Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 15 No. 1 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijep.v15i1.51985

PAPER

Personalized Guidance for Moroccan Students: An Approach Based on Machine Learning and Big Data

Morad Badrani[1](✉), Adil Marouan[1], Nabil Kannouf[2], Abdelaziz Chetouani[1]

[1]Mohammed First University, Oujda, Morocco

[2]Abdelmalek Essaadi University, Tetouan, Morocco

m.badrani@ump.ac.ma

ABSTRACT

Helping Moroccan students choose their high school presents significant challenges influenced by a variety of factors, including academic achievement, potential, and environmental influences. This study addresses these complexities using advanced data analytics and intelligent algorithms. We collected and examined authentic data from various secondary schools across Morocco, using the MASSAR system, a centralized education platform. To ensure robust model evaluation and optimized performance, we implemented 5-fold cross-validation and extensive hyper-parameter tuning for both support vector machine (SVM) and neural network models. Advanced classification algorithms, including hybrid learning techniques with neural networks and SVM algorithms, were applied, resulting in outstanding precision measures: 99.17% accuracy, 99.20% precision, 99.37% recall, 99.28% F1 score, and 0.99 area under the curve (AUC). By integrating this hybrid learning approach, powered by big data technologies such as Hadoop and Hadoop Distributed File System (HDFS), we accurately predict student choices and offer valuable academic advice. The use of a Hadoop cluster accelerated execution time by 40%. This pioneering merger underlines the adaptability and effectiveness of our approach to meeting the real-world educational challenges specific to the Moroccan context.

KEYWORDS

machine learning (ML), neural networks, student's orientation, support vector machines (SVM)

1 INTRODUCTION

In Morocco, first-year science students face a crucial decision when choosing between mathematical sciences (SM), physics (SF), and life and earth sciences (SVT). The current guidance process, which relies heavily on manual methods, often fails to provide students with the guidance they need to make informed decisions, resulting in a mismatch between their aptitudes and the chosen streams. This mismatch often leads to disengagement and contributes significantly to rising drop-out rates.

To address this challenge, our study proposes a data-driven approach to personalizing student guidance. Using machine learning (ML) and big data techniques,

Badrani, M., Marouan, A., Kannouf, N., Chetouani, A. (2025). Personalized Guidance for Moroccan Students: An Approach Based on Machine Learning and Big Data. *International Journal of Engineering Pedagogy (iJEP)*, 15(1), pp. 125–136. https://doi.org/10.3991/ijep.v15i1.51985

Article submitted 2024-09-14. Revision uploaded 2024-11-07. Final acceptance 2024-11-07.

including Hadoop and the Hadoop Distributed File System (HDFS), we will analyze extensive student data from the MASSAR system (a centralized education platform) to discover patterns and correlations predictive of academic success.

The importance of educational guidance goes beyond making informed choices and securing a future career. It encompasses a complete system of elements and dimensions that have a significant impact on the overall success of the educational system. A well-planned guidance system, based on early follow-up and a proactive approach, is essential to facilitate students' transitions between education systems, training, and the world of work.

Unfortunately, Moroccan high school students often follow a traditional approach based solely on choice, without taking into account their aptitudes and grades. This issue has been widely debated in the Moroccan parliament, with some parents forcing their children into scientific or physical fields solely because of perceived job prospects in engineering and medicine, disregarding their children's abilities and preferences. As a result, many students fail in their studies when they could have succeeded if their wishes and abilities had been respected. In addition, teachers are often frustrated when they have to deal with students who are forced to follow an orientation that does not correspond to their abilities, even if the student asks to change direction or despite the advice of family members.

This paper aims to address these challenges by exploring the possibility of providing predictions to secondary school students based on their results. By leveraging ML and big data, we aim to empower students to make informed choices and improve their chances of academic success.

