International Journal of Advanced Research and Review

www.ijarr.org

Hybrid Classification Methods for Assessing Academic Achievement

[1] DR. P. Ramasubramanian, [2] Thota Bharghavi,

[1] Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract-

The process of discovering useful information from large databases via natural selection is known as data mining. One well-known area of study in educational data mining is student performance prediction. Examining the methods for classification as a hybrid classification is the goal of the study. Random Forest, Multilayer Perceptron, C4.5, and Radial Basis Function Network were the categorization algorithms we used. The accuracy of the classifications was first calculated by each classification method separately. The classification accuracy for the Radial Basis Function network, Multilayer Perceptron, C4.5, and Random Forest Algorithm was 72.9167%, 75.4167%, 75%, and 73.125%, respectively. Combining a Radial Basis Function network with a multilayer perceptron increases the classification algorithm's accuracy. The classification accuracy achieved by this hybrid method is 75.625%. Then, we achieved a classification accuracy of 76.4583% by combining the C4.5 method with the random forest approach. Compared to separate classification algorithms, hybrid algorithms outperformed them in terms of accuracy.

Keywords:

Topics covered include voting, hybrid classification, multilayer perceptrons, performance prediction, and random forests.

I.INTRODUCTION

Data mining is the practice of discovering patterns and correlations in big data sets in order to solve issues [13]. Data Mining has found several uses. The educational sector has recently attracted the attention of researchers looking to extract useful insight from learning data repositories. When looking for valuable insights in educational datasets, data mining is the go-to technique. The term "Educational Data Mining" (EDM) refers to a collection of practices and tools developed for the purpose of automatically sampling data sets generated by instructional activities. EDM is a combination of teaching and predicting how well students will do in school so that teachers may propose changes to the way classes are currently taught [12]. Within the realm of educational data mining, several approaches have been developed and used with the aim of forecasting students' academic outcomes. Predicting terrorist attacks and patients' risk of cardiovascular disease are two applications of hybrid classification algorithms. Information on students, their schools, their parents, and their grades are all part of EDM. A classification system has used this data to forecast how well a pupil would do in school. Teachers and schools may use this forecast to gauge future students' performance and tailor their instruction accordingly. Related work using hybrid categorization in data mining and machine learning is presented in Section II of this article. The third section delves into the practical application of this study. We detail the findings and analysis in Section IV. This study comes to a close in Section V.

II. RELATED WORK

Using direct, indirect, and hybrid machine learning models, Vladislav Miskovic [5] assessed the predicted accuracy of pharmaceutics symptomatic, e-commerce, retailing, and economic analytical issues. For hybrid classification, he has employed algorithms such as C4.5, C45Rules, KNN, and Random Forest.
A model for predicting students' academic performance was developed by Thaddeus MatunduraOgwoka et al. [6] using a combination of k-means and decision tree algorithms. This algorithm enhances precision. As a result, the school is able to more accurately predict how well its kids would perform in class.The hybrid approach developed by Akanksha Ahlawat and Bharti Suri [7] makes use of clustering, pattern assessment, and decision trees to represent the results. They evaluated their results on real-world datasets and found that they improved accuracy. For the purpose of making performance predictions for students, Hamza Turabieh [8] used a hybrid feature selection method. Better predictions were achieved by him using KNN, CNN, NB, and C4.5 algorithms. They have demographic data, student grades, and school records in their data collection. Proceedings Mr. Ryan The work of S.J.D. Baker et al. [15] examined EDM's history and current trends. A greater emphasis on the results of predictions has been their focus. Using a hybrid approach based on clustering and classification, Bindhia et al. [9] developed a new method of academic performance prediction. Their findings demonstrate that the prediction accuracy is
significantly improved by the hybrid approach that combines clustering and classification. In order to classify the health data set, Abhishek Lal and Kumar [10] used Decision Tree and Naïve Bayes techniques to create a hybrid system. The data set's identity prevented them from reaching the desired level even when using separate classifiers. The classification accuracy was enhanced after the creation of the hybrid classifier. In order to foretell whether individuals will develop cardiac problems, Ankita Dewan and Meghna Sharma [11] used a hybrid classification method. For the purpose of illness prediction, they used top data mining methods as C4.5, Multilayer Perceptron, Back Propagation, Naïve Bayes, and Support Vector Machine. According to their findings, the back propagation method outperforms the competition. For the purpose of academic performance prediction, Deepali R. Vora and Kamatchi Rajamani [14] have created a hybridized classifier SVM with a deep learning model. By computing their prediction model with specificity and comparing the representation to other models, they were able to get improved results.

