outputs

Table 2   Reviews of Naïve Bayes classifier in student academic performance

S. no	Author	Year	Technique used	Findings of the study
1	Shaziya, Zaheer and Kavitha	2015	Naïve Bayes classifier	Predicts students performance in semester exams by using their former marks of semester exams and their overall performance in different tasks of the present semester.
2	Makhtar et al	2017	Naïve Bayes classifier	Provides accurate outputs which can be helpful for teachers to develop the learning outcomes of students for semester exams
3	Patil et al	2017	Naïve Bayes Algorithm	Best suited to predict the performance and grade of students
4	Razaque et al	2018	Naïve bayes algorithm	Helps students and teachers to develop the results of students who were at the risk of failure.
5	Divyabharathi and Someswari	2018	Naïve Bayes classification	Helps management to take timely decisions to develop the performance of students through additional counseling and coaching.

and the particular requirements of researches to development such as students accompaniment along their process of learning and the timely decision taking to hinder academic desertion and risk. Olaniyi et al. [24] presents the data mining method to study performance of students. Data mining offers several methods that could be employed to study performance of students and tas classification is used in this research to estimate the performance of students as there are several methods that can be used for classification involving the method of decision tree. This research also examines the accuracy of various algorithms of decision tree used. Hasan et al. [25] explores academic performance of students using the algorithm of decision tree having parameters like activity of students and academic data of students. The WEKA tool of data mining is used to estimate the algorithm of decision tree for discovery of performance of students along with access time of Moodle. The proposed research assists in developing grades of students in the module and support stakeholders to evaluate and analyze the results and delivery of module.

The below table shows the reviews of Decision Tree classifier in student academic performance (Table 3):

Neural network classifier in student academic performance

Zaldivar-Colado et al. [26] uses artificial neural network techniques for academic performance prediction. This is intended to make aspirants classification to enter to a career in university in various levels according to the probability of meeting a category of performance. This research predicts students shortcomings in different perspective of courses during the academic career and provide solutions to avoid them. Binh and Duy [27] stated that several education authors focused on styles of learning and its suggestions. They depend that students have various types of personality which tend to have different styles of learning which in turn impact performance of students in every subject type. This research have constructed an artificial neural network to predict the performance of academics based on the learning style of students. According to the research of Gerritsen [28] neural networks have been successful and widespread implementation in vast number of applications of data mining surpassing classifiers. This research main aim is to examine if neural networks are a proper classifier to predict performance of students from LMS (learning management system) in educational data mining context.

Table 3   Reviews of decision tree classifier in student academic performance

S. no	Author	Year	Technique used	Findings of the study
1	Kolo et al	2015	Decision tree	Predict academic performance of students as either fail or pass by using decision tree
2	Hamsa et al	2016	Decision tree and genetic algorithm	Students can see their academic updates and data so that reputed firms have a tie up with the institution can browse students according to their needs.
3	Raut and Nichat	2017	Decision tree classification	Helps to analyze the slow learning students who studies poor and develop their skills as early to accomplish their target
4	Olaniyi et al	2017	Decision tree	Helps in identifying dropouts and students require special attention and permit instructors to offer proper advising.
5	Hasan et al	2018	Decision tree algorithm	Develops grades of students

The characteristics used for training emerge from learning management system information acquired during every course length and range from usage data like time invested on every page of course to grades acquired for course quizzes and assignments.

Okubo et al. [29] in their study proposed an approach for predicting students final grades by a RNN (recurrent neural network) from the log information stored in the systems of education. The log information indicated the activities of learning of students who utilizes the LMS, the electronic book system and electronic portfolio system. This research used this approach to log information from students and investigated the prediction accuracy. Bendangnuksung and Prabu [30] proposed deep neural network for predicting student's performance. A neural network named deep neural network is proposed that shows students which type category it belongs to. This offers knowledge to educational institution so that they can provide a solution to essential failing students. The proposed deep neural network aims to predict students whether they exist under pass or fail category through logistic regression analysis.

The below table shows the reviews of Neural Network classifier in student academic performance (Table 4):

Methodology

This section explains the methodology which involves the phases used for the process of prediction of academic performance of the students through classification and clustering student data. The phases are depicted in the below figure (Fig. 1):

Each phase of the methodology proposed is explained below:

Phase 1:   Understanding of Business

The main aim of this study is to develop a model of prediction for academic performance of students using data mining classification and to decide which classifier performs good with the gathered data set of education.

Phase 2:   Attributes of Data collection

To build the model of prediction the students features and their description which was gathered was shown in the below table (Table 5):

The data has been gathered based on Demographic features, academic features, behavior features and extra features.

