Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 14 No. 4 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijep.v14i4.48205

PAPER

Predicting Global Education Quality: A Comprehensive Machine Learning Approach Using World Bank Data

Rabab El Hatimi[1](✉), Cherifa Fatima Choukhan[1], Abdellatif Moussaid[2], Mustapha Esghir[1]

[1]Laboratory of Mathematics, Computing and Applications, Faculty of Sciences, Mohammed V University in Rabat, Rabat, Morocco

[2]ENSIAS, Mohammed V University in Rabat, Rabat, Morocco

rababelhatimi@gmail.com

ABSTRACT

This paper introduces an innovative approach to predicting the quality of education on a global scale. It leverages a comprehensive dataset spanning multiple years and countries from the World Bank. Our methodology involves two key steps: first, unsupervised clustering using the K-means model to categorize countries based on their educational quality levels; and second, employing supervised classification techniques to develop a predictive model. Through training and optimizing various machine learning (ML) algorithms, we aim to identify the most accurate model for predicting education quality. The outcomes highlight the efficacy of our approach, with the KNN algorithm demonstrating superior performance after hyperparameter optimization. It achieved precision, recall, accuracy, and AUC values of 0.9740, 0.9721, 0.9711, and 0.9959, respectively. These findings provide valuable insights for policymakers, educational institutions, and researchers, helping to identify areas that require attention and to design targeted interventions.

KEYWORDS

education quality, machine learning (ML), World Bank data, predictive modeling, global education assessment, KNN algorithm

1 INTRODUCTION

Education is a fundamental pillar of human life, playing a crucial role in sustainable development and the well-being of societies. It serves as a catalyst for poverty reduction and the achievement of sustainable development goals, representing a fundamental human right that extends throughout life. However, the state of education globally, particularly for children, remains a major concern, especially in developing countries. Despite progress in literacy rates, there are various challenges such as school dropout, economic constraints, and global crises that hinder access to education [1] [2] [3] [4].

The utilization of machine learning (ML) in evaluating the progression of education quality globally provides numerous advantages. Firstly, ML models can analyze

El Hatimi, R., Choukhan, C.F., Moussaid, A., Esghir, M. (2024). Predicting Global Education Quality: A Comprehensive Machine Learning Approach Using World Bank Data. *International Journal of Engineering Pedagogy (iJEP)*, 14(4), pp. 24–37. https://doi.org/10.3991/ijep.v14i4.48205

Article submitted 2023-11-26. Revision uploaded 2024-01-15. Final acceptance 2024-02-07.

large amounts of data from various sources to identify key factors that contribute to the quality of education. These factors may include student performance, available educational resources, implemented educational policies, student-teacher ratios, teacher training levels, and more [5] [6] [7] [8] [9].

By utilizing supervised learning techniques, ML models can learn from historical data to predict the quality of education in various regions or countries. For example, by utilizing historical datasets containing information on key indicators of education quality, such as standardized test scores, graduation rates, and educational expenditures, ML models can be trained to establish relationships and trends among these variables. This enables predictions regarding the improvement or deterioration of education quality in the future [10] [11] [12]. Furthermore, the use of ML can help identify gaps and priority areas in educational systems. ML models can perform segmentation analyses to group countries or regions based on their shared educational characteristics. This helps identify countries that could benefit from specific policies or investments to improve their education quality. For instance, ML models could reveal that certain countries have abundant educational resources but poor learning outcomes, indicating a need to improve teaching practices or student support.

Prior research conducted by [13] explored the correlation between state regulation indicators in the education sector and the achievement of sustainable development. Focusing on Central and Eastern European countries, the study utilized World Bank data from 2006 to 2016. By employing a regression model, the researchers illustrated the substantial influence of a state's initiatives in the education sector on the attainment of sustainable development goals, offering guidelines for efficient policy formulation.

Luan and Tsai (2021) [10] conducted an in-depth review on the use of ML approaches in personalized education. They explored various applications of ML in educational contexts and discussed its potential to enhance personalized learning experiences. In another literature review, Alenezi and Faisal (2020) [6] examined the implementation of crowdsourcing and ML in education. They highlighted the benefits of incorporating crowd contributions and ML algorithms to improve educational outcomes. Additionally, Ho, Cheong, and Weldon (2021) [14] focused on predicting student satisfaction in emergency remote learning in higher education using ML techniques. Their study demonstrated the effectiveness of ML models in understanding and predicting student perceptions and satisfaction in the context of remote learning.

Al-Emran, Al-Nuaimi, Arpaci, Al-Sharafi, and Anthony Jnr (2023) [15] conducted a study focused on understanding the determinants influencing students' adoption of wearable technologies in education. They used ML algorithms to analyze data and identify factors influencing students' acceptance and use of wearable technologies. This research contributes to the development of a wearable technology-based educational environment and provides insights into the integration of these technologies in educational contexts.

These works provide valuable insights into the use of ML in education and serve as important foundations for understanding the potential impact of ML on improving the quality of education.

World Bank data provides a wealth of information on many countries, including economic indicators and demographic statistics. Numerous studies have utilized this data to inform decision-making processes on various education-related topics, such as dropout rates, student performance, and the correlation between education and sustainable development [16].

In this study, we contribute to the existing body of knowledge by developing an innovative strategy that classifies countries based on their percentage of access to education using an unsupervised classification model. Our approach involves statistical analysis and data preprocessing using the K-means clustering algorithm with different numbers of clusters. We utilize variable selection techniques to identify the most significant factors and statistical tests to determine the most effective clustering outcomes. Our results enable us to identify two at-risk groups and the common challenges faced by countries in these groups, providing valuable information for decision-makers. Furthermore, our study paves the way for the development of a semi-autonomous and rapid diagnostic system capable of predicting access to education during future country data updates, aligning with sustainable development goals.