III.IMPLENTATION

Data for this study came from the UCI repository. The prediction of student performance was based on a total of 480 data samples. The student's performance was predicted using the Weka Machine Learning technology. Section A: Priming
A good attribute for classification may be obtained by preprocessing. At this point, the weka tool has been fed the edudata.arff file. The data collection initially has seventeen characteristics. An improved classification is achieved by submitting these characteristics to the data preprocessing stage. To choose the optimal attribute for the Classification stage, the CfsSubsetEval attribute selection technique was used. You can see all the characteristics utilized for preprocessing in Figure 1.

Figure 1. Data Preprocessing

Evaluating CfsSubsets

In other words, it determines what a class of characteristics means. These characteristics account for the degree of recurrence across features as well as their individual predictive capabilities [5]. The feature selection technique CfsSubsetEval was used to choose 6 characteristics for the following phase. The CfsSubsetEval attribute selection technique yielded the six most important properties, as shown in figure 2.

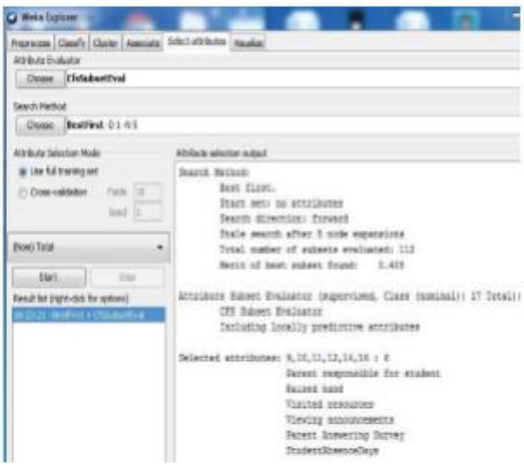

Figure 2. Attribute Selection

Features Explanation As a measure of parental involvement in their children's education, "parents responsible for students" is taken into account. In order to raise the level of performance among students, this component is crucial. Proceedings Students may raise their hands and ask questions while the instructor is teaching. Teachers can tell whether students are paying attention in class because they raise their hands. Students' engagement with the material is shown by the resources they have visited. It determines the frequency with which the materials are accessed by the pupils. The number of times students have seen the announcements is shown by this characteristic. This feature is used to determine whether parents are responding the survey questions or not. The number of days that students are present or absent from class is a strong predictor of how well they will do in class. All things considered, these characteristics indicate whether a student will perform at a high, medium, or low level. The data collection is expected to provide this result. Part B: Categorization This stage involves the individual classification of two methods for classification. A 10-fold cross-validation approach was used for data classification. After that, the Radial Basis Function Network was used. A network that uses radial basis functions: Depending on how far you go from a fixed point, the value of a Radial Basis Function (RBF) may go up or down. A Gaussian activation function is used by the RBF network. It is tri-layered.

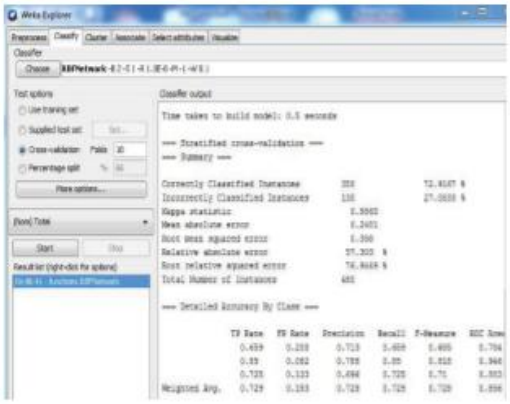

Figure 3.Classification of RBF Network

The input layer's previous function was to only receive data as input. The hidden layer makes use of the Gaussian activation function. The linear activation function is used by the Output layer [1]. The output was predicted using an RBF network.