Phase 3:   Data Pre-processing:

After the collection of dataset pre-processing methods are applied to develop the data set quality. The data pre-processing is regarded as an essential step in the process of knowledge discovery which involves cleaning of data, feature selection, data transformation and data reduction. Before applying the data mining algorithm the data preprocessing is the step which transforms the actual information into an applicable shape to be used by a specific algorithm of mining.

Phase 4:   Applying Data Mining Classification Algorithms

This study carried out the experiments using SVM, Naïve Bayes, Decision tree and Neural Network classifiers. These four classifiers have been chosen to evaluate the measures of dataset.

Support vector machine is used for solving the non linear function estimation and pattern recognition issues. The Support vector machine is used for representing the training information nonlinearly into a higher dimensional feature space then builds an isolating hyper plane with maximum margin. This yields a non linear boundary of decision in input space. Support vector machine solutions are acquired from issues of quadratic programming possessing a global solution [31].

Table 4   Reviews of neural network classifier in student academic performance

S. no	Author	Year	Technique used	Findings of the study
1	Zaldivar-Colado et al	2017	Neural network	Prediction of academic performance of drop out students and provide suggestions to avoid the drop out rates.
2	Binh and Duy	2017	Artificial neural network	Predicts academic performance based on learning style of students
3	Geritssen	2017	Neural Networks	Neural networks are useful for prediction of student performance and outperform than other classifiers
4	Okubo et al	2017	Recurrent Neural Network	Effective in predicting final grades of students
5	Bendangnuksung and Prabu	2018	Deep Neural network	Achieve greater accuracy and also helps to predict performance of students and recognize students who has greater chance of failing before to offer solution

Fig. 1 Flow diagram of the
proposed methodology
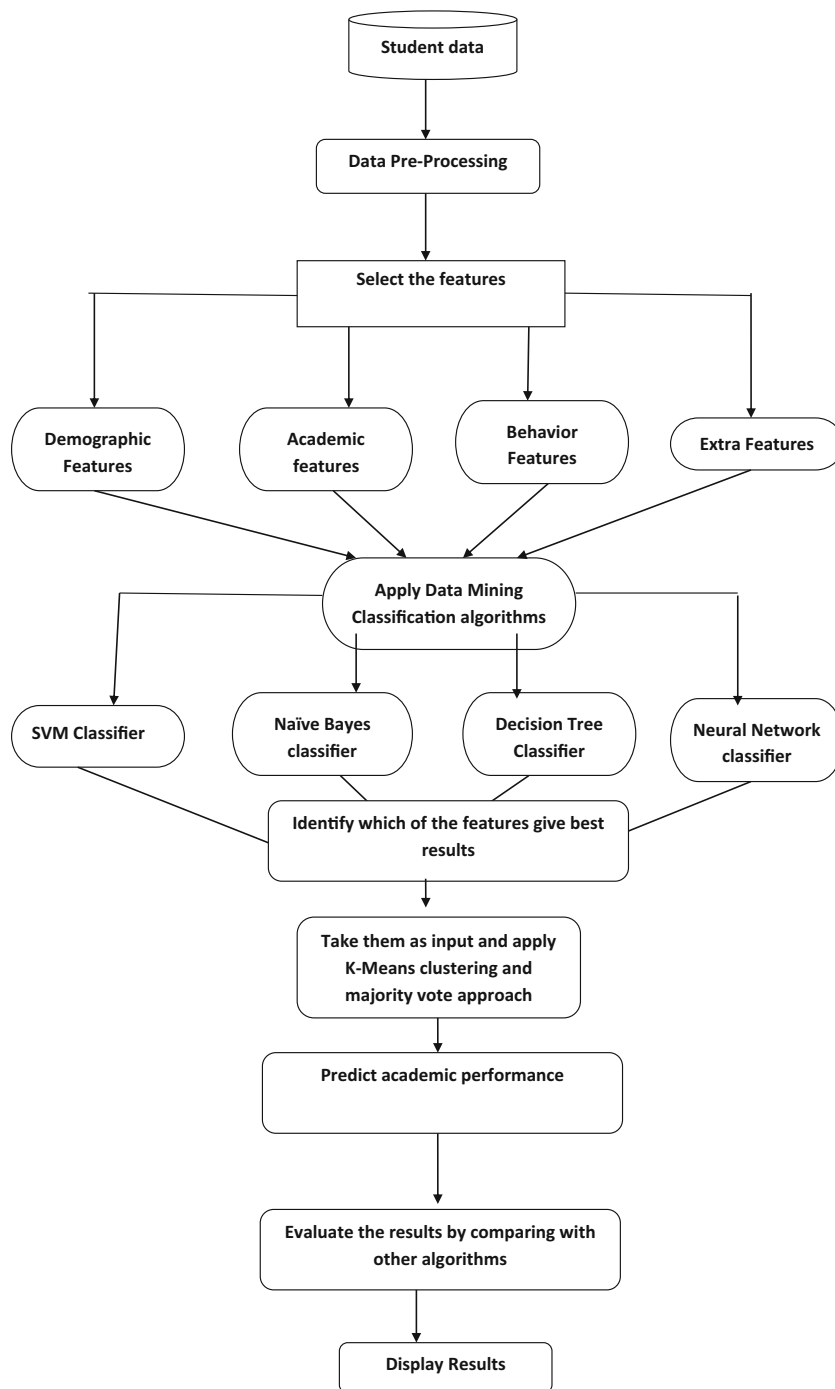
Decision trees are probably the most commonly used technique of data mining. It is a structure of flowchart where every internal node indicates an attribute test and every branch indicates a result of the test and class label is indicated by every terminal node. Decision tree uses a decision tree as a predicative model which represents observations about an item to inference about the target value of the item [32].

