Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Exploring the Efficacy of Methods used in Student Orientation Systems: A Review

Reguragui Samira
Abdelmalek Essaadi University
Tetuan, Morocco
samira.reguragui@etu.uae.ac.ma
https://orcid.org/0000-0002-3959-1907

Bendahmane Ahmed
Abdelmalek Essaadi University
Tetuan, Morocco
abendahman@uae.ac.ma

Abstract—**Guiding students in their academic and professional trajectories is an important aspect of education. This study explores the efficacy of various machine learning and Multi-Criteria Decision Making (MCDM) methods for student orientation systems. We analyze a comprehensive set of algorithms (Naïve Bayes, SVM, Neural networks, KNN, Random Forest, Decision Tree, Linear regression, Logistic regression, TOPSIS, AHP) and compare their performance based on accuracy, execution time, and other relevant criteria. Our key findings reveal that Naive Bayes, SVM, Random Forest, Decision Tree, and TOPSIS consistently deliver high accuracy. However, each method has limitations. To address these, we propose utilizing hybrid approaches that combine multiple algorithms. This research offers a novel two-fold contribution: (1) exploring diverse algorithms tailored for student orientation, and (2) providing a comparative performance analysis. We acknowledge the need for further research on refining existing methods, developing effective hybrid models, and addressing specific challenges to optimize student orientation system efficacy. Therefore, this study provides valuable insights for researchers and developers in advancing student guidance systems.**

Keywords—Student orientation; Machine learning; MCDM methods; Hybrid approaches; Artificial intelligence; Smart school guidance; Academic guidance

I. Introduction

Student orientation plays a critical role in shaping the academic trajectory of individuals, influencing their choice of academic specialization and, consequently, their future careers. The ramifications of a misguided academic decision are substantial, potentially leading to academic setbacks, dropout, and prolonged academic journeys. The conventional method of academic guidance, reliant on manual assessment by academic specialists, presents challenges due to its inherent slowness and limited consideration of factors beyond grades [1].

Indeed, the significance of informed academic decisions is highlighted by the observation that many students face challenges in determining their study areas at universities [1]. Factors such as familial influence, peer guidance, and family career history often overshadow individual passions and interests, leading to dissatisfaction, high dropout rates, major changes, and delayed graduations [2]. The lack of personalized guidance and limited online resources exacerbate the challenge, increasing the risk of students making suboptimal career choices [3].

To address these issues, artificial intelligence (AI)-based systems have been designed to provide personalized guidance to students. These systems include recommendation systems [3]–[8] expert systems [9], [10] and chatbots [11], [12].

Researchers have conducted comparative studies utilizing various machine learning and multi-criteria decision-making (MCDM) methods to identify the most effective algorithms for implementing optimal guidance systems.

This paper seeks to contribute to the existing literature by presenting and analyzing the performance evaluation results of different approaches in the context of academic and professional orientation for students. Drawing on a diverse range of machine learning algorithms and multi-criteria decision methods, the paper offers an overview of their effectiveness in student orientation.

The rest of the paper is structured as follows. In Section II, we briefly present the most popular and widely used algorithms in the field of educational guidance. The findings of comparative studies on the performance of various algorithms applied in the student orientation system are presented and discussed in section III. Finally, we conclude the paper with a brief summary and discuss the future work in Section IV.

II. Background

This section provides an overview of diverse machine learning and multi-criteria decision-making methods commonly used in student orientation systems.

A. K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a popular supervised learning algorithm that can be used for classification and regression tasks. It is based on the assumption that similar data points can be found close to each other.

To determine the class of a new case, the algorithm relies on the following principle: It searches for the k nearest neighbors of the new case. Then, it selects the closest and most frequent result among the found candidates [13]. This method primarily utilizes two parameters: the value of k and a similarity function for comparing already classified cases with the new case [13].

B. Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification, regression, and anomaly detection. It mainly relies on the use of functions called kernels, which facilitate data separation [14].

