Original Article

Student Performance Analysis using Bayesian Optimized Random Forest Classifier and KNN

Safira Begum[1], Sunita S Padmannavar[2]

[1]VTU-RRC, Belagavi, Karnataka,
[2]Gogte Institute of Technology, Belagavi, Karnataka, India

[1]Corresponding Author : safirabgm@gmail.com

Abstract - *The importance of Educational Data Mining (EDM), a new interdisciplinary study field that builds on various other disciplines, is rising. It is directly connected to data mining (DM), which is crucial to finding knowledge in databases (KDD). This data is expanding exponentially and may include valuable hidden information for users (both teachers and students). Such knowledge can easily be recognized as models, patterns, or any other type of representational scheme that enables improved system exploitation. It is discovered that data mining can be used to make similar discoveries, giving rise to EDM. To get the best outcomes in this complicated setting, many approaches and learning algorithms are typically applied. In recent times, educational systems have witnessed a surge in using artificial intelligence (AI) systems, especially for extracting relevant information. One such AI system is EDM, which combines various techniques to support the capture, processing, and analysis of these record sets. The primary method used in EDM is machine learning, which has been applied more frequently since the emergence of big data to extract useful information from a vast amount of data. Machine learning has been used for decades in data processing in various contexts. Educational data mining tools and algorithms can be used to assess student academic achievement. This study offers a fresh approach to forecasting student success in middle school Portuguese and mathematics subjects. Hyperparameter tuning of classifiers is essential to overcome the misclassification of conventional classifiers. In order to predict student performance in the UCI dataset, this work proposes a Bayesian-optimized KNN and a random forest classifier. For random forest and KNN, the attained accuracy is 87% and 73%, respectively.*

Keywords - Bayesian Optimization, Educational Data Mining, KNN, RF, UCI.

1. Introduction

Recent technological developments have resulted in the instrumentalization of the education sector, including the use of the Internet for learning, the teaching of software, and the computerised management of student performance by administrators of educational institutions. The amount of educational data has increased exponentially as a result of all these variables, particularly due to the popularity of e-learning, and computing resources are required to evaluate such a massive amount of data. Otherwise, the task will become impossible [1].

As such, data mining techniques are becoming increasingly important in education as a way to monitor, analyze and evaluate the learning process [2]. Data mining techniques have the potential to provide educational policymakers with models that support the goal of improving classroom efficiency and performance [2]. Additionally, many machine learning methods, such as individualized learning environments and solutions that contain effective decision-making frameworks, might have a good impact on resolving

certain issues in educational institutions and can be viewed as the foundation for a significant change in the way things are produced in educational institutions [1] [2] [3] [4]. In this situation, educational data mining (EDM) stands out because it uses (DMDM) approaches to draw out pertinent data from a variety of educational datasets. The International Educational Data Mining Society defines this subject as one that focuses on exploring distinctive and increasingly rich data from educational contexts and developing techniques to use these techniques better to understand students and the learning environments [5].

In other words, DM designates a collection of computational techniques for identifying facts among vast amounts of data. When the analyzed data comes from an academic context, it is called EDM [3]. DM is also described by the author of [6] as a field concerned with investigating data from educational settings and utilising them to create strategies for better-comprehending teaching and learning processes. According to the authors of [7], EDM is a branch of study that aims to advance and develop methods for

analysing data sets gathered in educational contexts. The authors claim that the nature of this data is more varied than the data typically used for mining operations, necessitating modifications and novel strategies. The variety of this data also has the potential to be a significant resource for enhancing education [6] [7] [8].

Most importantly, EDM contributes significantly to the forecasting of student performance. It seeks to ascertain how learners perform during the course so that required interventions can be made to enhance learning [9]. The authors of [9] mentioned 18 papers on this subject in the mapping above. After behavioural analysis, these were the topics that researchers examined the most. This is due to the fact that student performance is a crucial component of a school since it is one of the factors used to determine which colleges and universities have the highest academic performance records [10]. In order to help teachers and students improve their teaching and learning processes, the authors of [10] claim that forecasting student success can be quite beneficial.

EDM is a field that utilizes data mining techniques to extract relevant data from various educational datasets. It aims to understand students and learning environments better and develop models that support educational policymakers in enhancing classroom performance. EDM is also crucial in forecasting student performance, which is a crucial component in determining the academic success of institutions. However, existing studies mainly use consolidated EDM approaches, and there is a need to explore the use of machine learning methods in predicting student performance. This study aims to address this research gap by comparing the performance of machine learning methodology with previously consolidated EDM approaches. The study employs a common dataset from the UCI machine learning repository and uses KNN and random forest classifiers with Bayesian optimization to predict student performance. The study also investigates the adequacy of the database's characteristics in creating effective models for forecasting student performance.

