ISSN: 2835-7914

Research Article

Biomedical Science and Clinical Research

Hybrid Machine Learning Algorithms at the Service of Student Performance

Adil Korchi[1]* and Ahmed Abatal[2]

[1]*Chouaib Doukkali University, Faculty of Juridical, Economic and Social Sciences, El Jadida, Morocco*

[2]*Hassan Premier University, Faculty of Sciences and Techniques, Morocco*

***Corresponding Author**
Adil Korchi, Chouaib Doukkali University, Faculty of Juridical, Economic and Social Sciences, El Jadida, Morocco.

Submitted: 2023, Nov 16 **Accepted:** 2023, Dec 08 **Published:** 2023, Dec 12

Citation: Korchi, A., Abatal, A. (2023). Hybrid Machine Learning Algorithms at the Service of Student Performance. *Biomed Sci Clin Res, 2*(4), 409-417.

Abstract
The ability to alter and improve a student's status in order to get the greatest performance so that they pass their courses is an important component of today's educational landscape. This operation allows for the prediction of a student's performance in one or more disciplines. This has become possible nowadays through the use of Machine Learning algorithms that mine educational data to predict student performance by training the models and testing them with the available data set while using different algorithms.

In this study, we compared 9 algorithms in order to obtain the best models based on students' performance in well-defined disciplines in order to improve their results and success in their study. We started with the data collection and then we carried out a preprocessing process, after which, we built models to compare and evaluate them. After that, we compared the obtained results showed that the Random forest had the best ranking and this, in almost all the methods used monitored by SVM which had satisfactory results.

Keywords: Student's Performance, Machine Learning Algorithms, Dataset, Performance Metrics

1. Introduction

In order to forecast new data, machine learning algorithms create a predictive model using existing data. The question most often asked when developing a model is how to get better predictions namely: Improve performance through data, improve performance through machine learning algorithms, improve performance through model tuning and also, improve performance through set methods.

And for this, it is advisable to make changes to the data very often in order to obtain the greatest performance gains. Also, it is necessary to collect more data or even better data of better quality if it is possible because a model's ability to predict observations improves with the amount of training data it possesses. Then you have to think about cleaning the data in order to correct or delete missing or incorrect values in order to improve their quality [1]. Another point that is important is the projection of the data because a large number of input variables can lead to overfitting of the model and make it not generalizable on new data [2]. It is then advisable in this case to use methods to project the data into a lower dimensional space in order to reduce the model's overfitting.

Many choices of learning algorithms and their hyper parameters are available to Data Scientists. The nature of the problem to be solved partly guides this choice. Much comparative research in the world of student performance prediction using machine learning algorithms has been carried out to confirm the robustness of these algorithms. The machine can learn from previous studies and predict better results based on them.

2. Related Work

a) A conference paper by Maria Koutina with the title "Predicting Postgraduate Students' Performance Using Machine Learning Techniques" was published [3]. According to the author, using machine learning approaches to forecast student performance based on their backgrounds has proven to be a useful tool for predicting both excellent and bad performances at various educational levels. She choose five academic courses, each with its own dataset. Six well-known classification methods were tested on the selected data utilized in this study. The datasets utilized in this study were enhanced with attributes related to demographics, in-term performance, and classroom behavior. She addressed these problems for the performance prediction application using a variety of strategies, including resampling and feature selection. In her experimentation, the Nave Bayes and 1-NN produced the best prediction results, which were highly satisfying when compared to those of comparable techniques.

b) In a report published in 2021, Pallathadka (Pallathadka) addressed the kind of analysis that helps an institution minimize its failure rates. Based on their historical performance in comparable courses, his study makes predictions about students' achievement in a certain course. Data mining is a set of techniques used to find hidden patterns in enormous amounts of existing data. It can be useful for analysis and prediction.

c) Pallathadka used for categorization or prediction, Naîve Bayes, ID3, C4.5, and SVM, UCI machinery student performance data set was used in experimental study. Parameters like accuracy and error rate were also present in this research in order to analyze algorithms.

d) The study by employed a specific strategy for automatically observing and predicting student grades and marks [4]. By grouping students with the same academic session and educational history, it also seeks to improve classification accuracy and reduce root mean square error.