Few studies have been carried out on the prediction of students' orientation choices, particularly using ML techniques. Several studies have explored different classification algorithms for this task. Badrani et al. [1] compared these methods for college student orientation and found decision trees to be the most effective. Similar results were obtained by Badrani et al. [2], who found random forests to be optimal, and by Hicham et al. [3], who also found that decision trees outperformed the others. Other approaches by Elmrabat Hicham et al. [4] took advantage of Internet of Things (IoT) technology to develop a guidance system. Kemper et al. [5] found that decision trees were superior to other algorithms. Ouatik et al. [6] achieved high accuracy with SVMs to predict student success. In another work, Mimis Mohamed et al. [7] developed a system for Morocco's CPGE program. Other approaches were explored by Ouatik et al. [8], where Naive Bayes combined with MapReduce and Hadoop proved most effective. Zahour et al. [9] designed a chatbot system for student guidance. Boubaker et al. [10] studied the application of neural networks to predict student performance, with promising results. Nabizadeh et al. [11] explored ML techniques for predicting the performance of school-leavers in playful learning environments. Sánchez-Pozo et al. [12] compared supervised ML models to predict the academic success of high school students. Alghamdi et al. [13] used data mining to predict student success. Recent advances in ML have shown promising applications in various fields. Atia et al. [14] explored particle swarm optimization and two-way fixed-effects analysis of variance to achieve efficient segmentation of brain tumors. Kanetaki et al. [15] highlighted the role of multimedia in engineering education through their experimental study using YouTube, focusing on the analysis and interpretation of knowledge data. In another work, Kanetaki et al. [16] examined the modeling of grade prediction in hybrid learning contexts, revealing effective strategies for improving academic results. Arbaoui et al. [17] applied wavelet-based multi-resolution analysis integrated with deep learning to monitor cracks in concrete, demonstrating the flexibility of ML techniques. Benzaoui et al. [18] provided a comprehensive review of ear recognition, detailing the different methodologies and current challenges. Finally, Jacques et al. [19] examined the challenges and emerging trends in education following the COVID-19 conference, highlighting the need for innovative solutions to maintain educational standards in a changing landscape.

This paper contributes to the field by proposing a novel hybrid ML model to help assist Moroccan students in making informed orientation choices. Our study aims to address the critical issue of student dropout by providing personalized guidance and support. The following sections outline methodology, results, discussion, and conclusion. By revolutionizing the orientation process, we aim to empower Moroccan students to make informed choices, enhance their chances of academic success, and ultimately contribute to reducing dropout rates.

The remainder of this paper is organized as follows. Section 2 describes the methodology used. The main results obtained are presented and discussed in Section 3. Section 4 presents the main conclusions of this research work.

2 METHODOLOGY

We used a hybrid learning model combining support vector machines (SVMs) and neural networks to accurately predict students' orientation choices. This approach aimed to take advantage of the strong classification capabilities of SVMs and the ability of neural networks to capture complex patterns in student data. To manage the considerable computational demands, we used a Hadoop cluster configured with HDFS, comprising 3 Datanodes and 1 Namenode.

2.1 Data acquisition by the students

Student data were obtained from the MASSAR system, provided by the Education Department in Nador, Morocco. This system contains comprehensive information on students, including:

- **Core academic performance:** Grades in Mathematics, English, Physics, History and Geography, Life and Earth Sciences, French, Physical Education, Islamic Education, and Arabic.
- **Student engagement:** Number of absences and activities grades for Physics, Maths, and Life and Earth Sciences.
- **Expert opinion:** Teacher recommendations for Life and Earth, Mathematics, and Physics.
- **Outcome variable:** Selected academic orientation (SM: Mathematical Sciences, SF: Physical Sciences, SVT: Life and Earth Sciences).

The data set underwent rigorous pre-processing to deal with missing values and inconsistencies. Continuous variables, such as grades and number of absences, were standardized. Categorical variables, such as teacher recommendations, were converted into numerical representations using appropriate encoding techniques.

To prepare the data for modeling, the dataset was randomly divided into a training set (70%) and a test set (30%). To further validate the performance of our model and resolve potential over-fitting issues, we implemented a 5-fold cross-validation technique. This approach divides our data set into five equal subsets. The model is trained on four subsets and tested on the last, this process being repeated five times so that each subset serves once as a test set. This method enables a more robust assessment of our model's generalization capabilities.