This study's goal is to predict student performance in this situation. To do so, it employed a common dataset from the UCI machine learning repository, which allowed it to compare machine learning methodology with previously consolidated EDM approaches. The classification approach utilised was supervised learning [11]; rather than using the students' numerical grades, the grades were predicted but divided into four groups. With the use of these forecasts, it was also possible to determine whether the database's characteristics were adequate for creating effective models for forecasting student performance. In the suggested method, KNN and random forest classifiers were used to predict student performance. In addition, Bayesian optimization was employed to adjust the classifiers' hyperparameters for properly supervised classification training.

Following is the structure of this paper: A review of the literature is shown in section 2, a proposed approach for KNN and RF classifiers is presented in section 3, the study's results are presented in section 4, and the authors' conclusions are presented in section 5.

2. Literature Review

For a long time, EDM has been the focus of intensive investigation. To improve the precision of student data prediction, the authors of [12] combined an ensemble model with a number of categorization techniques (DT, RF and NB). In order to find association rules that have an impact on student performance, they also explored a number of rule-based methodologies [37]. The authors of [14] investigate how different classification methods are affected by data preparation on a dataset of student academic performance with a skewed classroom distribution. To solve the uneven class distribution issue in this direction, undersampling and oversampling strategies are utilised for the SVM, DT, and NB classification algorithms. In the experimental results, the SMOTE method, which is a member of the oversampling class, gets higher accuracy values [15].

The authors [16] also found that Machine Learning was the most widely used technique and that in terms of algorithm effectiveness, Neural Networks had the highest precision (98%) for predicting student performance, followed by Decision Trees (91%), Support Vector Machines and KNN with the same efficiency (83%) and Naive Bayes with the least (76 percent).

The authors of [17] Introduced a novel model in an optimised support vector machine to analyse students' learning profiles and engagement. The authors compared it to widely used algorithms – Linear Regression and Support Vector Machines – to validate its accuracy. In addition, [18] conducted research to predict student performance using a Learning Neural Network Profundo and compared it to Logistic Regression algorithms [38].

The feature selection method is used by the authors of [20] to estimate the performance of classifiers using the decision tree for the student data set. The best accuracy values are obtained by the RF algorithm [20]. The authors of [21] use the NB classification algorithm on the student dataset to assess the students' success in the courses. The author [22] assessed the DT and SVM algorithms using 10-fold cross-validation on the student performance dataset. They used a grid search technique for hyperparameter tuning to enhance the algorithm's performance [22]. The authors of [23] use regression, binary classification, and multilevel classification to analyse student datasets. For comparison, these methods make use of k-means, nearest neighbours, support vector machines, and naive Bayes algorithms [23]. In their analysis, the authors of [24] used SVM, multilayer perceptrons, and RF algorithms to predict the readiness of students to enter college.
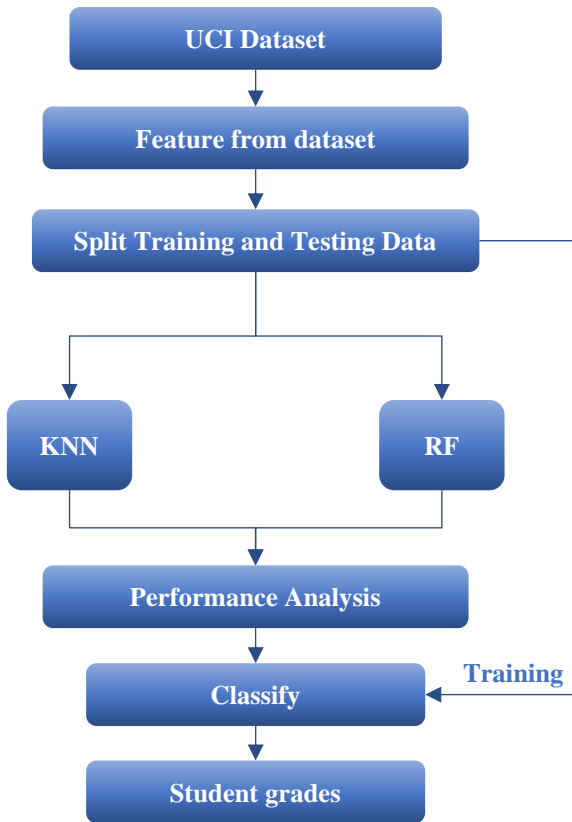
Fig. 1 Flow diagram of the proposed research

According to experimental findings, the RF algorithm is more effective than other algorithms [24]. The authors of [39] predicted student performance using classifiers like NB. They came to the conclusion that the student's school and work schedules significantly affected their final grades [39].