The attributes that the author used in his study contains students' confidential information and inapplicable attributes such as : Student's Address, Phone Number, Religion type, Roll_No …. This acquired data-set comprises 90000 historicals students' data.

e) The author used Decision Tree algorithm data-set comprising of 126 attributes as well as the KNN algorithm to get the best performance of students. Additionally, he determined the average accuracy attained by several categorization systems. The results were as follow:

a) Classifier	b) Accuracy
c) Decision-tree (DT)	d) 94.39
e) K-NN	f) 85.74
g) GA+DT	h) 96.64
i) GA+K-NN	j) 89.92

Table 1: Accuracy achieved by different classification algorithms

In the Industrial Engineering Department of Universitas Islam in Indonesia, Khasanah used feature selection to choose characteristics that had a significant impact on students' performance [5]. The Bayesian network and the decision tree, two well-known classification algorithms, were then put into use and contrasted to determine which produced the best prediction outcome. The outcome revealed that student attendance and GPA were the two feature selection techniques that performed best overall, with the Bayesian network outperforming the decision tree due to its greater accuracy rate.

Data cleaning, feature selection, and classification analysis were the three processes in this investigation. For the author, data cleaning was a crucial step since it enabled him to analyze the data effectively and find missing values, noisy data, and inconsistent data. The second one contributes to the feature selection algorithm by enabling the combination of a search method to suggest new feature subsets and a measurement that rates the various feature subsets [6]. Eliminating features that are irrelevant or provide less predictive information allows for the selection of a subset of input properties. This stage must be carried out in order to identify the key factors that will have a significant impact on predicting student achievement. Both Anuradha and Velmurugan and Ramaswami and Bhaskaran used a variety of techniques to choose the five common filter features for their studies.

By combining principal component analysis and machine learning algorithms, the author produced hybrid models [7]. He first presented the baseline models, and then used k-fold cross-validation to enhance the performance of the baseline models. Finally, he suggested the hybrid machine learning model as shown in Fig. 1 by fusing it with principal component analysis.

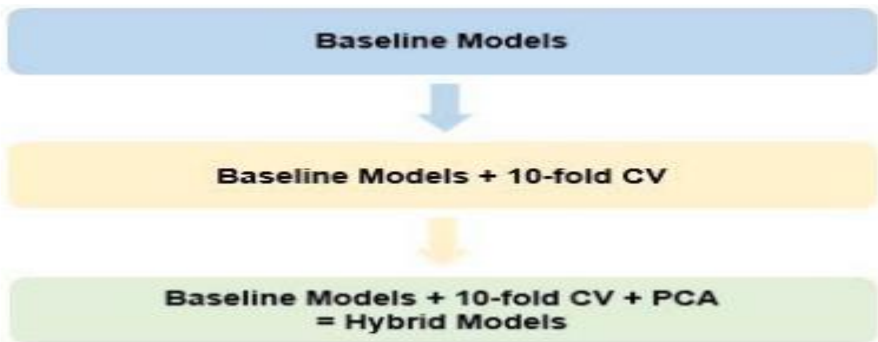

Figure 1: Illustration of Task Procedure

The author of this study lists all covert aspects that affect pupils' math performance. 43 attributes that detailed the details of each student's learning activities were included in the datasets, along with one objective variable that described the performance levels of students depending on their score. The predicted traits are those observed from the three main factors that were affected. The author of the essay displayed the 43 variables that make up these primary components. Also described were the predefined classes of the target variable.

He tested the robustness and effectiveness of our suggested methods using three datasets. The first two datasets, GDS1 (2000 samples) and GDS2 (4,000 samples), were developed utilizing recommended architectural layouts of predictive properties for the output variable.

3. Reseach Method

Several Machine Learning algorithms were used to make possible this research. The used hybrid algorithms are namely : ANN (Artificial neural network), DT (Decision tree), ELM based Model, KNN (K-nearest neighbour), LR (Logistic regression), LR 1 (Linear regression), NB (Naïve Bayes), RF (Random forest) and finaly SVM (Support vector machine).