Classification results using RBF Network are shown in figure 3. While the RBF classification method got 350 samples right, it got 130 examples wrong. The RBF network has a classification accuracy of 72.9167. There are several layers in a multilayer perceptron. Model-learning perceptrons form MLP. In this architecture, neurons are arranged in layers, and their outputs are directed towards the output layer at all times [1]. It is recommended that the multilayer perceptron be trained using the back propagation network technique. It undergoes training in a controlled environment to improve its production, cost, and resulting function. In order to find an approximate mathematical representation for the provided conditions, a multilayer perceptron may plot out and identify sufficient difficulty [2]. Figure 4 displays the output of the multilayer perceptron, which correctly categorizes 362 instances while incorrectly classifying 118. Multilayer perceptrons have a classification accuracy of 75.4167 percent.

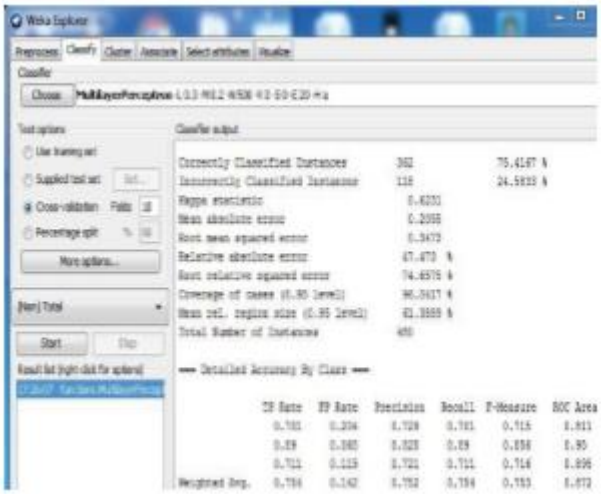

Figure 4.Classification of Multilayer Perceptron

J48: The J48 algorithm enhances the ID3 method by building a decision tree using a contract with both constant and definite characteristics. In order to store continuous properties, J48 splits the feature values in half. J48 employs the Gain Ratio Method for splitting. That guarantees an above-average improvement in knowledge [4]. It takes care of missing qualities as well. In Weka, the J48 algorithm is also known as the C4.5 algorithm.

The outcome of the J48 classification, which properly identifies 360 occurrences and wrongly classifies 120, is shown in figure 5. J48 has a 75% accuracy rate in its categorization.

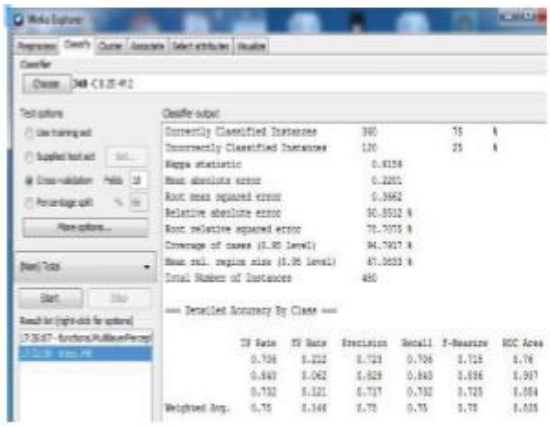

Figure 5.Classification of J48

The Random Forest method is a subset of the ensemble learning framework developed specifically for use in tree classification tasks. It belongs to the class of algorithms known as supervised learning. One great thing about Random Forest trees is that they can be used for both one-way classifications and regression. It has been shown to effectively address the issue of data overfitting in decision trees [2].

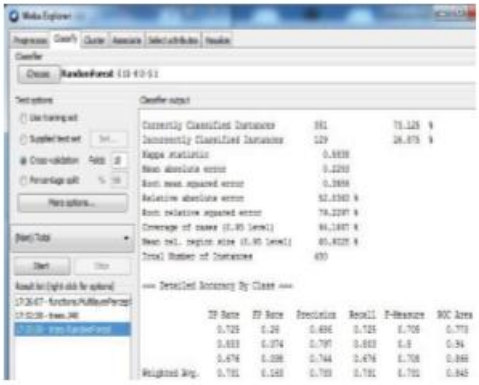

Figure 6.Classification of Random Forest

A random forest algorithm's output is seen in figure 6. Out of 351 occurrences, this algorithm got it right and 129 got it wrong. The accuracy rate for categorization is 73.1 25%. Methods of Casting a Vote We have used hybrid classification algorithms to improve the accuracy of categorization even more. To create the hybrid categorization, a voting mechanism was used. It is an approach to aggregation that is associated with combining the results of many classifiers. Each classifier casts a single vote in the majority-voting process. In order to determine the final prediction, the number of votes is added up; that is, the class with the most votes for each element is the one used [3]. The hybrid RBF/RBF network classification result is shown in Figure 7. Out of 363 cases that were successfully identified, 117 were wrongly classified, all because we combined two methods. A new level of 75.625% categorization accuracy has been achieved.