The best effective model for forecasting student performance was discovered by the authors [26] using k-nearest neighbours and a decision tree algorithm. The decision tree algorithm has a greater performance value, according to experimental findings [26]. The impacts of a quick correlation-based filtering strategy on SVM classification algorithms on three separate student performance record sets were examined by the authors of [27]. Feature selection methodology enhances models for predicting student academic achievement, according to experimental analysis [27]. Turkish language and mathematics classes, as well as the overall average success rates at the end of the semester, were evaluated by the authors of [28] in relation to the living situations and social milieu of students. Course grades were calculated using a classification method, and student performance ratings (0–100) were calculated using a regression method. According to experimental findings, the logistic-based classification algorithm employing the RF regression approach for assessing course grades and the correlation-based feature subset method for predicting course grades both produced good results [28].

To examine the efficacy of several categorization algorithms on student academic performance data, the authors of [40] use Weka open-source software. They found that particular attributes have a stronger impact on student performance during the data pre-processing stage and that the classification accuracy increases after this stage [30]. The authors of [31] employ decision tree-based classification algorithms [32] to predict student academic success. On two independent student record sets, the authors of [33] employed linear regression, decision trees, and naïve bayes algorithms to predict student performance. According to the experimental findings, the feature selection strategy increased the classification algorithm's accuracy value for both record sets. In order to forecast a student's performance level on a student success dataset.

3. Proposed Method

Through EDM, data gathered in the educational system is turned into useful information. EDM is widely employed to assess student performance. This is accomplished using feature extraction, feature selection, and classification techniques. In this work, the performance of higher secondary school students was assessed using K-NN and RF (Random Forest) classifiers, which were then further enhanced using Bayesian optimization. The issue of an unbalanced distribution of classes must be fixed in order to increase classification success. The research part goes over the aforementioned approaches. Higher secondary education in Portugal consists of three years of study, nine years of elementary school, and higher education. Most students enrol in public schools for a number of reasons.

3.1. K-Nearest Neighbors (KNN)

One of the simplest learning algorithms is the KNN algorithm, which is based on lazy learning. In other words, there is either no explicit training phase or a very short one. This indicates that the phase training is quite fast.

The data must exist in a space of features, according to the KNN algorithm. The data points are in metric space as a result. Scalars or even multidimensional vectors can make up the data. Regression and classification are done using this technique. In both situations, the entry is made from the k training data closest to the feature space. This algorithm uses the following approach to determine the class of a new case: it looks for the k nearest neighbours of the new case, then selects the candidate that is both closest to the original case and most frequently encountered. The method looks for the k nearest neighbours among the individuals who have already been classed to assign a new individual to a class. Consequently, the individual is put in the class.

3.1.1. Principle of the Algorithm

Consider a set E containing n labelled data: $E = \{(y_i, \vec{x}_i)\}$ with i included between 1 and n, where y_i corresponds to the class (the label) of the datum i and where

the vector $\vec{x}_i$ of dimension, p represents the predictor variables of datum i. $\vec{x}_i = (x_i^1, x_i^2, \ldots, x_i^p)$. Consider a data u which does not belong to E and which does not have a label (u is only characterized by a vector $\vec{x}_u$ of dimension p). Consider a function d which returns the distance between the datum u and any datum belonging to E, and K an integer less than or equal to n [28].

To apply this method, the steps to follow are as follows:
- Calculate the distances between the datum u and each datum belonging to E using the function d.
- Keep the k data of the data set E closest to u.
- Attribute to u the class which is the most frequent among the k nearest data.

The algorithm calculates the closest cases based on the distances between them (examples, vectors). The performance of the method depends heavily on the distance metric used. Examples of the most popular metrics include the Manhattan distance and the Euclidean distance:

$$d(p,q) = \sqrt{\sum_{i-1}^{N}(q_i - p_i)^2} \quad (1)$$

$$d(p,q) = \sum_{i-1}^{N}|q_i - p_i| \quad (2)$$

Where equation (1) represents Euclidean distance and equation (2) represents Manhattan distance.