The following diagram (Fig 2) summarizes the main steps of the methodology used:

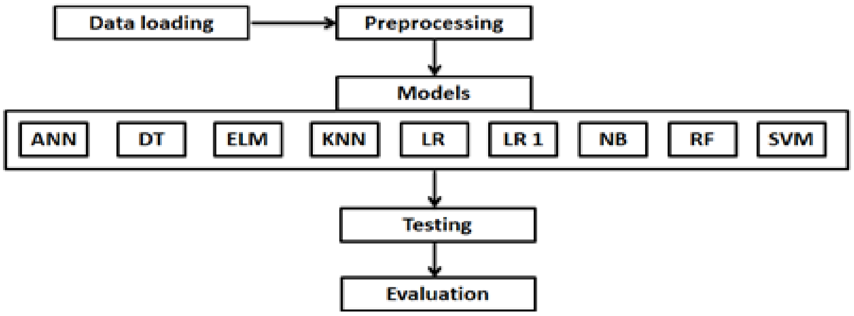

Figure 2: Steps of the used methodology

Pre-processing is done to improve the data once it has been loaded before building a machine learning model. In this section, we've cleaned the data to get rid of attributes that clog up and slow down learning. We've also dealt with missing data to enhance the effectiveness of the models we developed. Usually, when we discover different properties or missing values, etc., problems with the data start to occur. Separating the data into training and testing is another method of pre-processing.

Once the data was cleaned, we proceeded to the construction of the different learning models namely ANN, DT, ELM, SVM..... The training data is used to create models. The next step is to evaluate the different models built on a separate test set [8]. The final phase involves assessing the results in order to choose the optimal algorithm to forecast students' performance based on the model created.

4. Experimental Part

4.1 Dataset Information

The following description of Dataset gives an overview of the titles, their descriptions, the data type variables (Categorical or Continuous) as well as their categorization (binary, nominal, numeric).

These statistics represent an approximation of secondary student performance in two Portuguese schools. Student grades, demographic information, social features, and school characteristics are among the data attributes that were gathered via reports and surveys' school. There are two datasets available on performance in two distinct subjects: Portuguese and Mathematics (mat) (por).

The two datasets were modeled using binary/five-level classification and regression tasks in [9]. It should be noticed that the goal attribute G3 and the qualities G2 and G1 have a significant association. This is so that G1 and G2 correspond to the grades for the first and second periods, whereas G3 is the final annual grade (given in the third period). Without G2 and G1, it is more challenging to forecast G3, yet this prediction is considerably more valuable (see paper source for more details).

No	Feature Title	Description	Variable Data Type	Feature Categorization
1	school	student's- school	Categorical	binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira
2	sex	student's- sex	Categorical	binary: 'F' - female or 'M' - male
3	age	student's- age	Continuous	from 15 to 22
4	address	student's- home address type	Categorical	(binary: 'U' - urban or 'R' - rural
5	Fam-size	family size	Categorical	binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3
6	P-status	parent's- cohabitation-status	Categorical	binary: 'T' - living together or 'A' - apart
7	M-edu	mother's- education	Categorical	numeric: 0 - none, 1 - primary education (4th grade), 2- 5th to 9th grade, 3-secondary education or 4- higher education)
8	F-edu	father's- education	Categorical	(numeric: 0 - none, 1 – primary-education (4th grade), 2- 5th to 9th grade, 3- secondary-education or 4- higher-education)
9	M-job	mother's- job	Categorical	(nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
10	F-job	father's- job	Categorical	(nominal: 'teacher', 'health' care related, civil-'services' (e.g. administrative or police), 'at_home' or 'other')
11	reason	Reason- to- choose- this -school	Categorical	(nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12	guardian	student's- guardian	Categorical	(nominal: 'mother', 'father' or 'other')
13	traveltime	Home- to- school- travel- time	Categorical	(numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14	Study-time	Weekly- study- time	Categorical	(numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15	failures	Number- of- past- class- failures	Categorical	(numeric: n if 1<=n<3, else 4)
16	School-sup	extra educational support	Categorical	(binary: yes or no)
17	Fam-sup	Family- educational- support	Categorical	(binary: yes or no)
18	paid	Extra- paid- classes- within- the course subject	Categorical	(Portuguese or Math) (binary: yes or no)
19	activities	extra-curricular activities	Categorical	(binary: yes or no)
20	nursery	Attended- nursery school	Categorical	(binary: yes or no)
21	higher	Wants- to- take higher- education	Categorical	(binary: yes or no)
22	internet	Internet- access- at- home	Categorical	(binary: yes or no)
23	romantic	With- a- romantic- relationship	Categorical	(binary: yes or no)
24	famrel	Quality- of- family relationships	Categorical	(numeric: from 1 - very bad to 5 - excellent)