Neural networks developed as an essential classification tool. Neural networks are a promising alternative to different methods of conventional classification. Neural networks are data driven and self adaptive methods where they can adjust themselves to data without any explicit specification of distributional or functional form for the model [33].

Naïve Bayes is among the entire easiest probabilistic classifiers. It always performs well in real world applications despite the powerful assumption that entire features are independent conditionally. In the classifiers learning process with the known structure, conditional probabilities and class probabilities are estimated using training information and then these probabilities values are used to classify new observations [34].

Table 5 Student features and their description

Category of feature	Sub-features	Description
Demographic features	Gender	Students gender
	Nationality	Students nationality
	Place of birth	Students place of birth
	Relation	Relation who is responsible for students
Academic features	Stage ID	Educational stage of students
	Grade ID	Grade level of students
	Section ID	Classroom section which the student belongs to
	Topic	Topic of the course
	Semester	Semester of school year
Behavioral features	Raised hands	Behavior of students during learning of education
	Visited resources	
	Announcement view	
	Discussion	
Extra features	Parent answering survey	Survey answered by parents provided from school
	Parent school satisfaction	Parents satisfaction regarding schools
	Student absence days	Absence days of students in schools

Phase 5:   Apply K-means clustering plus majority voting

The data mining classification algorithms are applied to K-means clustering plus majority voting which is proposed in this study. K-means clustering is the most vastly used algorithm of clustering which is used in several areas such as pattern recognition, computer vision and information retrieval. K-means clustering allots n points of data into k number of clusters so that same points of data can be grouped together. It is an iterative process which allots every point to cluster whose centroid is the closest. Then it again evaluates these groups centroid by taking its average. In this research a new algorithm is proposed by integrating K-means clustering plus majority voting which predicts the best accuracy of students

Phase 6:   Find the results

After applying the K-means clustering plus majority voting the four classifiers are compared and the best accuracy (i.e. greater number of accuracy) of students is found. The algorithm used for decision tree is ID3 whereas the algorithm used for neural network is MLP which has two layers with size 2 and 5. According to Bedi (2015) ID3 algorithm uses information gain as criterion of splitting. Topmost decision node is the good predictor and it is also known as root node. The attribute with greater gain of information is chosen as split attribute.

Information gain is employed to create tree from instances of training. This tree is employed to categorize test information. When information gain methods to zero or entire instances belong to individual target then growth of tree terminates. Wankhede [35] has stated that the most relevant neural network model is the multilayer perceptron. This kind of neural network is referred as a supervised network because it needs a desired result in order to learn. The purpose of this kind of network is to create a model that maps the input correctly to the output using historical information so that the model can be used to generate the output when desired result is unknown.

Phase 7:   Evaluation of the algorithm proposed

In this research four measures have been used for the evaluation of the quality of classification. The four measures are precision, recall, fscore and accuracy. Precision is the ratio of the properly classified cases to the total number of misclassified cases and properly classified cases. Recall is the proportion of correctly classified cases to total number of correctly classified cases and unclassified ones. F-score integrates the precision and recall measure which is regarded as a good indicator of relationship between them. Lastly accuracy is the ratio of the total number of predictions where calculated properly. The equations of precision, recall, fscore and accuracy is stated below:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} Fscore$$

$$= 2\frac{Precision\ c * Recall\ c}{Precision\ c + Recall\ c} Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Negative + False\ Positive + True\ Negative}$$

Phase 8:   Display the Result

The last phase is the result display which provides the details of the features and algorithm that gives the best accuracy in terms of graphical representation.

Results and discussion

Layer 1: Identifying essential features based on the accuracy from baseline machine learning algorithms

For every student the features are chosen by the first testing of accuracy boost by using those characteristics on standard baseline machine learning algorithms namely SVM, KNN, DT etc.