2.2 Application of machine learning methods

Machine learning techniques [20] are one of the subdivisions of artificial intelligence (AI) based on coding computers in all their forms so that they have the ability to respond to commands and complete assigned tasks on the basis of existing data and its analysis, with the limitation of human intervention in direction or lack thereof. It should be noted that the machine in this situation must depend on the analysis of input data in advance to respond to commands, apply the decision once it is needed, and decide which tasks to perform; the role of human beings will ultimately be much reduced. ML algorithms can be divided according to the learning modes used (see Figure 1): supervised, unsupervised, and reinforcement learning paradigms.

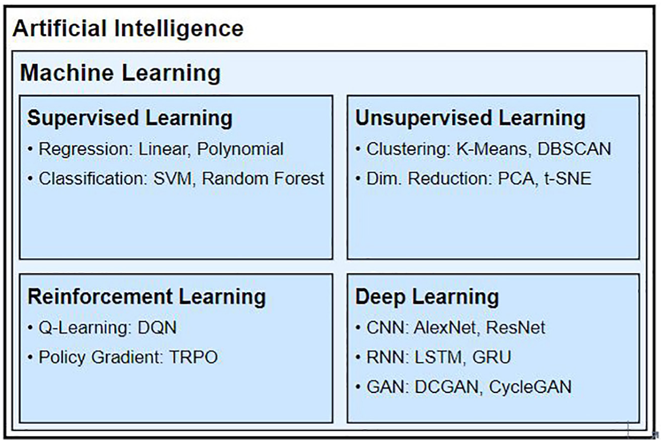

Fig. 1. Artificial intelligence and machine learning techniques

Additionally, deep learning, an advanced subset of ML, has emerged as a powerful approach. It utilizes multi-layered neural networks to extract complex features from data, enabling more sophisticated pattern recognition and decision-making capabilities.

Support vector machines (SVMs). Support vector machines, as referenced in [21] and [22], represent a model used to classify data into different categories. During the training phase, SVM identifies a boundary line that separates a given dataset into distinct classes while maximizing the margin between individual classes, which refers to the distance between the boundary and the closest data point of each class. Once the classification boundaries are established, the algorithm can then classify new data based on these learned patterns. In our implementation, we optimized the following key hyperparameters for the SVM model:

- Kernel type: We explored linear, radial basis function (rbf), and polynomial (poly) kernels to determine the optimal hyperplane for class separation.
- C (regularization parameter): This controls the trade-off between achieving a low training error and a low testing error.
- Gamma: For non-linear kernels, this parameter defines the influence reach of each training example.

These parameters were fine-tuned using grid search with 5-fold cross-validation to find the optimal configuration for our dataset.

Neural networks. In our approach, we leverage the MLPClassifier, a neural network architecture specifically designed for classification tasks. This model excels at identifying complex patterns within data by using deep learning techniques [23], [24]. Neural networks consist of interconnected processing units, called neurons, organized into layers. Through a training process called backpropagation, the model iteratively adjusts the connections between these neurons, enabling it to learn and extract intricate relationships within the dataset. For our MLPClassifier, we optimized several key hyperparameters:

- Hidden layer sizes: We experimented with various architectures to balance model complexity and performance.
- Activation function: 'relu' (Rectified Linear Unit) and 'tanh' (hyperbolic tangent) functions were considered.
- Optimizer: We compared 'adam' (Adaptive Moment Estimation) and 'sgd' (Stochastic Gradient Descent) algorithms.
- Learning rate: Various values were tested to find the optimal learning speed.
- Maximum iterations: This was set to balance between sufficient training and computational efficiency.

These hyperparameters were also tuned using grid search with 5-fold cross-validation.

Hybrid models. Within our hybrid learning framework, we integrate SVM and neural networks as base classifiers. Each base classifier, namely SVM and MLPClassifier, undergoes training on distinct subsets of the provided dataset D, labeled as D1 and D2, respectively. The SVM classifier is specifically trained using D1, while the MLPClassifier utilizes D2 for its training process. After training both base classifiers, we employ a voting mechanism (see Figure 2) to combine their predictions. When presented with a new data point for classification, each base classifier independently predicts the student's orientation choice. The hybrid classifier then aggregates these predictions using a majority vote, resulting in the final orientation prediction for the new student. This approach, while currently using standard neural networks, is designed to potentially incorporate more advanced deep learning techniques in future iterations.