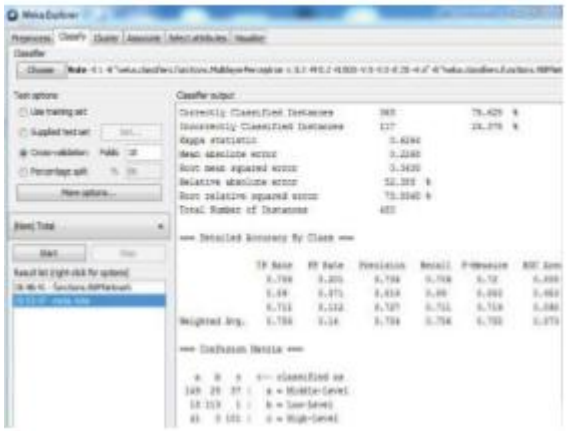

Figure 7.Classification of RBF and Multilayer Perceptron

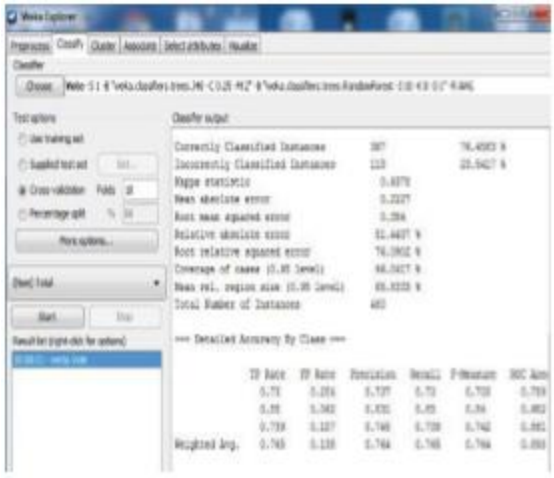

Figure 8.Classification of Random Forest and J48

The outcome of a hybrid classification using the J48 and random forest methods is shown in Figure 8. With the help of two algorithms working together, the classification accuracy reached 76.4583%. Out of 367 cases, this hybrid algorithm gets the classifications right, while 113 get them wrong.

IV.RESULTS AND DISCUSSION

In this suggested study we have picked 4 classification methods for prediction. We started by sorting each algorithm separately. For improved prediction, we merged two algorithms. Here we evaluate four algorithms based on their accuracy and the number of occurrences they properly categorized.

Algorithm	Correctly Classified Instances	Accuracy
RBF	350	72.9167%
MLP	362	75.4167%
J48	360	75%
Random Forest	351	73.125%

Table 1: Individual Performance of Algorithms

You can see how well four different algorithms did when it came to classification in Table 1. The multilayer perceptron method outperforms the other three algorithms in terms of accuracy. With 362 cases successfully categorized, it has an accuracy rate of 75.417%. With respective accuracies of 72.9167%, 75%, and 73.125%, the RBF, J48, and Random Forest algorithms successfully identified 350, 360, and 351 instances, respectively.

Hybrid Algorithm	Correctly Classified Instances	Accuracy
RBF + MLP	363	75.625%
J48+Random Forest	367	76.4583

Table 2: Hybrid Performance of Algorithms

As seen in Table 2, hybrid classification algorithms provide the following results. With an accuracy of 75.625%, the RBF and MLP algorithms provide 363 properly identified cases. A total of 367 occurrences were accurately categorized using the J48 and Random Forest algorithms, yielding an accuracy rate of 76.4583%. Tables 1 and 2 showed the results of the individual and hybrid algorithms used for performance classification. The approach enhances accuracy and increases the number of properly categorized examples while developing the hybrid classification algorithm. With this method, future qualities may be accurately predicted.