Layer 2: Clustering students based on these features into three clusters

The students are clustered based on these features into three clusters. The obtained features are then passed into K-Means clustering algorithm to acquire clusters indicating high, medium and low performing students. New students are represented to these clusters and allotted labels based on majority voting of students in these clusters. Then the accuracy is computed on this test set.

Results of the features

The classification results by using all features are represented by a table with a bar graph:

Demographic features (Table 6, Fig. 2)

Table 6   Demographic features

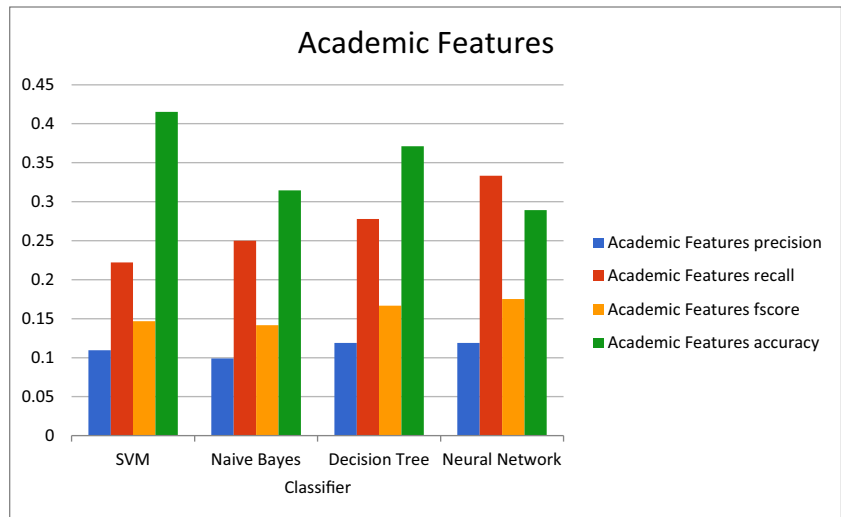| Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.253165	0.555556	0.347826	0.528302
Naive Bayes	0.083333	0.194444	0.116667	0.333333
Decision Tree	0.267442	0.638889	0.377049	0.522013
Neural Network	0.195122	0.444444	0.271186	0.459119

Fig. 2 Demographic features
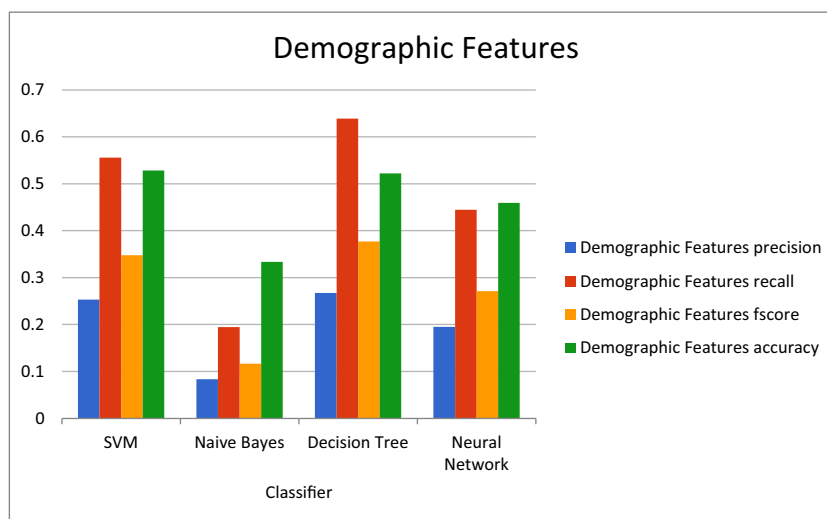
Academic features (Table 7, Fig. 3)

Table 7   Academic features

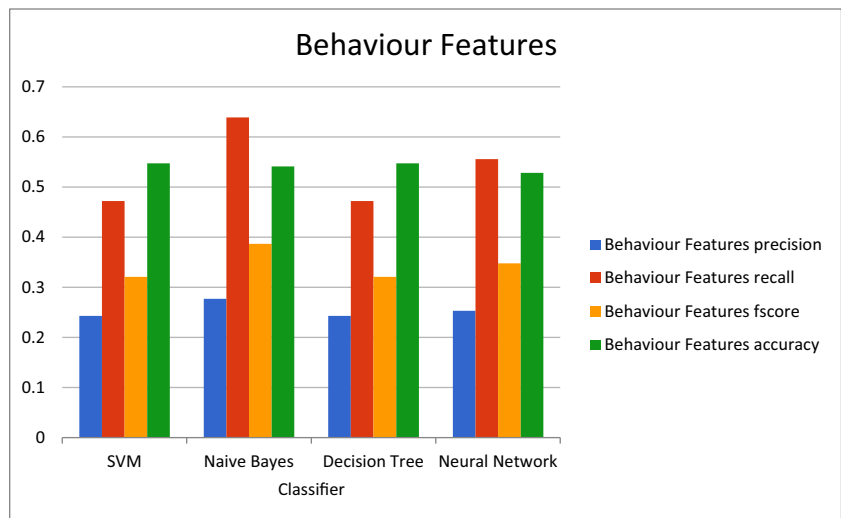| Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.109589	0.222222	0.146789	0.415094
Naïve Bayes	0.098901	0.25	0.141732	0.314465
Decision Tree	0.119048	0.277778	0.166667	0.371069
Neural Network	0.118812	0.333333	0.175182	0.289308