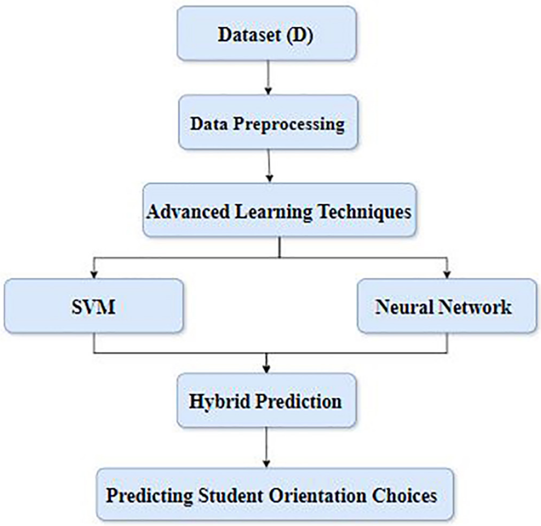

Fig. 2. Workflow diagram for using ensemble learning to predict student orientation

Weka. In our study, Weka [25] serves a crucial role in facilitating data preprocessing, feature selection, and the implementation of the hybrid algorithm combining neural networks and SVM. As an essential tool for data analysis, Weka provides a robust framework for handling various machine-learning tasks, including supervised classification. Despite potential challenges with large datasets, Weka's integration with complementary frameworks such as MOA enhances its capability for effective analysis of large datasets, closely aligned with the objectives of our study.

2.3 Tools for managing big data

To meet the challenge of managing vast student data sets, we turned to Big Data, which requires specialized technologies due to its sheer volume, speed, and variety. Although various platforms exist, we opted for Hadoop, the predominant open-source solution. Hadoop, with its components such as MapReduce and HDFS, facilitates the efficient storage and processing of large datasets, making it the ideal solution for our study.

Hadoop. Hadoop [26] is an open-source software framework for storing and deploying applications on conventional machine clusters. Based on Java and supported by the Apache Software Foundation, Hadoop offers extensive storage capabilities, robust processing power and scalability to handle diverse workloads. Based on the MapReduce framework, Hadoop efficiently manages large datasets by enabling processing software to operate directly on the data, eliminating the need to transfer data across networks for processing.

MapReduce. MapReduce (see Figure 3) is a programming model initially devised by Google to handle and produce extensive datasets within computer clusters. It serves as a fundamental element of Apache Hadoop, facilitating the distributed processing of extensive unstructured datasets across nodes within a cluster. The framework delegates tasks across nodes (mapping) and subsequently aggregates results into cohesive outputs (reducing), facilitated by the HDFS distributed file system.

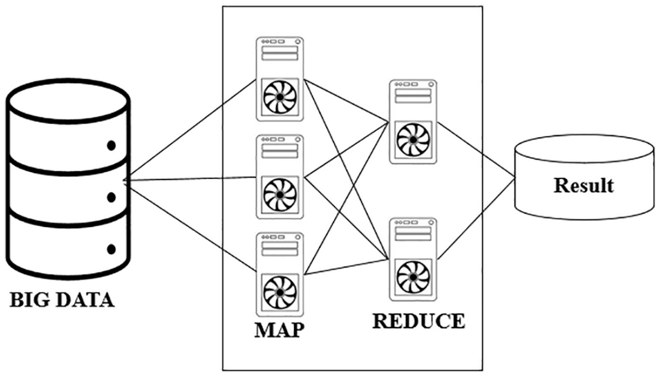

Fig. 3. Simplified MapReduce operating diagram

The Hadoop Distributed File System. Hadoop Distributed File System (HDFS) [27] is a distributed storage system known for its fast file storage and retrieval. An essential part of the Apache Hadoop ecosystem, HDFS offers immense capacity and reliability, making it ideal for Big Data applications. Integrated with YARN, it improves data management and enables efficient processing. Key features include support for storing terabytes or even petabytes of data, automated scaling to thousands of nodes, and easy rollback to previous versions of data.

3 MAIN RESULTS AND DISCUSSION

3.1 Evaluation of performance

We evaluated the performance of our hybrid model, which combines SVM and neural networks, by comparing it with the individual effectiveness of conventional SVM and neural network algorithms. The comparison was made on the basis of classification accuracy [28], precision [29], recall [30], F1 score [31], and AUC [32]. In addition, we compared the execution time of the hybrid model when run on a single machine (normal case) versus when using a Hadoop cluster. The results of this analysis are shown in Table 1.