25	freetime	Free- time- after- school	Categorical	(numeric: from 1 - very low to 5 - very high)
26	goout	Going- out- with- friends	Categorical	(numeric: from 1 - very low to 5 - very high)
27	Dalc	Workday- alcohol- consumption	Categorical	(numeric: from 1 - very low to 5 - very high)
28	Walc	Weekend- alcohol consumption	Categorical	(numeric: from 1 - very low to 5 - very high)
29	health	Current- health-status	Categorical	(numeric: from 1 - very bad to 5 - very good)
30	absences	Number- of- school- absences	Continuous	(numeric: from 0 to 93)
These grades are related with the course subject, Math or Portuguese:				
31	G1	First- period- grade	Continuous	(numeric: from 0 to 20)
32	G2	Second- period- grade	Continuous	(numeric: from 0 to 20)
33	G3	Final- grade	Continuous	(numeric: from 0 to 20, output target)

Table 2: Dataset informations (Cortez and Silva, 2008)

3. Performance Metrics (Assessment Metric)

The performance of a classification model directly depends on the metric used to evaluate it. For each task, the evaluation metric is different. It is important to employ a metric that best captures the requirements of the problem. To better evaluate and compare algorithms, it is important to clearly define a simple model used as a benchmark to compare the performance of your machine learning models.

An evaluation metric quantifies the performance of a predictive model. Choosing the right metric is therefore crucial when evaluating machine learning models and the quality of a classification model directly depends on the metric used to evaluate it. It provides an overview of correct and incorrect predictions.

To give more value to our study, we used several performance measurement criteria. In the field of machine learning, there are several metrics. We have opted for the best known in this field, namely: Mean Absolute Error, Root Mean Squared Error (RMSE), Accuracy (Acc), The area under the roc curve (AUC), Relative Absolute Error (RAE), Relative Squared Error (RSE).

• **Mean Absolute Error (MAE):** is most commonly used evaluation metric for regression problems. It tries to calculate the difference between the actual and the predicted values. This difference is the error. The formula for calculating the MAE is as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \qquad 1$$

Is thus an arithmetic mean of the absolute errors |**yi** – **xi**|, where yi is the prediction and xi is the true value. **n** stands for the number of observations.

• **Root Mean Square Error (RMSE):** or Root Mean Square Deviation is one of the most commonly used measures to assess the quality of forecasts. It shows how far predictions deviate from measured true values using Euclidean distance.

For each data point, the RMSE calculates the residual (difference between forecast and truth), the norm of the residual, the mean of the residuals, and the square root of that mean. Because it uses and necessitates real measurements at each projected data point, the RMSE is frequently utilized in supervised learning applications. The following is an expression for the root mean square error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}} \qquad 2$$

The number of data points in this example is N, the ith measurement is y(i), and the related prediction is ^y(i).

Because RMSE is not scale invariant, the scaling of the data has an impact on how models are compared using this measure. This is why RMSE is frequently chosen over normalized data.

• **Accuracy (Acc):** The proportion of occurrences that were correctly categorised in relation to all instances. The most used metric for assessing the effectiveness of classifiers is this one. This formula is used to compute it:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \qquad 3$$

• **The Area Under the ROC Curve (AUC)** provides a summary of the model's performance for all conceivable thresholds by measuring how well the positive examples align ahead of the negative cases. It can be thought of as the likelihood that a random positive example will be ranked higher than a random negative example by the model.

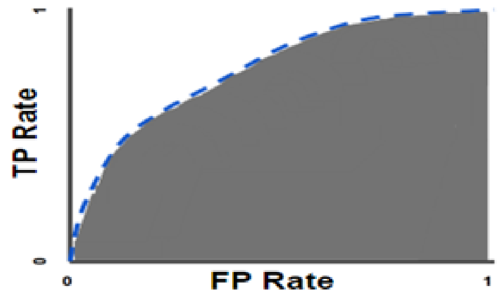

Figure 3: AUC (Area under the ROC curve

• **Relative Absolute Error (RAE):** This statistic can be used to assess how well a prediction model is performing. It is mostly utilized in the disciplines of operations management, data mining, and machine learning. Relative error, a generic indicator of the precision or accuracy of tools like clocks, rulers, or scales, should not be confused with RAE.
The ratio of a mean error (residual) to errors produced by a simplistic or naive model is used to indicate the relative absolute error. A ratio less than one is produced by a credible model (one that outperforms a trivial model).