3.2. Random Forest Classifiers

Random forests are composed (as the term "forest" implies) of a set binary decision tree in which randomness has been introduced. These trees distinguish from each other by the subsample of data on which they are trained. These sub-samples are drawn at random (hence the term "random") from a dataset. The random forest technique modifies the Bagging method used here for the trees by adding a decorrelation criterion between these trees. This approach aims to lessen correlation without substantially raising variance. The idea is to select a subset of factors at random that will be taken into account when determining which node on the tree is the best at each level.

3.2.1. Principle and Algorithm of Random Forests

Let S be a training set containing m instances of the student dataset, represented as $S = \{x_i | y_i \ldots \ldots \ldots x_m, y_m\}$. Additionally, let S_t be a bootstrap sample consisting of m instances obtained by resampling S with replacement. Let h denote a set of T decision trees denoted by H_t, each of which is constructed from S_t. To create each node in the tree, an attribute is selected for partitioning by randomly selecting a subset of attributes. Finally, a new instance is classified using a random forest classifier, which employs a uniformly weighted majority vote of the classifiers in the set h. This principle is demonstrated by the algorithm.

Pseudocode
randomforest (S,T)
Entrance: $S = \{x_i | y_i \ldots \ldots \ldots x_m, y_m\}$, the training set.
Input: T the number of random forest decision trees.
For $t = 1, \ldots \ldots T$
1. Generate Bootstrap size S_t sample m from S.
2. By reiterating the process, create a decision tree.
Recursively, do the following actions for each node of the tree:
a) Randomly pick attributes among the attributes
b) Select the partitioning attribute among the
c) Divide the node into two child nodes
End for
Output: the RF classifier

3.3. Bayesian Optimization for Random Forest Classifier

The objective function for Bayesian optimization of the random forest classifier can be defined as the cross-validation accuracy on a training dataset. Let X be the space of hyperparameters and Y be the space of cross-validation accuracies. The objective function $f: X \rightarrow Y$ can be defined as:

$$f(x) = mean\ (cross\ validation\ accuracies\ using\ hyperparameters\ x) \quad (3)$$

The hyperparameters "x" might consist of various factors such as the count of trees in the forest, the highest depth limit of each tree, and the least amount of samples that are necessary to split an inner node. The Gaussian process model is used to model the objective function f. The mean function $\mu(x)$ and covariance function $k(x, x')$ are defined as:

$$\mu(x) = 0, k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right) \quad (4)$$

Where, l is the length scale parameter that controls the smoothness of the function.

The acquisition function for Bayesian optimization of the random forest classifier can be defined as expected improvement (EI). The expected improvement of a hyperparameter x is defined as:

$$EI(x) = E[\max(0, f(x) - f(x_{best}) - \xi)] \quad (5)$$

Where, x_{best} is the hyperparameter with the highest observed accuracy so far, and ξ is a trade-off parameter that controls the exploration-exploitation trade-off. The acquisition function is then defined as:

$$a(x) = E[EI(x)] \quad (6)$$

Where, the expectation is taken with respect to the posterior distribution over functions.

The next hyperparameter to evaluate is then chosen by maximizing the acquisition function:

$$x_{next} = \arg\max a(x) \quad (7)$$

The optimization process continues until a stopping

criterion is met, such as a maximum number of evaluations or a threshold on the expected improvement.

3.4. Bayesian Optimization for KNN Classifier

The Bayesian-optimized KNN classifier is a machine learning algorithm used in student performance analysis. The KNN algorithm is a type of instance-based learning where the model predicts the class of a new instance by searching through the training set for the K-nearest neighbors and assigning the class of the majority of the neighbors to the new instance.

The KNN algorithm necessitates choosing the number of nearest neighbors (K) for prediction. Choosing the appropriate value of K is crucial for the model's performance. In this study, the Bayesian optimization approach is utilized to determine the optimal K value that maximizes the KNN classifier's performance.

The Bayesian optimization algorithm leverages Bayesian inference to develop a probabilistic model of the function to be optimized. The algorithm establishes a surrogate model approximating the true objective function by modeling the relationship between the hyperparameters and the objective function. The surrogate model is updated with each new observation, and the algorithm employs the updated model to select the next set of hyperparameters to evaluate. This process continues until the algorithm converges to the optimal set of hyperparameters.