Fig. 3 Academic features

Behavior features (Table 8, Fig. 4)

Table 8   Behavior features

Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.242857	0.472222	0.320755	0.54717
Naïve Bayes	0.277108	0.638889	0.386555	0.540881
Decision Tree	0.242857	0.472222	0.320755	0.54717
Neural Network	0.253165	0.555556	0.347826	0.528302

Fig. 4 Behavior features

Extra features (Table 9, Fig. 5)

Table 9   Extra features

Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.333333	0.75	0.461538	0.603774
Naive Bayes	0.351064	0.916667	0.507692	0.597484
Decision Tree	0.366667	0.916667	0.52381	0.622642
Neural Network	0.366667	0.916667	0.52381	0.622642

Fig. 5 Extra features
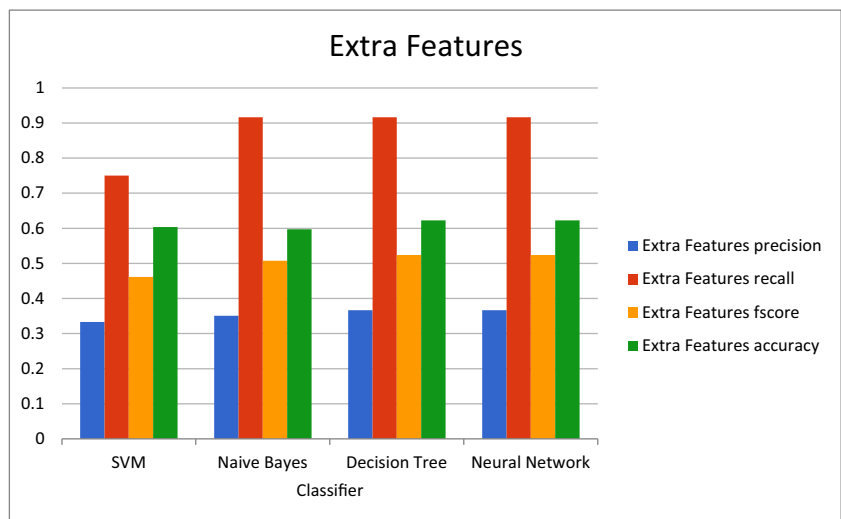
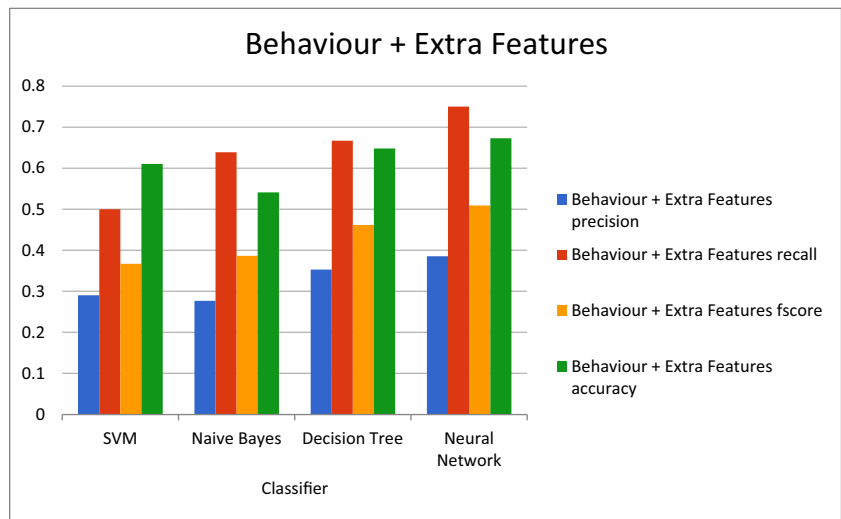## Behavior + Extra features (Table 10, Fig. 6)

Table 10 Behavior + Extra features

Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.333333	0.75	0.461538	0.603774
Naive Bayes	0.351064	0.916667	0.507692	0.597484
Decision Tree	0.366667	0.916667	0.52381	0.622642
Neural Network	0.366667	0.916667	0.52381	0.622642