• **Relative Squared Error (RSE):** This straightforward metric is used to assess how well regression and other predictive models work. Even though this metric is simple, it's an excellent place to start when assessing the quality of a model. The popular R-squared metric can also be simply calculated using the RSE. The following is its formula:

$$RSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{y})^2} \qquad \boxed{4}$$

- n: Number of observations
- x_i: Realized value
- y_i: Predicted value
- $\bar{y}$: Average of the realized values

4. Classification Results

The model's predictions were evaluated by five cross-validations. We randomly selected 80% of the student data into the training set and 20% of the student data into and during the test set on each run. Table 2 below presents the evaluation models, we compare the real values with the predicted ones for each student.

Algorithms / Metrics	MAE	RMSE	ACC	AUC	RAE	RSE
ANN (Artificial neural network)	0.35	0.32	76.58	70.82	0.36	0.34
DT (Decision tree)	0.22	0.25	76.51	72.66	0.19	0.24
ELM based Model	0.25	0.26	72.33	70.23	0.27	0.28
KNN (K-nearest neighbour)	0.28	0.30	77.92	74.72	0.33	0.29
LR (Logistic regression)	0.29	0.26	80.77	74.71	0.28	0.30
LR 1 (Linear regression)	0.33	0.35	69.40	65.78	0.31	0.34
NB (Naïve Bayes)	0.33	0.30	79.43	74.52	0.34	0.33
RF (Random forest)	0.15	0.19	86.53	78.50	0.22	0.21
SVM (Support vector machine)	0.17	0.19	76.92	73.83	0.23	0.22

Table 3: Comparison between the different algorithms according to different metrics

To find out which algorithm best predicts student performance, we compared 9 different machine learning algorithms namely: ANN (Artificial Neural Network), DT (Decision Tree), ELM based Model, KNN (K-nearest neighbor), LR (Logistic Regression), LR 1 (Linear Regression), NB (Naïve Bayes), RF (Random forest) and finally SVM (Support vector machine). This operation allowed us to know which algorithm had the best performance.

The following Figure (Fig 4.) shows the ACC and MAE performance according to the 9 algorithms used.

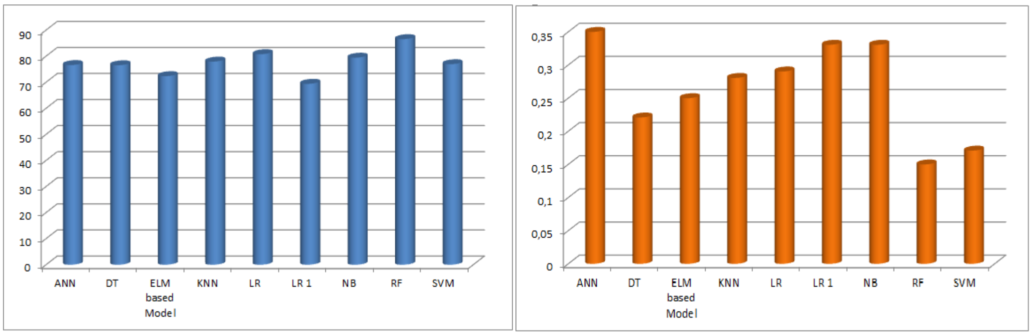

Figure 4: Performance of the ACC and MAE according to the used algorithms

5. Ranking of Differents Methods

For more visibility and based on the results obtained in Table 2, we have classified the scores obtained by each algorithm according to the 6 chosen metrics.

The following table (Tab 3) gives an overview of the performance ranking of algorithms according to the different metrics used.