The key contributions of our study are as follows:

- Exploitation of a comprehensive World Bank database: We utilized an extensive database from the World Bank, encompassing multiple years and most countries. This approach allows us to obtain a global view of education quality and provide specific predictions for each country.
- Two-step approach: We implemented a two-step approach by combining unsupervised and supervised classification. This enables us to consider both economic and educational factors in our model for predicting education quality.
- Hyperparameter optimization and algorithm comparison: We optimized the hyperparameters of our models and compared several ML algorithms to select the one with the best educational quality prediction performance. This step enables us to obtain reliable and accurate results.

The remaining sections of this paper are structured as follows: Section 2 provides an overview of the dataset used, explains the data preprocessing steps, and outlines the classification algorithm employed. Next, Section 3 presents and discusses the findings of the experiment. Finally, Section 4 provides insights and perspectives on the potential implications of this analytical approach.

2 MATERIALS AND METHODS

2.1 Dataset

The data used in this study was sourced from the World Bank database [17] [18], encompassing a range of economic and educational variables as outlined in Table 1. These variables encapsulate aspects such as school accessibility (primary or secondary), student-to-class ratio, and schools per habitat ratio, among others. Additionally, economic indicators such as the unemployment rate, poverty rate, gross domestic product, per capita domestic product, foreign direct investment, net inflows, balance of payments, inflation, prices, consumption, and population were considered.

Given the wealth of pertinent data provided by the World Bank, including the mentioned indicators, we meticulously selected 60 variables relevant to our study. These variables cover 217 countries over a 20-year period, from 2001 to 2020, reflecting the latest statistics for each indicator. This comprehensive dataset ensures a consistent representation across various country categories, including developed, developing, impoverished, and marginalized nations.

After obtaining the data from the World Bank website, the initial dataset, which included 217 countries worldwide, exhibited numerous missing values. To address this issue, we initiated data preprocessing by excluding variables with over 50% missing values. Subsequently, the remaining missing values were imputed using the average values of their respective variables. Finally, we obtained a cleaned dataset of 217 countries with 21 variables, including 12 educational factors and nine economic factors (refer to Table 1).

Table 1. Data variables

Education Variables	Unschooled adolescents; Unschooled kids; Higher education inscriptions; Preschool inscriptions; Secondary school inscriptions; Ratio female/male in higher education; Ratio girls/boys in primary school; Ratio girls/boys in secondary school; Primary school achievement rate; Secondary school achievement rate; Youth literacy rate; Total literacy rate.
Economic Variables	Financing capacity; Unemployment; GPD growth; GPD per capita growth; Gini index; Labor force by the level of education; Employment rate 15+; Employment rate 15–24 years; Gross saving; Income share held by highest 20%; Income share held by lowest 20%; Gross domestic saving (% of GPD).

2.2 Data clustering

To preprocess the data for both input and output, we utilized the K-means algorithm along with the elbow method to determine meaningful labels. Illustrated in Figure 1, our exploration involved testing clusters ranging from one to nine using the elbow method [19], [20], [21], [22]. The optimal clustering emerged at three clusters, indicating a distinctive categorization for countries. These clusters represent three discernible types: countries with high-quality education, those with average-quality education, and those with low-quality education. This methodical approach ensures a nuanced classification that captures the diverse educational landscapes across nations.

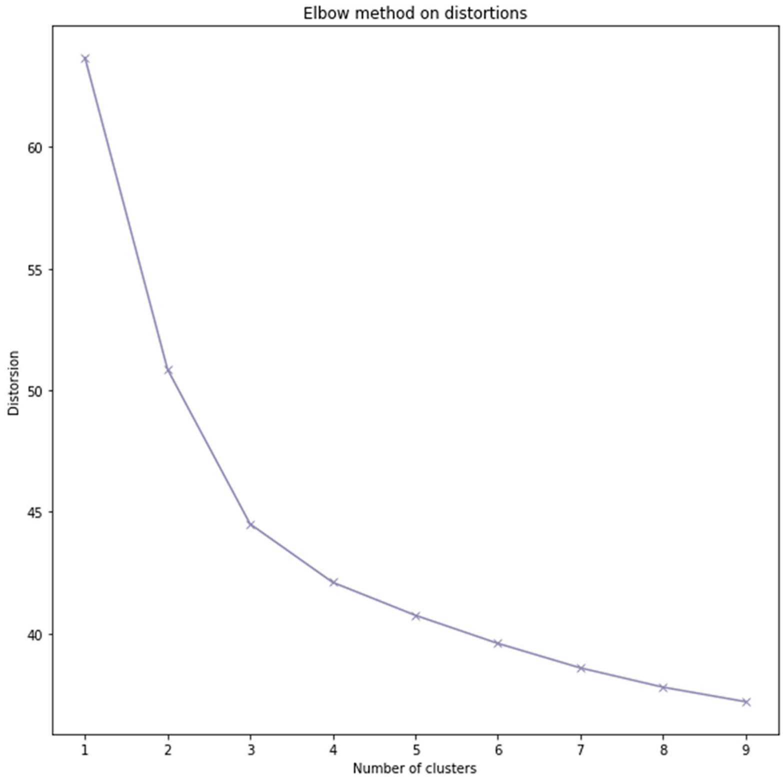

Fig. 1. Elbow results

After determining the optimal number of clusters, we applied the K-means algorithm to the complete dataset. The map of countries among the established clusters is visually illustrated in Figure 2 [23], [24], [25]. The color scheme distinguishes the clusters: green denotes countries with high-quality education, purple represents those with average-quality education, and red signifies countries with low-quality education. This crucial step signifies the successful preparation of our data, priming it for integration as input in predictive modeling initiatives. The structured clustering lays the foundation for accurate and targeted predictions, enabling a thorough analysis of educational dynamics across countries. This strategic approach enhances our ability to uncover nuanced insights into global educational patterns and disparities.