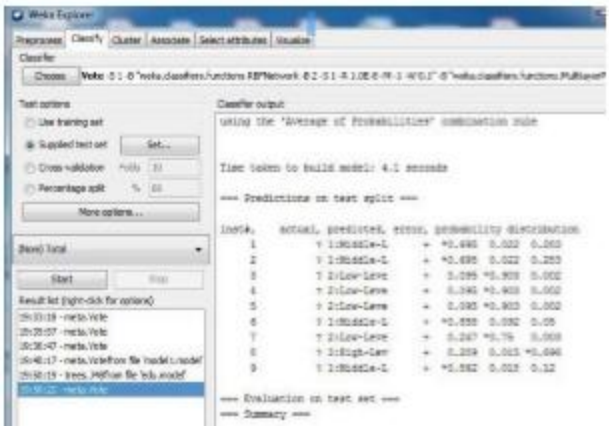

Figure 9.Output Prediction

The results that the hybrid classification algorithm is expected to provide are shown in Figure 9. We may deduce that the hybrid classification technique is effective from this image. Improved forecasting is possible with its help.

V.CONCLUSION AND FUTURE WORK

To predict the students' academic success, this research used Random Forest, J48, Multilayer Perceptron, and Radial Basis Function networks. We calculated the classification accuracy for each of these four techniques separately. After that, the accuracy was calculated by combining the RBF and MLP algorithms. J48 and Random Forest were then combined to form a new hybrid method for classification. In comparison to RBF and MLP hybrid classification, this approach has a higher accuracy of 76.4583. Since J48 and Random Forest outperformed RBF and MLP, we deduced that they were the superior hybrid classification algorithms. Teachers may get a head start on knowing how their pupils are doing and how to help them study better according to this research. In Future additional hybrid classification methods may be studied to increase the accuracy of classification algorithm.

REFERENCES

[1] Margaret H.Dunham, “Data Mining Intoductory and Advanced Topics”, Pearson Education.

[2] Akagra Jain, Kushagra shah, Pradhyumn Chaturvedi” Prediction and Analysis of Student Performance using Hybrid Model of Multilayer Perceptron and Random Forest”,IEEE, 2018.

[3] Muhammad Sufyian Bin Mohd Azmi, lkmal Hisyam Bin Mohamad Paris, “Academic Performance Prediction Based on Voting Technique”, IEEE, 2011.

[4]. A.Dinesh Kumar, R.Pandi Selvam, V. Palanisamy, “Prediction of Student Performance using Hybrid Classification”, International Journal of Recent Technology and Engineering (IJRTE),Vol8,2019.

[5] Vladislav Miskovic, “Machine Learning of Hybrid Classification Models for Decision Support”, SinteZa, 2014.

[6] Thaddeus MatunduraOgwoka, Wilson Cheruiyot, George Okeyo, “A Model for Predicting Students’ Academic Performance using a Hybrid of K-means and Decision tree Algorithms”- International Journal of Computer Applications Technology and Research, Volume 4, 2019. [7] Akanksha Ahlawat, Bharti Suri,“ Improving Classification in Data mining using Hybrid algorithm”, IEEE, 2016.

[8] Hamza Turabieh, ”Hybrid Machine Learning Classifiers to Predict Student Performance”, IEEE, 2019.

[9] Bindhia ,K. Francis , Suvanam Sasidhar Babu, “Predicting Academic Performance of Students Using a Hybrid Data Mining Approach”, Journal of Medical Systems ,2019, Springer.

[10] LT Col Abhishek Lal, C.R.S Kumar, “Hybrid Classifier for Increasing Accuracy of Fitness Data Set”, IEEE, 2017.

[11] Ankita Dewan and Meghna Sharma, “Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification”, IEEE, 2015.

[12] Jai Ruby,K. David, “An Analysis on Academic Performance of Students using a Hybrid Model for Higher Education”, International Journal of Engineering and Technology (IJET) , 2017.

[13] Xiaofeng Ma, Zhurong Zhou, “Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree”, IEEE, 2018.

[14] Deepali et al "A hybrid Classification Model for prediction of academic performance of students: a big data Applications", Springer, 2019.

[15] BakerRSJd, Yacef K. "The state of educational data mining in 2009: A review and future visions". JEduData Min, 2009.