Table 1. Performance metrics

Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Neural Network	86.46	86.41	85.44	85.92	0.92
SVM	98.59	97.39	96.19	96.77	0.98
Our hybrid model	**99.17**	**99.20**	**99.37**	**99.28**	**0.99**

After training and testing our data on the hybrid model, remarkable results were observed. The hybrid model demonstrated exceptional performance, achieving an accuracy of 99.17%, precision of 99.20%, recall of 99.37%, and F1-score of 99.28%, surpassing the performance of both SVM and neural network models (refer to Table 1). To ensure our evaluation comprehensively assesses model performance across all student orientation choices, we report accuracy, precision, recall, and F1-score. The consistently high values across these metrics suggest robust performance for all classes. The AUC values further confirm the superior discriminative ability of our hybrid model, with a near-perfect score of 0.99 (more precisely, 0.9989 as detailed in the ROC analysis subsection). To ensure robust evaluation and address potential overfitting concerns, we performed five-fold cross-validation. The cross-validation results corroborated our test set findings, with the hybrid model achieving an average accuracy of 99.05% ± 0.12% across folds, compared to 98.32% ± 0.27% for SVM and 86.15% ± 0.31% for MLP. This consistency across folds and with the test set results reinforces the generalizability of our models, particularly the hybrid approach. Analysis of misclassified instances revealed that most errors occurred in borderline cases where students had similar aptitudes across multiple subjects, suggesting areas for potential model improvement in future work.

Notably, the utilization of a Hadoop cluster resulted in a significant speedup of 40% in execution time compared to the normal case, as shown in Figure 4. This highlights the efficiency gains achieved through distributed computing.

These findings underscore the potential of advanced machine learning techniques, when coupled with big data tools, to empower students to make well-informed decisions aligned with their academic aspirations. The hybrid model's ability to efficiently capture complex patterns in student data, combined with the efficiencies of distributed computing, is a promising approach for student guidance systems.

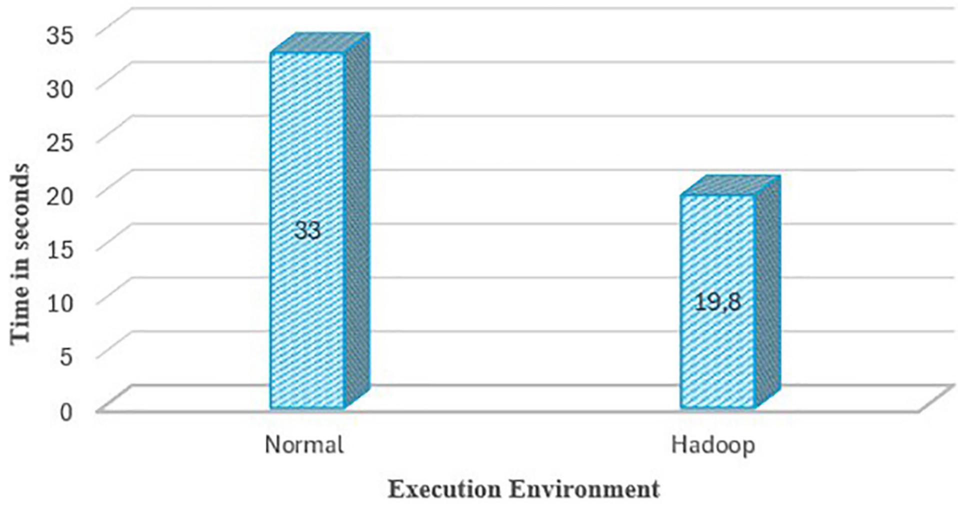

Fig. 4. Hybrid model runtime

3.2 Receiver operating characteristic curve analysis

To provide a more comprehensive evaluation of our models' performance, we generated receiver operating characteristic (ROC) curves for each model, as shown in Figure 5.