The Bayesian optimization algorithm is mathematically represented as follows: Given a set of hyperparameters θ and a set of training data D, the objective is to identify the hyperparameters that optimize the KNN classifier's performance. The KNN classifier's performance is measured using a metric such as accuracy or F1-score. The objective function $f(\theta)$ represents the metric's value for a given set of hyperparameters θ.

$$f(\theta) = performance\ metric\ for\ KNN\ classifier\ with\ hyperparameters\ \theta \quad (8)$$

The surrogate model used in Bayesian optimization is a Gaussian process that models the objective function as a distribution over functions. The surrogate model is defined as follows:

$$y = f(\theta) + \varepsilon \quad (9)$$

Where, y is the observed performance metric, $f(\theta)$ is the true objective function, and ε is a Gaussian noise term.

The abovementioned model is utilized to make predictions of the objective function for a given set of hyperparameters. The acquisition function is defined as the utility of the next hyperparameters to be evaluated. The acquisition function balances the exploration of the

hyperparameter space and the exploitation of promising regions of the space. The acquisition function used in this study is the expected improvement (EI), which is defined as:

$$EI(\theta) = E[\max(0, f(\theta) - f(\theta^*))] \quad (10)$$

Where θ^* is the current best set of hyperparameters.

The next set of hyperparameters to be evaluated is selected by maximizing the acquisition function:

$$\theta_{next} = \arg\max[EI(\theta)] \quad (11)$$

The algorithm iteratively selects the next set of hyperparameters to evaluate until convergence is achieved.

3.5. Pseudocode for the Research Work
Begin
Import necessary libraries and modules
Load the dataset from the UCI repository
Divide the dataset into training and testing sets
Perform data pre-processing (handle missing data, normalize the data, etc.)
Implement Bayesian optimization for hyperparameter tuning of KNN and random forest classifiers
Train the KNN and random forest classifiers on the training set
Use the trained classifiers to predict student performance on the testing set
Calculate the accuracy of the random forest and KNN classifiers
End

4. Results and Discussion
4.1. Dataset
The database was created utilising two sources: the questionnaire used to gather the aforementioned data and paper references of varying quality. Portuguese (395 samples) and Mathematics (395 samples) were the two datasets created from the data (649 datasets). Numerous attributes go unnoticed because there are not any distinctive values. Some data elements include school name, age, gender, travel time, distance from home to school, interests, health information, and others. They were gathered through surveys and school reports. The UCI repository [41] has further information about characteristics.

4.2. Results
This research paper provides the simulation results in terms of confusion matrix plots generated in MATLAB. The observed value of the objective function at the end of the optimization process was 0.30243, while the estimated value was 0.30237, indicating that the optimization algorithm performed well in finding the optimal value of the objective function. Figure 5 shows the progression of the minimum objective function value over the course of the optimization process, with the number of function evaluations on the x-axis and the minimum objective function value on the y-axis.

Table 1. Classification system's five levels [41]

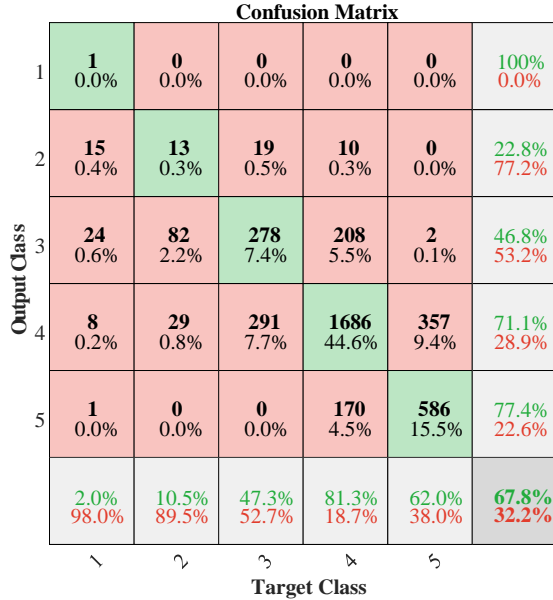|	I	II	III	IV	V
Country	(excellent/very good)	(good)	(satisfactory)	(sufficient)	(fail)
Portugal/France	16-20	14-15	12-13	10-11	0-9
Ireland	A	B	C	D	F