Algorithms / Metrics	MAE	RMSE	ACC	AUC	RAE	RSE
ANN (Artificial neural network)	9	8	3	6	5	5
DT (Decision tree)	3	3	7	9	9	9
ELM based Model	4	4	8	4	4	4
KNN (K-nearest neighbour)	5	6	2	5	7	8
LR (Logistic regression)	6	5	4	8	8	7
LR 1 (Linear regression)	7	9	6	3	1	1
NB (Naïve Bayes)	8	7	9	7	6	6
RF (Random forest)	1	1	1	1	2	2
SVM (Support vector machine)	2	2	5	2	3	3

Table 4: Performance ranking of algorithms

6. Exécution Time

Good performance on the training dataset does not always guarantee good results on new data. We must therefore seek to model an analysis that reflects the complexity of the nature of the data by avoiding the problems of underfitting and overfitting.

Algorithms / Metrics	Execution Time	
	Time taken for Training (s)	Time taken for Classification (s)
ANN (Artificial neural network)	6,345	0,048
DT (Decision tree)	0,444	0,141
ELM based Model	102,135	1,325
KNN (K-nearest neighbour)	1,464	0,567
LR (Logistic regression)	2,236	1,476
LR 1 (Linear regression)	2,362	1,521
NB (Naïve Bayes)	23,567	1,167
RF (Random forest)	4,976	0,780
SVM (Support vector machine)	545,881	0,153

Table 5: Overview of algorithms' execution time

It is also important to define a machine learning model that has a suitable computation time and a reasonable use of memory resources so that the analysis algorithm is usable.

The following table (Tab. 5) gives an overview of the execution time of our 9 algorithms chosen to know which of them records a good performance with respect to the couple: result obtained / execution time.

7. Results Analysis

The accuracy results give different results than previous comparisons (Tab 2). The Random Forest algorithm has the best student prediction accuracy (86%) compared to logistic regression (80%) and Naïve Bayes (79%). Recorded KNN (78%), SVM (77%) while ANN (76%), decision tree (76%), ELM-based model (72%) and finally, regression linear (69%).

8. Regarding the Results for the Area Under the Roc Curve (AUC), the Results were as Follows

The Random Forest algorithm has the best prediction accuracy (78%) compared to KNN (74%) and logistic regression (74%). Naïve Bayes obtained (74%), SVM (73%) while decision tree (72%), ANN (70%), ELM-based model (70%) and finally, linear regression (65%).

From the result obtained in Table 2, we can say that Random forest remains the best in terms of errors and accuracy of predictions followed by SVM which was a little less efficient than RF. These results are totally logical because these 2 algorithms are the most efficient in almost the majority of classification problems [10].

We also notice that DT has the best training and classification time followed by KNN and therefore, it remains the best place for this operation as long as their construction process is simple. Logistic regression, linear regression, SVM and Random forest performed less well than DT and KNN.

9. Conclusion

In this comprehensive study, our primary aim was to examine and compare the effectiveness of nine widely recognized algorithms, with the goal of identifying optimal models for predicting and enhancing students' performance in carefully defined academic disciplines. The overarching objective was to contribute meaningfully to the advancement of educational outcomes, creating an environment conducive to heightened success in students' academic pursuits. The initial phase of our investigation focused on meticulous preprocessing of gathered data, a critical step pivotal in laying the foundation for constructing, comparing, and evaluating various predictive models. Through the adoption of a robust preprocessing strategy, we ensured that the data input into the algorithms was refined and standardized, thereby facilitating a fair and accurate assessment of their respective performances.

Upon executing the diverse algorithms selected for this research, a noteworthy outcome emerged—the Random Forest algorithm demonstrated unparalleled student prediction accuracy, reaching an impressive rate of 86% and surpassing all other algorithms under consideration. This finding not only holds promise for the practical application of the Random Forest algorithm in predicting students' academic success but also underscores its superiority in overall performance metrics. With a prediction accuracy of 78%, it outshone its counterparts, showcasing a robust capacity for precisely forecasting students' outcomes. Additionally, the algorithm exhibited the most favorable training and classification times, affirming its efficiency in navigating the complexities of the dataset. In conclusion, this study contributes valuable insights to educational research by establishing the Random Forest algorithm as an optimal choice for predicting student performance. The robustness of its predictions, along with efficient training and classification times, positions it as a powerful tool for educators and policymakers aiming to implement data-driven strategies to enhance educational outcomes, extending the implications of this research beyond mere algorithmic comparison [11-22].