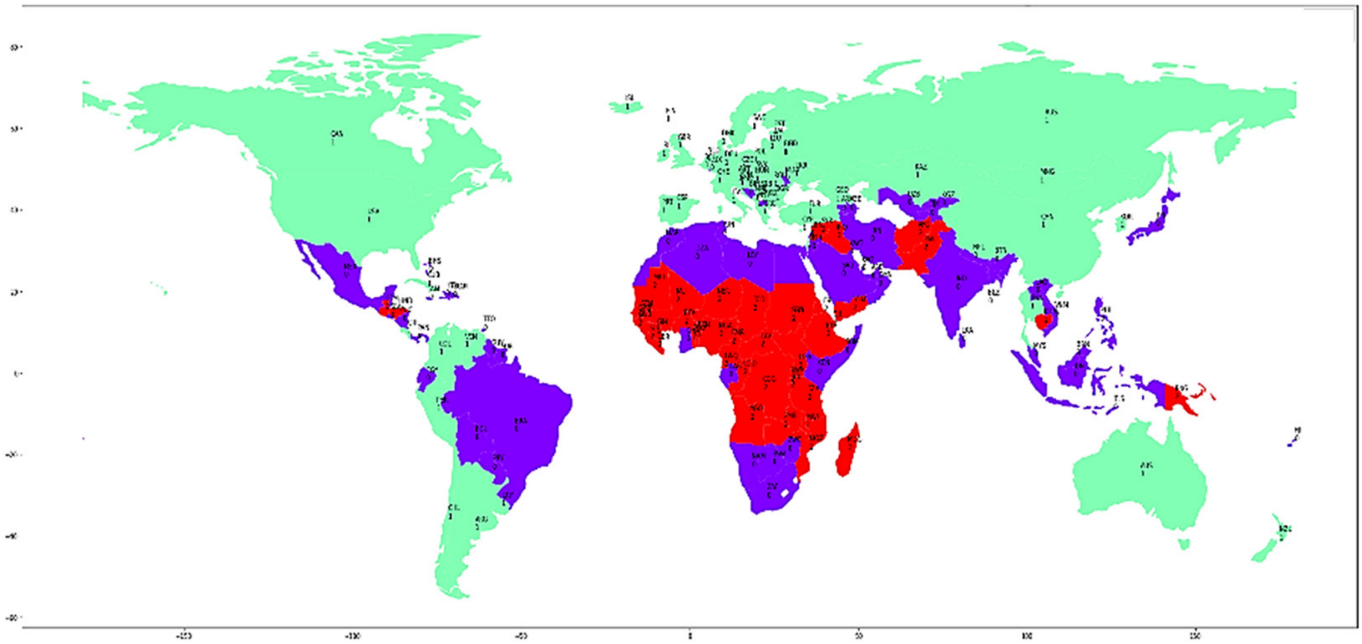

Fig. 2. K-means results (data labeling)

In order to summarize all the data comprehensively, we aimed to create a correlation matrix that would reveal the hidden information related to the features, their interrelationships, and their connection with the target [26]. Figure 3 displays a correlation matrix (heatmap) of the input variables. As evident from the heatmap, numerous variables are positively correlated (red color), while others are negatively correlated (blue color), indicating a significant dependency among them. The results demonstrate a high correlation among all variables, suggesting that the selected variables are reliable and follow a similar distribution. This is a positive indicator for our study. For instance, variables related to non-schooled individuals (non-schooled adolescents, non-schooled children, and enrollments in higher education) exhibit strong correlations with all variables. Additionally, we observe favorable correlations between several variables, such as primary and secondary school success rates with overall school success rates, enrollments in higher education, and success in secondary education, among others.

Furthermore, we note that specific variables reveal significant correlations, uncovering hidden information. For example, a country's funding capacity is positively correlated with gross savings. Additionally, variables associated with income distribution demonstrate a negative correlation with the GINI index, suggesting that countries with a more equitable income distribution tend to have a lower GINI index. We also observe a positive impact of higher education on development and on the employment of the population aged 15 and above.