Fig. 2 Student prediction using RF

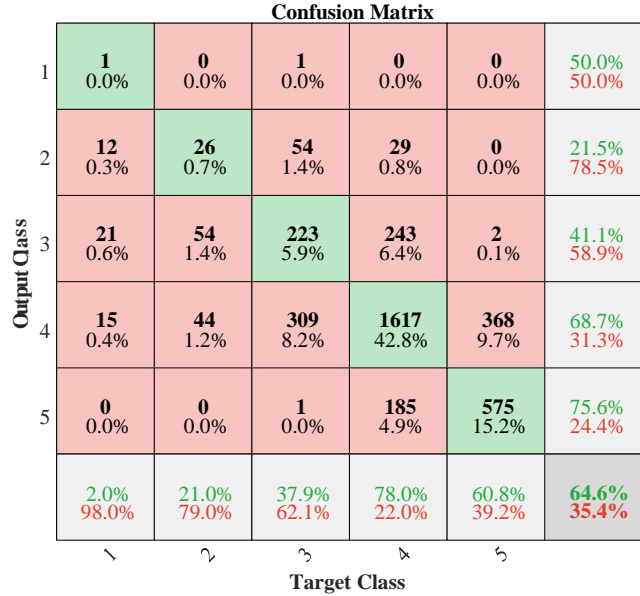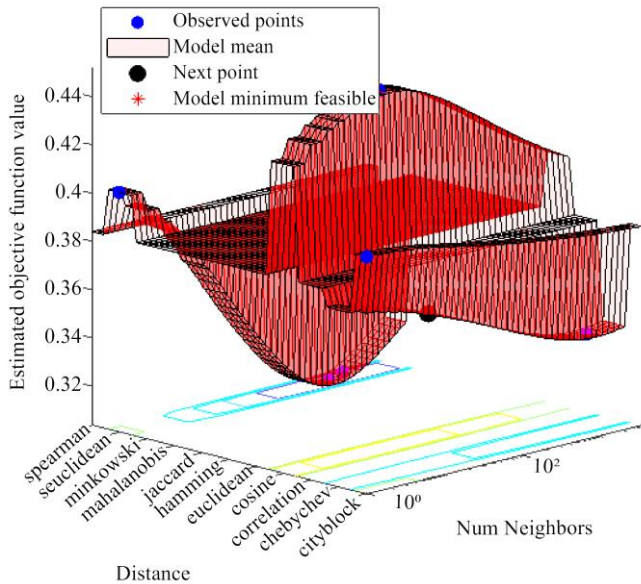**Fig. 3 Student performance using KNN**

Fig. 4 Bayesian optimization using a different distance measure
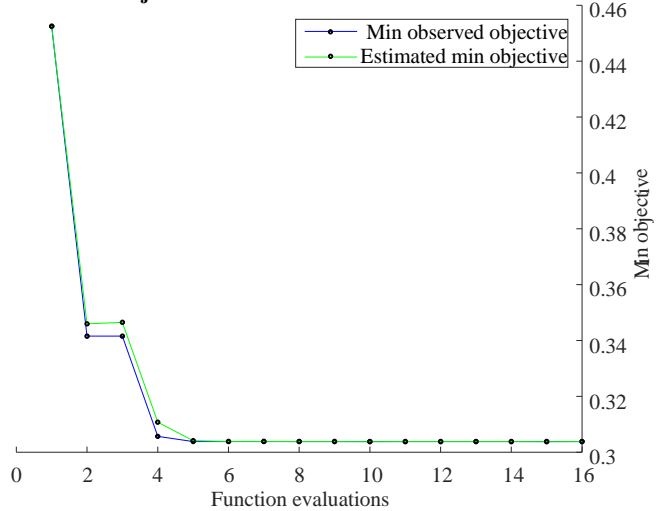
Fig. 5 Fitness achieved by the proposed algorithm

Table 2. Comparison of accuracy

Method	Accuracy of datasets used	
	Portuguese	**Maths**
KNN	64.61%	69%
RF	72%	78%
KNN-Bayesian optimization	67.8%	73%
RF-Bayesian optimization	79%	87%

Table 3. For UCI Maths data, performance analysis of existing and planned with various classifiers

Methods	Accuracy
Logistic Regression (SVM) [35]	62.05%
SVM [36]	72.3%
KNN [36]	62.3%
Proposed-KNN	69%
Proposed KNN- Bayesian optimization	73%
Proposed RF-Bayesian optimization	87%

As the number of function evaluations increases, it is possible that the minimum objective function value decreases as the optimization algorithm attempts to find the optimal solution. The time taken for each function evaluation was 0.36098 units, which may be indicated in the figure in some way, such as through the spacing of the data points on the x-axis.