References

1. Wiyono, S., Abidin, T., Wibowo, D. S., Hidayatullah, M. F., & Dairoh, D. (2019). Comparative study of machine learning knn, svm, and decision tree algorithm to predict students performance. International Journal of Research-Granthaalayah, 7(1), 190-196.
2. Song, J., Gao, Y., Yin, P., Li, Y., Li, Y., Zhang, J., ... & Pi, H. (2021). The random forest model has the best accuracy among the four pressure ulcer prediction models using machine learning algorithms. Risk Management and Healthcare Policy, 1175-1187.
3. Koutina, M., & Kermanidis, K. L. (2011, September). Predicting postgraduate students' performance using machine learning techniques. In International Conference on Engineering Applications of Neural Networks (pp. 159-168). Berlin, Heidelberg: Springer Berlin Heidelberg.
4. Yousafzai, B. K., Khan, S. A., Rahman, T., Khan, I., Ullah, I., Ur Rehman, A., ... & Cheikhrouhou, O. (2021). Student-performulator: Student academic performance using hybrid deep neural network. Sustainability, 13(17), 9775.
5. Khasanah, A. U. (2017, June). A comparative study to predict student's performance using educational data mining techniques. In IOP Conference Series: Materials Science and Engineering (Vol. 215, No. 1, p. 012036). IOP Publishing.
6. Anuradha, C., & Velmurugan, T. (2016, January). Feature selection techniques to analyse student acadamic performance using Naïve Bayes classifier. In The 3rd international conference on small & medium business (pp. 345-350).
7. Sokkhey, P., & Okazaki, T. (2020). Hybrid machine learning algorithms for predicting academic performance. Int. J. Adv. Comput. Sci. Appl, 11(1), 32-41.
8. Mothilal, R. K., Sharma, A., & Tan, C. (2020, January). Explaining machine learning classifiers through diverse counterfactual explanations. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 607-617).
9. Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.
10. Teles, G., Rodrigues, J. J., Rabelo, R. A., & Kozlov, S. A. (2021). Comparative study of support vector machines and random forests machine learning algorithms on credit operation. Software: Practice and Experience, 51(12), 2492-2500.
11. Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017, April). Predicting student performance using advanced learning analytics. In Proceedings of the 26th international conference on world wide web companion (pp. 415-421).
12. Sa, C. L., Hossain, E. D., & bin Hossin, M. (2014, November). Student performance analysis system (SPAS). In The 5th International Conference on Information and Communication Technology for The Muslim World (ICT4M) (pp. 1-6). IEEE.
13. Singh, I., Sabitha, A. S., & Bansal, A. (2016, January).

Student performance analysis using clustering algorithm. In 2016 6th International Conference-Cloud System and Big Data Engineering (Confluence) (pp. 294-299). IEEE.
14. Wiyono, S., Abidin, T., Wibowo, D. S., Hidayatullah, M. F., & Dairoh, D. (2019). Comparative study of machine learning knn, svm, and decision tree algorithm to predict students performance. International Journal of Research-Granthaalayah, 7(1), 190-196.
15. Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings, 80, 3782-3785.
16. Nugrahaeni, R. A., & Mutijarsa, K. (2016, August). Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification. In 2016 International Seminar on Application for Technology of Information and Communication (ISemantic) (pp. 163-168). IEEE.
17. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & education, 113, 177-194.
18. Conijn, R., Van den Beemt, A., & Cuijpers, P. (2018). Predicting student performance in a blended MOOC. Journal of Computer Assisted Learning, 34(5), 615-628.
19. Devasia, T., Vinushree, T. P., & Hegde, V. (2016, March). Prediction of students performance using Educational Data Mining. In 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE) (pp. 91-95). IEEE.
20. Patil, V., Suryawanshi, S., Saner, M., Patil, V., & Sarode, B. (2017). Student performance prediction using classification data mining techniques. International Journal of Scientific Development and Research, 2(6), 163-167.
21. Paul, Y., Goyal, V., & Jaswal, R. A. (2017, September). Comparative analysis between SVM & KNN classifier for EMG signal classification on elementary time domain features. In 2017 4th international conference on signal processing, computing and control (ISPCC) (pp. 169-175). IEEE.
22. Zhang, J., & Feng, S. (2021). Machine learning modeling: A new way to do quantitative research in social sciences in the era of AI. Journal of Web Engineering, 281-302.