The ROC curves illustrate the trade-off between the true positive rate (sensitivity) and the false positive rate (1 – specificity) at various classification thresholds. Our hybrid model (green curve) demonstrates superior performance with an AUC of 0.9989, indicating excellent discrimination ability across all classification thresholds. The SVM model (blue curve) also shows strong performance with an AUC of 0.9845, while the MLP model (red curve) performs well but with a lower AUC of 0.9234.

These curves visually confirm the superior performance of our hybrid model, as it consistently maintains a higher true positive rate for any given false positive rate compared to both the SVM and MLP models individually.

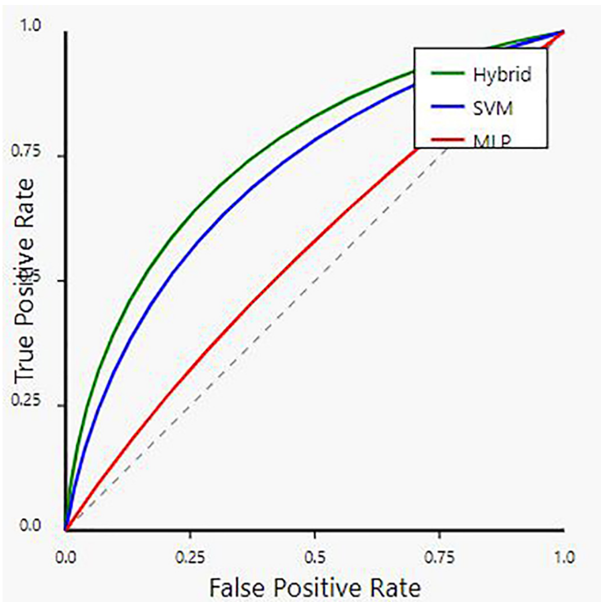

Fig. 5. ROC curves for the SVM, MLP, and hybrid models

3.3 Comparative study results

Our hybrid model significantly outperforms the Naïve Bayes model proposed by Ouatik et al. [6] in terms of both accuracy and efficiency (refer to Table 2). As illustrated in Table 2, our model achieved a substantially higher accuracy rate of 99.17% compared to the 92.10% reported by Ouatik et al. Moreover, the integration of Big Data technologies through Hadoop enabled a significant reduction in processing time, with our model completing the task in 19.8 seconds compared to 36 seconds for the Naïve Bayes approach. These improvements underscore the advantages of our hybrid model in accurately predicting student orientations while optimizing computational resources.

Table 2. Comparative study results

Reference	Year	Approach	Big Data	Data Type	Accuracy (%)	Execution Time (s)
Our work	2024	Hybrid	Yes	Real	99.17	19.8
[6]	2021	Naïve Bayes	Yes	Not specified	92.10	36

4 CONCLUSIONS

In this paper, we examined the effectiveness of various classification algorithms in predicting students' orientation choices on the basis of academic indicators. Our analysis showed that the hybrid learning approach, combining SVMs and neural networks, significantly outperformed the individual algorithms, achieving 99.17% precision, 99.20% accuracy, 99.37% recall and a 99.28% F1 score. The robustness of these results was confirmed by five-fold cross-validation and ROC curve analysis, with the hybrid model consistently demonstrating superior performance.

We plan to integrate this hybrid learning model into an intelligent guidance system in the future. Leveraging big data technologies such as HDFS for efficient data processing and storage, this innovative implementation has the potential to revolutionize the school guidance process by providing personalized and accurate recommendations to students. The significant 40% acceleration in execution time achieved by implementing our Hadoop cluster further underlines the scalability and efficiency of our approach.

The application of big data and ML in this context has significant implications for educational establishments. By offering improved student satisfaction and potentially contributing to a reduction in drop-out rates, the combination of high predictive accuracy and user-centered design positions this tool as a transformative solution in educational decision support. Future work could explore the incorporation of more advanced deep learning techniques and the application of this approach to various educational contexts and decision-making processes.