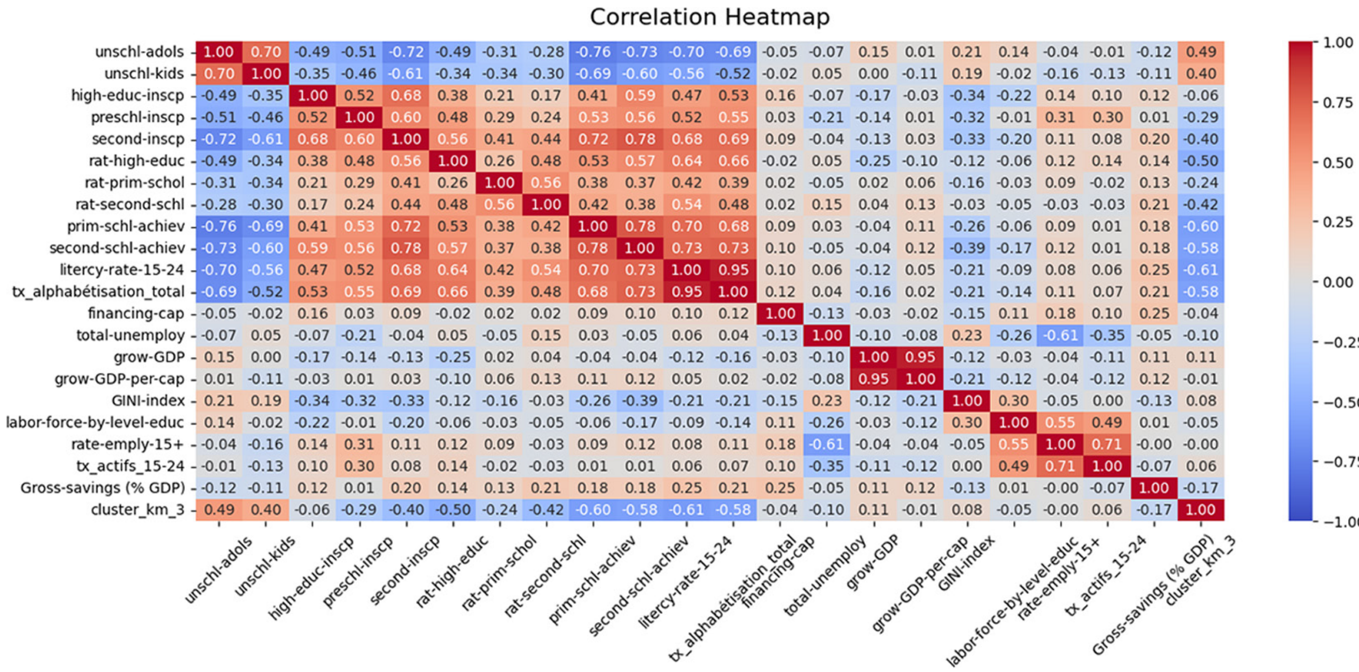

Fig. 3. Correlation matrix between input variables

Regarding the correlation between input variables and the target (see Figure 4), we observed that certain variables related to non-schooled individuals (non-schooled adolescents, non-schooled children) as well as other economic variables show a strong correlation, reaching up to 0.5 and 0.6. This constitutes a positive indicator before the onset of the prediction phase.

It is crucial to highlight that a strong correlation among variables is essential to enabling ML algorithms to train effectively and demonstrate the quality of the data used.

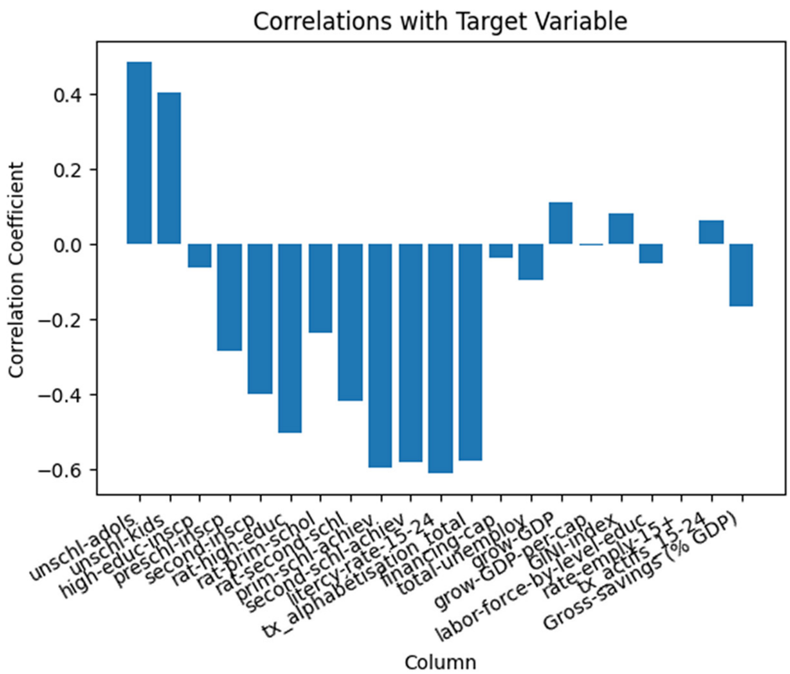

Fig. 4. Correlation between input variables and target

2.3 Modeling

In recent years, ML techniques have demonstrated significant efficacy across various domains [27] [28]. In our study, we utilized these techniques to predict global education quality after conducting a comprehensive data cleaning and preparation process. The task at hand involves a classification challenge, constituting a supervised learning problem where algorithms such as K-neighbors, random forests, and support vector machines (SVMs) play a crucial role in examining the relationships between the target variable (Y) and the set of variables (X), enhancing the precision of our predictions [29] [30].

Among the commonly used algorithms is the decision tree classifier, which identifies optimal features in the training input data by partitioning it into subsets based on each data point's features. These subsets are then used to construct a decision tree capable of predicting the class or label of a new dataset, concluding the modeling process when further predictions are no longer feasible [31].

Machine learning advancements have given rise to efficient algorithms such as gradient boosting classifiers, random forest classifiers, K neighbors' classifiers SVMs with a linear kernel, and others [32]. Employing a cross-validation technique with parameter fitting is essential for determining the best-performing model [33].

In our specific case, we partitioned the dataset into training and test sets. We implemented and evaluated various ML classification algorithms to determine the accuracy of the results. The methodology is outlined in Figure 5.