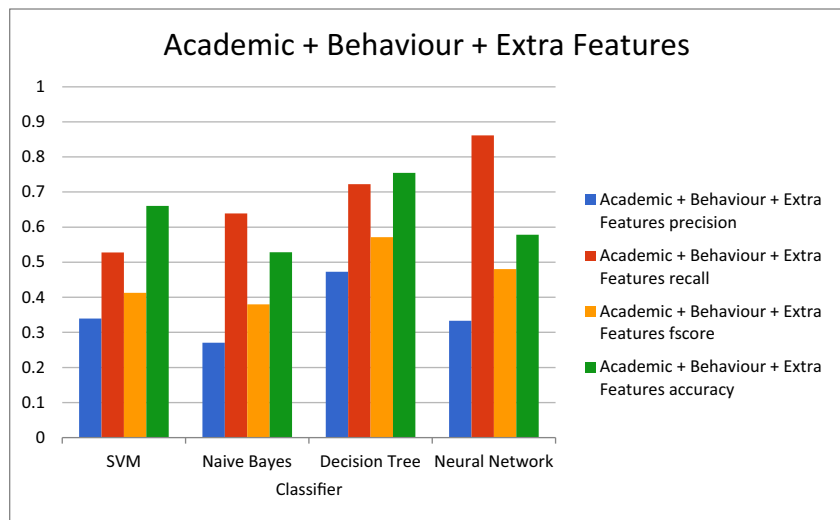**Fig. 6** Behavior + Extra features

Academic + Behavior + Extra features (Table 11, Fig. 7)

Table 11 Academic + Behavior + Extra features

Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.339286	0.527778	0.413043	0.660377
Naive Bayes	0.270588	0.638889	0.380165	0.528302
Decision Tree	0.472727	0.722222	0.571429	**0.754717**
Neural Network	0.333333	0.861111	0.48062	0.578616

Fig. 7 Academic + Behavior +
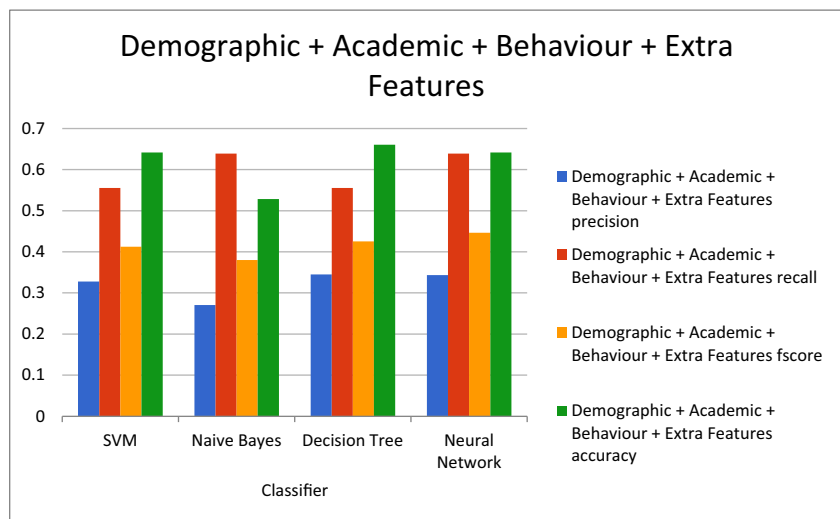Extra features

Demographic + Academic + Behavior + Extra features
(Table 12, Fig. 8)

Table 12   Demographic + Academic + Behavior + Extra features

Classifier	Precision	Recall	Fscore	Accuracy
SVM	0.327869	0.555556	0.412371	0.641509
Naive Bayes	0.270588	0.638889	0.380165	0.528302
Decision Tree	0.344828	0.555556	0.425532	0.660377
Neural Network	0.343284	0.638889	0.446602	0.641509

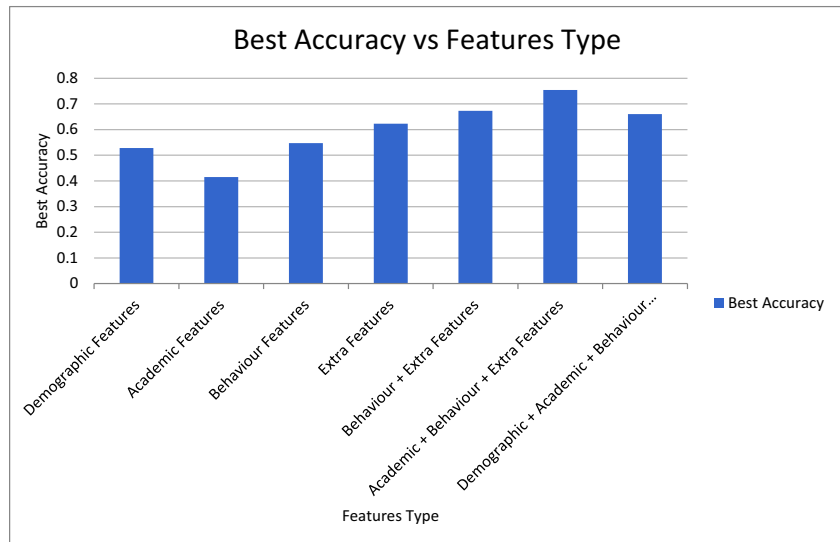**Fig. 8** Demographic + Academic
+ Behavior + Extra Features