To assess the efficiency of the approaches suggested in this investigation, experiments were performed using classification techniques utilized in prior studies. The comparison was carried out according to the approach described below. The results are presented in Table 3. Our KNN-based method outperformed the method presented in [35] by 7% in terms of accuracy. Additionally, the Bayesian-optimized KNN and Bayesian-optimized RF methods achieved promising results in the mathematics dataset, with accuracy rates of 73% and 87%, respectively, which compares favorably to the approaches outlined in [36].

5. Conclusion

Numerous studies have been conducted utilising data mining algorithms to estimate students' academic performance. It is feasible to improve prediction outcomes using the suitable data pre-processing method and selecting the right algorithm. In this study, a range of classifier techniques is suggested to forecast students' performance in math and Portuguese classes in higher secondary schools. After normalising the data to the [0, 1] interval, the proposed classifier algorithms for five-level classification are tuned during the training phase. The accuracy that was attained was 87% for the random forest classifier and 73% for the KNN classifier. The future potential of the data mining strategy is to evaluate student academic performance using various classification algorithms and learn these algorithms with attribute selection.

References

[1] R. S. Baker, "*Big Data and Education*," New York: Teachers College, Columbia University, 2015. [Google Scholar]

[2] Cristobal Romero, and Sebastian Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Hanan Aldowah, Hosam Al-Samarraie, and Wan Mohamad Fauzy, "Educational Data Mining and Learning Analytics for 21st Century Higher Education: A Review and Synthesis," *Telematics and Informatics*, vol. 37, pp. 13-49, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[4] K. Sutha, and J. Jebamalar Tamilselvi, "A Review of Feature Selection Algorithms for Data Mining Techniques," *International Journal on Computer Science and Engineering*, vol. 7, no. 6, pp. 63-67, 2015. [Google Scholar] [Publisher Link]

[5] Harsh H. Patel, and Purvi Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *International Journal of Computer Sciences and Engineering*, vol. 6, no. 10, pp. 74-78, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[6] Maria Tsiakmaki et al., "Transfer Learning from Deep Neural Networks for Predicting Student Performance," *Applied Sciences*, vol. 10, no. 6, pp. 2145, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Tiago Luís de Andrade, Sandro José Rigo, and Jorge Luis Victória Barbosa, "Active Methodology, Educational Data Mining and Learning Analytics: A Systematic Mapping Study," *Informatics in Education*, vol. 20, no. 2, pp. 171, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] E. Mangina, and G. Psyrra, "Review of Learning Analytics and Educational Data Mining Applications," *EDULEARN21 Proceedings, IATED Digital Library*, pp. 949-954, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Zenum Kastrati et al., "Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study," *Applied Sciences*, vol. 11, no. 9, pp. 3986, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] A. Namoun, and A. Alshanqiti, "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review," *Applied Sciences*, vol. 11, no. 1, pp. 237, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Christian Fischer et al., "Mining Big Data in Education: Affordances and Challenges," *Review of Research in Education*, vol. 44, no. 1, pp. 130-160, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Cristobal Romero, and Sebastian Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Francisca Onaolapo Oladipo, Ogunsanya Funmilayo Blessing, and Ezendu Ariwa, "Terrorism Detection Model using Naive Bayes Classifier," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 12, pp. 9-15, 2020. [CrossRef] [Publisher Link]

[14] Muhammad Haziq Bin Roslan, and Chwen Jen Chen, "Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021)," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Anupam Khan, and Soumya K. Ghosh, "Student Performance Analysis and Prediction in Classroom Learning: A Review of Educational Data Mining Studies," *Education and Information Technologies*, vol. 26, pp. 205-240, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Y. K. Salal, S. M. Abdullaev, and M. Kumar, "Educational Data Mining: Student Performance Prediction in Academic," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 4C, pp. 54-59, 2019. [Google Scholar] [Publisher Link]

[17] Wen Xiao, Ping Ji, and Juan Hu "A Survey on Educational Data Mining Methods Used for Predicting Students' Performance," *Engineering Reports*, vol. 4, no. 5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of students' Academic Performance Based on Courses' Grades Using Deep Neural Networks," *IEEE Access*, vol. 9, pp. 140731-140746, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[19] Vandana Mulye, and Atul Newase, "A Review: Recruitment Prediction Analysis of Undergraduate Engineering Students Using Data Mining Techniques," *SSRG International Journal of Computer Science and Engineering*, vol. 8, no. 3, pp. 1-6, 2021. [CrossRef] [Publisher Link]