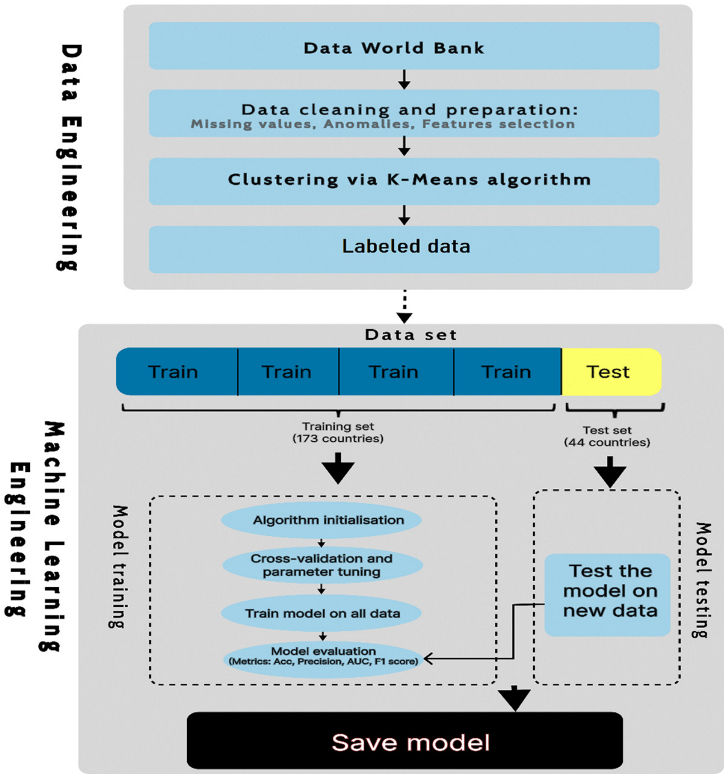

Fig. 5. The approach includes data engineering and machine learning steps

To prevent overfitting and select the most suitable algorithm, we utilized a five-fold stratified cross-validation technique, which was recursively applied and repeated for five classes. In each round, a grid-search technique was employed to identify optimal parameters for each algorithm [34] [35]. After five iterations, an average score was obtained, which served as reliable validation. Subsequently, we tested the models on a separate portion of the data (the test set) to obtain the prediction score. The cross-validation steps are summarized in Figure 6.

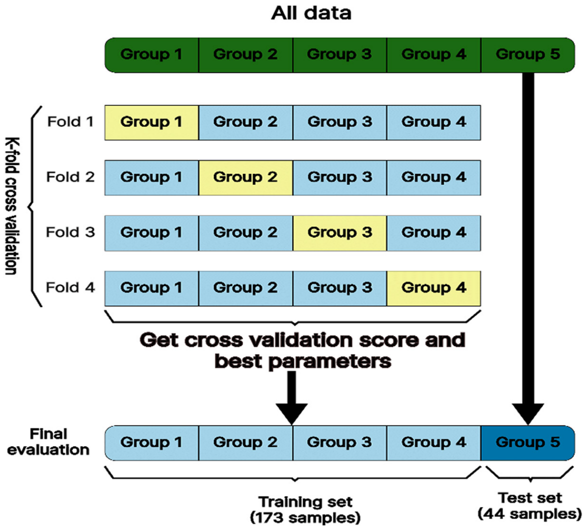

Fig. 6. Cross-validation and parameters tuning

3 RESULTS AND DISCUSSION

The purpose of this study is to improve the ranking of countries based on quality by making accurate predictions through classification algorithms using data from the World Bank. Given the sensitivity of this prediction, relying solely on accuracy may not be adequate without considering other performance measures. To evaluate the performance of classification models, we utilized a confusion matrix in our study. The confusion matrix measures four indicators: accuracy, recall, F1-score, and precision. The definitions of these indicators are provided in Table 2.

Table 2. Confusion matrix

	Positive Prediction	Negative Prediction
Positive Cases	TP	FN
Negative Cases	FP	TN

We utilized accuracy, recall, F1-score, and precision (Equations 1, 2, 3, 4) as evaluation metrics to gauge the effectiveness of the classification treatment results [36]. These metrics collectively provide a comprehensive evaluation of the model's performance in accurately predicting treatment protocols. Accuracy provides an overall measure of correctness; recall evaluates the model's ability to capture all relevant instances; and the F1-score balances precision and recall for a comprehensive assessment.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1_{score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

After performing cross-validation and parameter tuning using the Grid search algorithm, the results of the cross-validation, including accuracy, precision, recall, AUC, and F1 score for each model, are presented in Table 3.

Table 3. Metrics evaluation after tuning parameter

Model	Accuracy	AUC	Recall	Precision	F1-Score
K-Neighbors Classifier	**0.9711**	**0.9959**	**0.9721**	**0.9740**	**0.9710**
Extra Trees Classifier	0.9244	0.9890	0.9324	0.9279	0.9235
Ridge Classifier	0.9131	0.0000	0.8947	0.9251	0.9108
Random Forest Classifier	0.9074	0.9895	0.9078	0.9205	0.9067
Gradient Boosting Classifier	0.9012	0.9713	0.8954	0.9124	0.9011
SVM Linear Kernel	0.8845	0.0000	0.8781	0.9050	0.8821
Naive Bayes	0.8437	0.9599	0.8751	0.8589	0.8414
Decision Tree Classifier	0.8437	0.8740	0.8436	0.8519	0.8429
Quad Discriminant Analysis	0.7916	0.8991	0.7682	0.8159	0.7888
Ada Boost Classifier	0.7521	0.8494	0.7394	0.7425	0.7232
Dummy Classifier	0.4509	0.5000	0.3333	0.2035	0.2804

In Table 3, the tested algorithms demonstrate favorable results, with KNN particularly standing out as the top performer. Given the nature of our dataset, where we don't have a substantial volume of data, KNN proves effective in yielding good results. The obtained scores of 0.97 accuracy and 0.79 F1 score indicate the absence of overfitting or imbalance issues. This underscores the model's proficiency in classifying countries into categories of good, average, or low education.