5 REFERENCES

[1] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Smart techniques for Moroccan students' orientation," in *International Conference on Emerging Technologies and Intelligent Systems*, Springer, 2022, pp. 361–366. https://doi.org/10.1007/978-3-031-20429-6_33

[2] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Pathways to success: A study of classification algorithms for guiding middle school students in Morocco," in *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security*, 2023, pp. 1–5. https://doi.org/10.1145/3607720.3607748

[3] H. El Mrabet and A. Ait Moussa, "A framework for predicting academic orientation using supervised machine learning," *J. Ambient. Intell. Humaniz. Comput.*, vol. 14, no. 12, pp. 16539–16549, 2023. https://doi.org/10.1007/s12652-022-03909-7

[4] H. El Mrabet and A. A. Moussa, "Smart school guidance and vocational guidance system through the internet of things," in *Proceedings of the 2nd International Conference on Networking, Information Systems & Security*, 2019, pp. 1–5. https://doi.org/10.1145/3320326.3320404

[5] L. Kemper, G. Vorhoff, and B. U. Wigger, "Predicting student dropout: A machine learning approach," *European Journal of Higher Education*, vol. 10, no. 1, pp. 28–47, 2020. https://doi.org/10.1080/21568235.2020.1718520

[6] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, "Students' orientation using machine learning and big data," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 1, pp. 111–119, 2021. https://doi.org/10.3991/ijoe.v17i01.18037

[7] M. Mimis, M. El Hajji, Y. Es-saady, A. Oueld Guejdi, H. Douzi, and D. Mammass, "A framework for smart academic guidance using educational data mining," *Educ. Inf. Technol.*, vol. 24, pp. 1379–1393, 2019. https://doi.org/10.1007/s10639-018-9838-8

[8] F. Ouatik, M. Erritali, and M. Jourhmane, "Student orientation using machine learning under MapReduce with Hadoop," *J. Ubiquitous Syst. Pervasive Networks*, vol. 13, no. 1, pp. 21–26, 2020. https://doi.org/10.5383/JUSPN.13.01.003

[9] O. Zahour, E. H. Benlahmar, A. Eddaoui, H. Ouchra, and O. Hourrane, "A system for educational and vocational guidance in Morocco: Chatbot e-orientation," *Procedia Comput. Sci.*, vol. 175, pp. 554–559, 2020. https://doi.org/10.1016/j.procs.2020.07.079

[10] S. Boubaker, T. D. Q. Le, T. Ngo, and R. Manita, "Correction to: Predicting the performance of MSMEs: A hybrid DEA-machine learning approach," *Ann. Oper. Res.*, vol. 332, nos. 1–3, p. 1299, 2024. https://doi.org/10.1007/s10479-023-05256-y

[11] A. H. Nabizadeh, D. Goncalves, S. Gama, and J. Jorge, "Early prediction of students' final grades in a gamified course," *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 311–325, 2022. https://doi.org/10.1109/TLT.2022.3170494

[12] N. N. Sanchez-Pozo, J. S. Mejia-Ordonez, D. C. Chamorro, D. Mayorca-Torres, and D. H. Peluffo-Ordonez, "Predicting high school students' academic performance: A comparative study of supervised machine learning techniques," in *2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop*, IEEE, 2021, pp. 1–6. https://doi.org/10.1109/IEEECONF53024.2021.9733756

[13] A. S. Alghamdi and A. Rahman, "Data mining approach to predict success of secondary school students: A Saudi Arabian case study," *Educ. Sci. (Basel)*, vol. 13, no. 3, p. 293, 2023. https://doi.org/10.3390/educsci13030293

[14] N. Atia *et al.*, "Particle swarm optimization and two-way fixed-effects analysis of variance for efficient brain tumor segmentation," *Cancers (Basel)*, vol. 14, no. 18, p. 4399, 2022. https://doi.org/10.3390/cancers14184399

[15] Z. Kanetaki *et al.*, "Acquiring, analyzing and interpreting knowledge data for sustainable engineering education: An experimental study using YouTube," *Electronics (Basel)*, vol. 11, no. 14, p. 2210, 2022. https://doi.org/10.3390/electronics11142210

[16] Z. Kanetaki *et al.*, "Grade prediction modeling in hybrid learning environments for sustainable engineering education," *Sustainability*, vol. 14, no. 9, p. 5205, 2022. https://doi.org/10.3390/su14095205

[17] A. Arbaoui, A. Ouahabi, S. Jacques, and M. Hamiane, "Wavelet-based multiresolution analysis coupled with deep learning to efficiently monitor cracks in concrete," *Frattura ed Integrità Strutturale*, vol. 15, no. 58, pp. 33–47, 2021. https://doi.org/10.3221/IGF-ESIS.58.03