[20] Ferda Ünal, "Data Mining for Student Performance Prediction in Education," *Data Mining-Methods, Applications and Systems*, vol. 28, pp. 423-432, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Suhas S. Athani et al., "Student Academic Performance and Social Behavior Predictor Using Data Mining Techniques," *International Conference on Computing, Communication and Automation (ICCCA),* pp.170-174, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Xiaofeng Ma, and Zhurong Zhou, "Student Pass Rates Prediction Using Optimized Support Vector Machine and Decision Tree," *IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC),* pp. 209-215, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[23] C. Troussas, M. Virvou, and S. Mesaretzidis, "Comparative Analysis of Algorithms for Student Characteristics Classification Using a Methodological Framework," *6th International Conference on Information, Intelligence, Systems and Applications (IISA),* pp. 1-5, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[24] Mandeep Singh et al., "Towards Enthusiasm Prediction of Portuguese School's Students towards Higher Education in Realtime," *International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, pp. 421-425, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Akanksha Pandey, and L S Maurya, "Career Prediction Classifiers based on Academic Performance and Skills using Machine Learning," *SSRG International Journal of Computer Science and Engineering*, vol. 9, no. 3, pp. 5-20, 2022. [CrossRef] [Publisher Link]

[26] Akhilesh Kumar Srivastava et al., "Prediction of Students Performance Using KNN and Decision Tree-A Machine Learning Approach," *Strad Reserch*, vol. 7, no. 9, pp. 119-125, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] Maryam Zaffar et al., "Role of FCBF Feature Selection in Educational Data Mining," *Mehran University Research Journal of Engineering & Technology*, vol. 39, no. 4, pp. 772-778, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[28] Xing Xu et al., "Prediction of Academic Performance Associated with Internet Usage Behaviors Using Machine Learning Algorithms," *Computers in Human Behavior,* vol. 98, pp. 166-173, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[29] Paniah Emeko Kodzo, "The Determinants of Educational Performance in Togo," *SSRG International Journal of Economics and Management Studies*, vol. 9, no. 11, pp. 6-16, 2022. [CrossRef] [Publisher Link]

[30] Charu C. Aggarwal, "*Mining Text Data,*" Springer International Publishing, pp. 429-455, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[31] Alfredo Daza et al., "Predicting Academic Performance through Data Mining: A Systematic Literature," *TEM Journal*, vol. 11, no. 2, pp. 939, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[32] Jurgen Schmidhuber, "Deep Learning in Neural Networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[33] Cindi Mason et al., "Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression," *Research in Higher Education*, vol. 59, pp. 382-400, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[34] C. Managwu, D.Matthias, and N. Nwaibu, "Random Forest Regression Model for Estimation of Neonatal Levels In Nigeria," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 7, pp. 41-44, 2020. [CrossRef] [Publisher Link]

[35] K. Deepika, N. Sathyanarayana, and N. Sathyanarayana, "Relief-F and Budget Tree Random Forest Based Feature Selection for Student Academic Performance Prediction," *International Journal of Intelligent Engineering and Systems*, vol. 12, no.1, pp. 30-39, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[36] Safira Begum, and Sunita S Padmannavar, "Prediction of Student Performance using Genetically Optimized Feature Selection with Multiclass Classification," *International Journal of Engineering Trends and Technology*, vol. 70, no. 4, pp. 223-235, 2022. [CrossRef] [Publisher Link]

[37] G. Lampropoulos, "Educational Data Mining and Learning Analytics in the 21st Century", *Encyclopedia of Data Science and Machine Learning, IGI Global*, pp. 1642-1651, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[38] M. Yağcı, "Educational Data Mining: Prediction of Students' Academic Performance Using Machine Learning Algorithms," *Smart Learning Environments*, vol. 9, no. 1, pp. 11 2022. [CrossRef] [Google Scholar] [Publisher Link]

[39] Nidhi Walia et al., "Student's Academic Performance Prediction in Academic Using Data Mining Techniques," *International Conference on Innovative Computing & Communications (ICICC),* 2020. [CrossRef] [Google Scholar] [Publisher Link]

[40] S.aid A. Salloum et al., "Mining in Educational Data: Review and Future Directions," *International Conference on Artificial Intelligence and Computer Vision (AICV2020)*, *Springer International Publishing*, pp. 92-102, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[41] Student Performance Data Set, UCI Machine Learning Repository, Online [available]: https://archive.ics.uci.edu/ml/datasets/student+performance