Also, the AUC metric serves as additional validation, demonstrating that our model does not make predictions randomly, as evidenced in Figure 7. This underscores the robust predictive capability of our model.

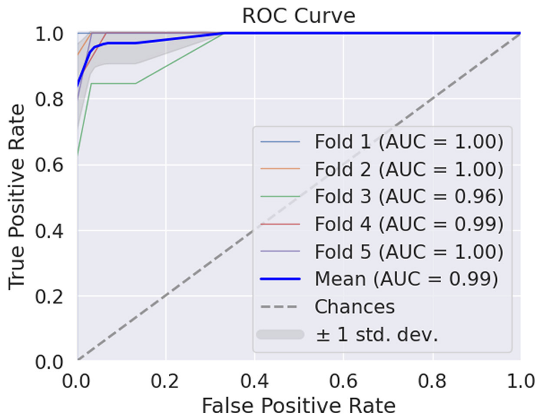

Fig. 7. AUC ROC curve

The learning and validation curves during the training process are depicted in Figure 8. As evident from the graphs, our model shows no signs of overfitting, emphasizing its robustness and effectiveness.

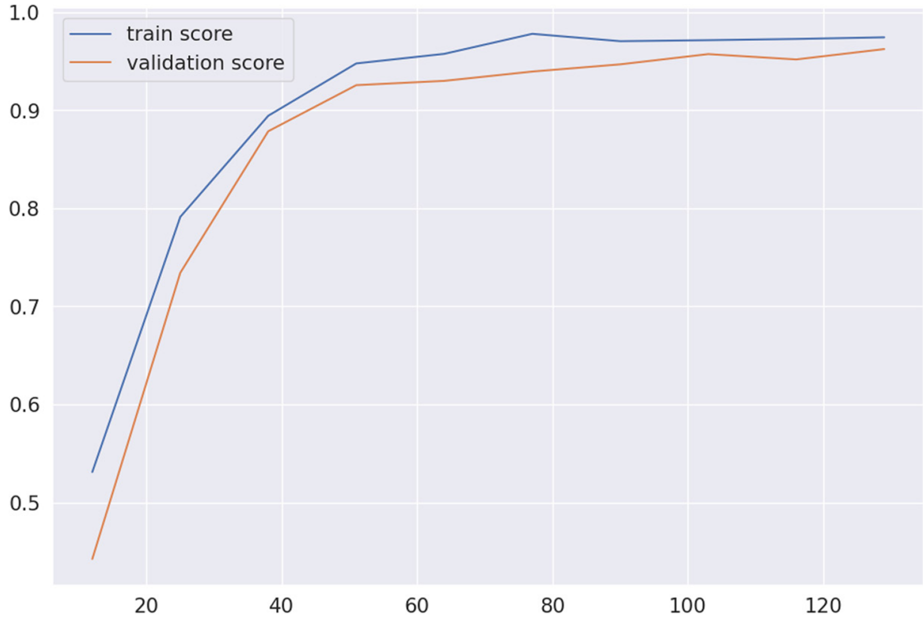

Fig. 8. Learning and validation curves of the KNN model

To delve deeper into the obtained results, we constructed a graph depicting the three metrics—accuracy, recall, and F1-score—to meticulously analyze the precision of predictions against the actual values.

In Figure 9, the confusion matrix is meticulously presented. Upon testing our model on 44 countries, it successfully classified 43 of them accurately, with only one misclassification. This outcome underscores the model's robust predictive capabilities and its ability to make accurate predictions.

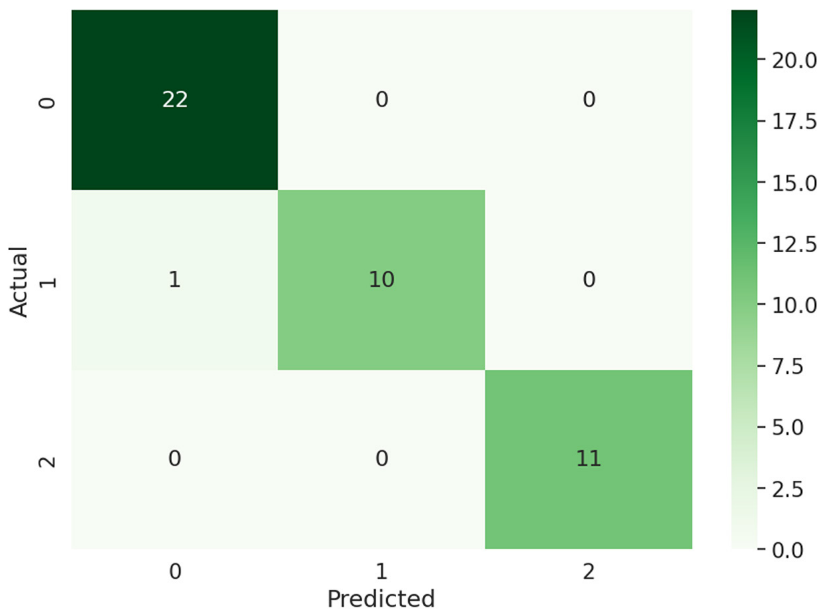

Fig. 9. Confusion matrix

In conclusion, our developed strategy offers an efficient and expedient method for predicting the quality of education globally across countries. Through the rigorous evaluation of ML algorithms on our dataset, the algorithm, when applied with hyperparameter optimization, emerges as the most accurate predictor. This refined KNN technique yields precision, recall, accuracy, and AUC values of 0.9740, 0.9721, 0.9711, and 0.9959, respectively.