Table 13   Feature type vs. best accuracy

Features type	Best accuracy
Demographic features	0.528302
Academic features	0.415094
Behaviour features	0.54717
Extra features	0.622642
Behaviour + Extra features	0.672956
Academic + Behaviour + Extra features	**0.754717**
Demographic + Academic + Behaviour + Extra Features	0.660377

Table 14   Result of the new approach

Algorithm	Precision	Recall	Fscore	Accuracy
Clustering	0.641509	0.641509	0.641509	0.641509
Decision tree	0.622642	0.622642	0.622642	0.622642
Neural network	0.484277	0.484277	0.484277	0.484277

From the previous analysis and base paper this research found that behavioral and extra features work the best in boosting the accuracy of the system. In this research a clustering based approach is used employing behaviour and extra features to obtain clusters and also employed them to classify the users in the test dataset into one of the categories using a majority vote on the clusters they belonged. After applying the new algorithm proposed in this study the result of the new approach is (Table 14, Fig. 10):

Feature type vs. best accuracy

The below table and figure shows the features type vs best accuracy of students (Table 13, Fig. 9):

Fig. 9 Feature type vs. best accuracy

Fig. 10 Result of the new approach

From the above graph it was clear that the clustering technique has performed much better than a decision tree or a neural network based technique.

Conclusion

The achievement in academics of students is a huge concern for academic institutions all over the globe. The vast use of learning management system generates huge number of data about learning and teaching interactions. This data comprises of hidden knowledge that could be employed to develop the academic performance of students. The results of the application of the proposed hybrid algorithm on student data set shows that there is a strong relation between behavior of learner and their academic performance of the students. The accuracy of the proposed hybrid model combining clustering and classification is 0.7547 when applied to the academic, behavior and extra features of the student data set and is found to be superior to that of the other existing algorithms. This model can help educators to perceive learners recognize weak learners to develop the process of learning and reducing the failure rates of academics and also helps administrators to manage better based on the results of learning system. In future the model could be extended further to support huge varieties of features of the student dataset.

Compliance with ethical standards

Conflict of interest   This paper has not communicated anywhere till this moment, now only it is communicated to your esteemed journal for the publication with the knowledge of all co-authors.