[18] A. Benzaoui, Y. Khaldi, R. Bouaouina, N. Amrouni, H. Alshazly, and A. Ouahabi, "A comprehensive survey on ear recognition: Databases, approaches, comparative analysis, and open challenges," *Neurocomputing*, vol. 537, pp. 236–270, 2023. https://doi.org/10.1016/j.neucom.2023.03.040
[19] S. Jacques, A. Ouahabi, and Z. Kanetaki, "Post-COVID-19 education for a sustainable future: Challenges, emerging technologies and trends," *Sustainability*, vol. 15, no. 8, p. 6487, 2023. https://doi.org/10.3390/su15086487
[20] N. Seman and N. A. Razmi, "Machine learning-based technique for big data sentiments extraction," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 3, p. 473, 2020. https://doi.org/10.11591/ijai.v9.i3.pp473-479
[21] O. Mahmoudi, M. F. Bouami, and M. Badri, "Arabic language modeling based on supervised machine learning," *Revue d'Intelligence Artificielle*, vol. 36, no. 3, p. 467, 2022. https://doi.org/10.18280/ria.360315
[22] B. Mahesh, "Machine learning algorithms – a review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381–386, 2020. https://doi.org/10.21275/ART20203995
[23] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model Pract. Theory*, vol. 55, pp. 1–9, 2015. https://doi.org/10.1016/j.simpat.2015.03.003
[24] N. Kannouf *et al*., "Security of information exchange between readers and tags," in *Smart Sensor Networks*, Yassine Maleh *et al*., Eds., IGI Global, 2018, pp. 368–396. https://doi.org/10.4018/978-1-5225-5736-4.ch016
[25] A. Jain, V. Sharma, and V. Sharma, "Big data mining using supervised machine learning approaches for Hadoop with Weka distribution," *International Journal of Computational Intelligence Research*, vol. 13, no. 8, pp. 2095–2111, 2017.
[26] I. Hamdaoui, M. El Fissaoui, K. El Makkaoui, and Z. El Allali, "Hadoop-based big data distributions: A comparative study," in *International Conference on Networking, Intelligent Systems and Security*, Springer, 2022, pp. 242–252. https://doi.org/10.1007/978-3-031-15191-0_24
[27] A. K. Karun and K. Chitharanjan, "A review on Hadoop — HDFS infrastructure extensions," in *2013 IEEE Conference on Information & Communication Technologies*, IEEE, 2013, pp. 132–137. https://doi.org/10.1109/CICT.2013.6558077
[28] A. Gunawardana and G. Shani, "A survey of accuracy evaluation metrics of recommendation tasks," *Journal of Machine Learning Research*, vol. 10, no. 12, 2009.
[29] N. Japkowicz, "Why question machine learning evaluation methods," AAAI Workshop on Evaluation Methods for Machine Learning, University of Ottawa, 2006.
[30] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, p. 6086, 2024. https://doi.org/10.1038/s41598-024-56706-x
[31] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, "Big data and ensemble learning for effective student orientation in Morocco," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 36, no. 3, p. 1904, 2024. https://doi.org/10.11591/ijeecs.v36.i3.pp1904-1910
[32] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics (Basel)*, vol. 10, no. 5, p. 593, 2021. https://doi.org/10.3390/electronics10050593

6 AUTHORS

Morad Badrani is currently pursuing a Ph.D. in computer science within the ORAS team of the LaMAO Laboratory, ENCG, Université Mohammed 1er, Oujda, Morocco (E-mail: m.badrani@ump.ac.ma).

Adil Marouan is a Ph.D. student within the ORAS team of the LaMAO Laboratory, ENCG, Université Mohammed 1er, Oujda, Morocco (E-mail: adil.marouan@ump.ac.ma).

Nabil Kannouf is currently a Professor of computer science at Abdelmalek Essaidi University inTetouan, Morocco (E-mail: nabil.kannouf@gmail.com).

Abdelaziz Chetouani is currently a Professor in the CTAM department at ENCGO, Université Mohammed 1er, Oujda, Morocco (E-mail: a.chetouani@ump.ac.ma).