Interestingly, other classifiers such as K neighbors' classifier, extra trees classifier, and ridge classifier exhibit nearly identical metrics, regardless of hyperparameter optimization or selection techniques employed. This consistency in metrics may be attributed to the challenge of handling limited data registration, which affects and elucidates the classification results of these models. Interpreting standardized data with low registration rates underscores the intricacies associated with it.

4 CONCLUSION

Our study emphasizes the significance of prioritizing the enhancement of the quality of education on a global scale. By utilizing ML techniques and leveraging data from the World Bank, we have successfully developed an accurate predictive model to evaluate the quality of education in various countries. These findings provide valuable information to policymakers, such as the Council of Europe, UNICEF, and UNESCO, enabling them to make informed decisions to enhance the quality of education. These organizations play a crucial role in promoting effective educational policies globally and coordinating efforts to ensure quality education for all children.

It is essential to continue efforts to guarantee quality education for all children. This necessitates a continuous focus on teacher training, the creation of pertinent curricula, the provision of high-quality educational resources, and frequent evaluation of student performance. Furthermore, it is important to recognize that the quality of education is influenced by complex factors such as educational policies, financial resources, school infrastructure, and pedagogical practices. Therefore, efforts to improve the quality of education must be approached holistically, taking these different aspects into account.

In the future, it is crucial to continue researching ways to enhance the quality of education. In-depth studies on the impact of social and cultural factors, the integration of technology in education, long-term evaluation of the benefits of investing in quality education, sharing best practices, and strengthening teacher capacities are necessary. These future endeavors will provide a better understanding of specific challenges, identify effective solutions, and guide educational policies to ensure quality education for all learners. This will contribute to sustainable economic and social development.

5 REFERENCES

[1] C. B. Lloyd and A. K. Blanc, "Children's schooling in sub-Saharan Africa: The role of fathers, mothers, and others," *Population and Development Review*, vol. 22, no. 2, pp. 265–298, 1996. https://doi.org/10.2307/2137435

[2] A. Siebrits and V. van de Heyde, "Towards the sustainable development goals in Africa: The African space-education ecosystem for sustainability and the role of educational technologies," in *Embedding Space in African Society: The United Nations Sustainable Development Goals 2030 Supported by Space Applications*, *Springer,* 2019, pp. 127–180. https://doi.org/10.1007/978-3-030-06040-4_10

[3] D. Filmer and L. Pritchett, "The effect of household wealth on educational attainment: Evidence from 35 countries," *Population Development Review*, vol. 25, no. 1, pp. 85–120, 1999. https://doi.org/10.1111/j.1728-4457.1999.00085.x
[4] I. Ozturk, "The role of education in economic development: A theoretical perspective," *Journal of Rural Development and Administration*, vol. XXXIII, pp. 39–47, 2001. https://doi.org/10.2139/ssrn.1137541
[5] H. Luan and C. C. Tsai, "A review of using machine learning approaches for precision education," *Educ. Technol. Soc.*, vol. 24, no. 1, pp. 250–266, 2021.
[6] H. S. Alenezi and M. H. Faisal, "Utilizing crowdsourcing and machine learning in education: Literature review," *Education Information Technologies*, vol. 25, no. 4, pp. 2971–2986, 2020. https://doi.org/10.1007/s10639-020-10102-w
[7] B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques," *Educ. Sci.*, vol. 11, no. 9, p. 552, 2021. https://doi.org/10.3390/educsci11090552
[8] A. A. Enughwure and M. E. Ogbise, "Application of machine learning methods to predict student performance: A systematic literature review," *Int. Res. J. Eng. Technol.*, vol. 7, no. 5, pp. 3405–3415, 2020.
[9] R. M. Martins and C. G. Von Wangenheim, "Findings on teaching machine learning in high school: A ten-year systematic literature review," *Informatics in Education*, vol. 22, no. 3, pp. 421–440, 2023. https://doi.org/10.15388/infedu.2023.18
[10] N. R. Devi and A. Ramakrishna, "Artificial Intelligence in Education," *J. Huazhong Univ. Sci. Technol.,* vol. 1671, p. 4512.
[11] P. Blikstein *et al.*, "The future of maker education," in *Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023*, International Society of the Learning Sciences, 2023, pp. 1650–1658. https://doi.org/10.22318/icls2023.337879
[12] E. S. Bhutto, I. F. Siddiqui, Q. A. Arain, and M. Anwar, "Predicting students' academic performance through supervised machine learning," in *2020 International Conference on Information Science and Communication Technology (ICISCT),* IEEE, 2020, pp. 1–6. https://doi.org/10.1109/ICISCT49550.2020.9080033
[13] A. S. Vorontsova, T. A. Vasylieva, Y. V. Bilan, G. Ostasz, and T. Mayboroda, "The influence of state regulation of education for achieving the sustainable development goals: Case study of Central and Eastern European countries," 2020.
[14] I. M. K. Ho, K. Y. Cheong, and A. Weldon, "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques," *Plos One*, vol. 16, no. 4, p. e0249423, 2021. https://doi.org/10.1371/journal.pone.0249423
[15] M. Al-Emran, M. N. Al-Nuaimi, I. Arpaci, M. A. Al-Sharafi, and B. Anthony Jnr, "Towards a wearable education: Understanding the determinants affecting students' adoption of wearable technologies using machine learning algorithms," *Education and Information Technologies*, vol. 28, no. 3, pp. 2727–2746, 2023. https://doi.org/10.1007/s10639-022-11294-z
[16] H. A. Patrinos, G. Psacharopoulos, and A. Tansel, "Private and social returns to investment in education: The case of Turkey with alternative methods," *Applied Economics*, vol. 53, no. 14, pp. 1638–1658, 2021. https://doi.org/10.1080/00036846.2020.1841086
[17] K. Mundy and A. Verger, "The World Bank and the global governance of education in a changing world order," *The Handbook of Global Education Policy*, pp. 335–356, 2016. https://doi.org/10.1002/9781118468005.ch18
[18] T. L. Morgan, D. Zakhem, and W. L. Cooper, "From high school access to postsecondary success: An exploratory study of the impact of high-rigor coursework," *Education Science*, vol. 8, no. 4, p. 191, 2018. https://doi.org/10.3390/educsci8040191