Ethical approval   This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Soni, A., Kumar, V., Kaur, R., and Hemavath, D., Predicting student performance using data mining techniques. *International Journal of Pure and applied Mathematics* 119(12):221–227, 2018.
2. Yassein, N. A., Helall, R. G. M., and Mohomad, S. B., Predicting student academic performance in KSA using data mining techniques. *Journal of Information Technology and Software Engineering* 7(5):1–5, 2017.
3. Saa, A. A., Educational data mining & students' performance prediction. *Int. J. Adv. Comput. Sci. Appl.* 7(5):212–220, 2016.
4. Shahiri, A. M., Husain, W., Rashid, N. A. A review on predicting students performance using data mining techniques. The 3rd Information System International Conference, Elsevier, 72, p 414–422, 2015.
5. Ahmad, F., Ismail, N., and Aziz, A. A., The prediction of students' academic performance using classification data mining techniques. *Appl. Math. Sci.* 9(129):6415–6426, 2015.
6. Agaoglu, M., Predicting instructor performance using data mining techniques in higher education. *IEEE Access* 4:2379–2387, 2016.
7. Acharya, A., and Sinha, D., Early prediction of student performance using machine learning techniques. *Int. J. Comput. Appl.* (1):107, 2017.
8. Mueen, A., Zafar, B., and Manzoor, U., Modeling and predicting students academic performance using data mining techniques. *International Journal of Mod. Education, Computer Science* 8: 36–42, 2016.
9. Ahmad, A. B. E. D., and Elaraby, I. S., Data mining: A prediction for students performance using classification method. *World Journal of Computer Applications Technology* 2:43–47, 2014.
10. Kaur, M, and Sinngh, G. S. J. Classification and prediction based data mining algorithms to predict slow learners in education sector. 3rd International Conference on Recent Trends in Computing (ICRTC-2015), 2015.
11. Asogbon, M. G., Samuel, O. W., Omisore, M. O., and Ojokoh, B., A multi-class support vector machine approach for students academic performance prediction. *International Journal of Multidiscplinary and Current Research* 4:210–215, 2016.
12. Pratiyush, G., and Manu, S., Classifying educational data using support vector machines: A supervised data mining technique. *Indian J. Sci. Technol.* 9(34), 2016.
13. Kadambande, A., Thakur, S., Mohol, A., and Ingole, A. M., Predicting students performance system. *International Research Journal of Engineering and Technology* 4(5):2814–2816, 2017.
14. Oloruntoba, S. A., and Akinode, J. L., Student academic performance prediction using support vector machine. *International Journal of Engineering Sciences and Research Technology* 6(12): 588–597, 2017.
15. Raihana, Z., and Farah Nabilah, A. M., Classification of students based on quality of life and academic performance by using support vector machine. *Journal of Academia UiTM Negeri Sembilan* 6(1): 45–52, 2018.
16. Shaziya, H., Zaheer, R., and Kavitha, G., Prediction of students performance in semester exams using a Naïve Bayes classifier. *International Journal of Innovative Research in Science, Engineering and Technology* 4(10):9823–9829, 2015.
17. Makhtar, M., Nawang, H., and Shamsuddin, S. N. W., Analysis on students performance using Naïve Bayes classifier. *J. Theor. Appl. Inf. Technol.* 95(16):3993–3999, 2017.
18. Patil, V., Suryawanshi, S., Saner, M., Patil, V., and Sarode, B., Student performance prediction using classification data mining techniques. *International Journal of Scientific Development and Research* 2(6):163–167, 2017.
19. Razaque, F., Soomro, N., Shaikh, S. A., Soomro, S., Samo, J. A., Kumar, N., and Dharejo, H. Using naïve bayes algorithm to students' bachelor academic performances analysis. In: *Engineering Technologies and Applied Sciences (ICETAS), 2017 4th IEEE International Conference.* p 1-5, 2017.
20. Divyabharathi, Y., and Someswari, P., A framework for student academic performance using naive Bayes classification technique. *J. of Advancement in Engineering and Technology* 6(3):1–4, 2018.
21. Kolo, K. D., Adepoju, S. A., and Alhassan, J. K., A decision tree approach for predicting students academic performance. *IJ Education and Management Engineering* 5:12–19, 2015.
22. Hamsa, H., Indiradevi, S., and Kizhakkethottam, J. J., Student academic performance prediction model using decision tree and fuzzy genetic algorithm. *Procedia Technology* 25:326–332, 2016.
23. Raut, A. B., and Nichat, M. A. A., Students performance prediction using decision tree. *Int. J. Comput. Intell. Res.* 13(7):1735–1741, 2017.
24. Olaniyi, A. S., Kayode, S. Y., Abiola, H. M., Tosin, S. I. T., and Babatunde, A. N., Students performance analysis using decision tree algorithms, *Annals. Computer Science Series* 15(1), 2017.
25. Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S., and Sarker, K. U. Student Academic Performance Prediction by using

Decision Tree Algorithm, 4th International Conference on Computer and Information Sciences. 2018.

26. Zaldivar-Colado, A., Aguilar-Calderon, J. A., Garcia-Sanchez, O. V., Zurita-Cruz, C. E., Moncada Estrada, M., and Bernal-Guadiana, R., Artificial neural networks for the prediction of students academic performance, 8th International Technology, Education and Development Conference, 2014.

27. Binh, H. T., and Duy, B. T., Predicting students' performance based on learning style by using artificial neural networks, 9th International Conference on Knowledge and Systems Engineering, Vietnam, 2017.

28. Gerritsen, L. Predicting student performance with Neural Networks (Doctoral dissertation, Tilburg University). 2017.

29. Okubo, F., Yamashita, T., Shimada, A., and Ogata, H., A neural network approach for students' performance prediction. In: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pp. 598-599, 2017.

30. Bendangnuksung and Prabu, Students performance prediction using deep neural network. *Int. J. Appl. Eng. Res.* 13(2):1171–1176, 2018.

31. Tripathy, B. K., Sahu, S. K., Prasad, M. B. N. V., and Barik, K. K., A support vector machine binary classification and image segmentation of remote sensing data of Chilika Lagloon. *International Journal of Research in Information Technology* 3(5):191–204, 2015.

32. Sharma, H., and Kumar, S., A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)* 5(4):2094–2097, 2016.

33. Jebaseeli, A. N., Neural network classification algorithm with M-learning reviews to improve the classification accuracy. *Int. J. Comput. Appl.* 71(23), 2014.

34. Taheri, S., and Mammodov, M., Learning the naives Bayes classifier with optimization models. *Int. J. Appl. Math. Comput. Sci.* 23(4):787–795, 2014.

35. Wankhede, S. B., Analytical study of neural network techniques: SOM, MLP and classifier-a survey. *IOSR Journal of Computer Engineering* 16(3):86–92, 2015.