[19] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and k-means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, 2014.

[20] M. Cui, "Introduction to the k-means clustering algorithm based on the elbow method," *Account. Audit. Finance*, vol. 1, no. 1, pp. 5–8, 2020.

[21] E. Umargono, J. E. Suseno, and S. Gunawan, "K-Means clustering optimization using the elbow method and early centroid determination based-on mean and median," in *Proceedings of the International Conferences on Information System and Technology – CONRIST*, SCITEPRESS—Science and Technology Publications, 2019, vol. 1, pp. 234–240. https://doi.org/10.5220/0009908402340240

[22] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," in *IOP Conference Series: Materials Science and Engineering*, 2018, p. 012017. https://doi.org/10.1088/1757-899X/336/1/012017

[23] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020. https://doi.org/10.3390/electronics9081295

[24] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," in *IEEE Access*, vol. 8, pp. 80716–80727, 2020. https://doi.org/10.1109/ACCESS.2020.2988796

[25] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019. https://doi.org/10.3390/j2020016

[26] J. M. Rickman, G. Balasubramanian, C. J. Marvel, H. M. Chan, and M.-T. Burton, "Machine learning strategies for high-entropy alloys," *J. Appl. Phys.*, vol. 128, no. 22, 2020. https://doi.org/10.1063/5.0030367

[27] F. Emmert-Streib, O. Yli-Harja, and M. Dehmer, "Explainable artificial intelligence and machine learning: A reality rooted perspective," *Wiley Interdisciplinary Reviews Data Mining Knowledge Discovery*, vol. 10, no. 6, p. e1368, 2020. https://doi.org/10.1002/widm.1368

[28] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020. https://doi.org/10.3390/e23010018

[29] S. R. Jiao, J. Song, and B. Liu, "A review of decision tree classification algorithms for continuous variables," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 012083. https://doi.org/10.1088/1742-6596/1651/1/012083

[30] P. Wang, S. Mou, J. Lian, and W. Ren, "Solving a system of linear equations: From centralized to distributed algorithms," *Annual Reviews in Control*, vol. 47, pp. 306–322, 2019. https://doi.org/10.1016/j.arcontrol.2019.04.008

[31] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science And Technology Trends (JASTT)*, vol. 2, no. 1, pp. 20–28, 2021. https://doi.org/10.38094/jastt20165

[32] A. Kadiyala and A. Kumar, "Applications of Python to evaluate the performance of decision tree-based boosting algorithms," *Environmental Progress and Sustainable Energy*, vol. 37, no. 2, pp. 618–623, 2018. https://doi.org/10.1002/ep.12888

[33] N. E. Siedhoff, U. Schwaneberg, and M. D. Davari, "Chapter Twelve – Machine learning-assisted enzyme engineering," *Methods in Enzymology*, vol. 643, pp. 281–315, 2020. https://doi.org/10.1016/bs.mie.2020.05.005

[34] D. M. Belete and M. D. Huchaiah, "Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results," *International Journey of Computer and Applications*, vol. 44, no. 9, pp. 875–886, 2022. https://doi.org/10.1080/1206212X.2021.1974663

[35] S. Prabu, B. Thiyaneswaran, M. Sujatha, C. Nalini, and S. Rajkumar, "Grid search for predicting coronary heart disease by tuning hyper-parameters," *Computer Systems Science and Engineering (CSSE)*, vol. 43, no. 2, pp. 737–749, 2022. https://doi.org/10.32604/csse.2022.022739

[36] K. Jyoti, S. Sushma, S. Yadav, P. Kumar, R. B. Pachori, and S. Mukherjee, "Automatic diagnosis of COVID-19 with MCA-inspired TQWT-based classification of chest X-ray images," *Computers in Biology and Medicine*, vol. 152, p. 106331, 2023. https://doi.org/10.1016/j.compbiomed.2022.106331

6 AUTHORS

Rabab El Hatimi, Laboratory of Mathematics, Computing and Applications, Faculty of Sciences, Mohammed V University in Rabat, Rabat 10000, Morocco (E-mail: rababelhatimi@gmail.com).

Cherifa Fatima Choukhan, Laboratory of Mathematics, Computing and Applications, Faculty of Sciences, Mohammed V University in Rabat, Rabat 10000, Morocco.

Abdellatif Moussaid, ENSIAS, Mohammed V University in Rabat, Rabat 10000, Morocco.

Mustapha Esghir, Laboratory of Mathematics, Computing and Applications, Faculty of Sciences, Mohammed V University in Rabat, Rabat 10000, Morocco.