The goal of this method is to find an optimal hyperplane that separates the data and maximizes the distance between classes [14]. The closest points, which are solely used for hyperplane determination, are called support vectors.

SVM was developed by Vapnik in 1995 and is known for its ease of use, high flexibility, and strong theoretical guarantees.

C. Linear Regression

Linear regression is a supervised learning algorithm used to understand relationships within data. It predicts the value of a dependent variable based on the value of an independent variable.

The terms dependent/independent come from the assumption that the dependent variable depends on the independent variables, which, in turn, are not dependent on the others.

D. Logistic Regression

Logistic regression, like linear regression, is a supervised learning algorithm used to understand relationships within data. It is employed when the dependent variables are binary.

E. Naïve Bayes

Naïve Bayes is a popular supervised machine learning algorithm used for classification. It is primarily based on Bayes' theorem. This probabilistic classifier is termed naïve because it assumes that the features of input data (variables) are conditionally independent. Naïve Bayes is known for its simplicity and speed.

F. Decision Tree

Decision tree is one of the most popular and widely used supervised learning methods for classification and result visualization. It serves as a decision-making and Data mining tool, representing a set of classification rules in tree form.

There are several algorithms for constructing a decision tree, including CART (Classification and Regression Tree), ID3, C4.5, Rnd Tree, and RF. The basic idea of all algorithms is to partition the attribute space into branches and leaves until the data is classified, satisfying a stopping condition [15].

G. Random Forest

Random Forest is a supervised learning algorithm used for classification and regression problems. It relies on randomization to create a large number of decision trees. The output of these trees is aggregated into a single output using voting for classification problems or averaging for regression problems.

H. Neural Network

Neural Network is a set of interconnected nodes (artificial neurons) allowing the solution of complex problems. It is organized into multiple layers, including an input layer, one or more hidden layers, and an output layer. Nodes in one layer are connected to all nodes in the next layer.

Artificial neural networks are systems inspired by the functioning of biological neurons, with the multilayer perceptron (MLP) being the most famous—one capable of learning from experience.

I. Technique for Order Preference by Similarity to Ideal Solution

Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), originated by Yoon and Hwang in 1981, is a popular method of Multi-Criteria Decision Making (MCDM) used to make optimal decisions in multicriteria problems.

It involves choosing an alternative from a set of alternatives based on the shortest distance to the positive ideal alternative (the best alternative in all criteria) and the greatest distance to the negative ideal alternative (the one degrading all criteria).

J. Analytical Hierarchy Process

Analytical Hierarchy Process (AHP), developed by Saaty in 1992, is one of the strongest and most used MCDM methods in multi-criteria evaluation. It is based on constructing a hierarchical model between criteria, alternatives, and the objective.

Application of this method occurs at two levels: the hierarchical structure and the evaluation. The drawback of this method is the instability of ranking various alternatives when dealing with a large number of alternatives.

III. RESULTS AND DISCUSSION

Comparative studies were conducted to identify the most effective approach for the student orientation system, involving various machine learning and MCDM methods. The results presented in Table 1 offer a comprehensive overview of the performance of different techniques, considering accuracy, execution time, and other relevant criteria.

According to the table, certain methods stand out for their superior performance compared to others, namely Naïve Bayes, SVM, Random Forest, Decision Tree, and TOPSIS. The Naïve Bayes classifier demonstrated exceptional accuracy, surpassing even more sophisticated methods such as Neural Networks (MLP) [16], [17]. It is recognized for its simplicity of implementation and speed.

However, its performance may be constrained in scenarios with a high number of dependencies between variables.
In another comparative study [20], SVM surpassed Naïve Bayes and other machine learning algorithms such as Neural Network and Decision Tree. This classification technique is known for its ease of use, high flexibility, solid theoretical guarantees, and its ability to handle large amounts of data. It yields relevant results but is time-consuming compared to other methods, such as Logistic regression and Decision tree [19]. It requires a significant amount of time to find the best parameters during the testing phase.

Regarding the two methods, Random Forest and Decision Tree, they presented accurate and promising results. Decision Tree demonstrated greater precision compared to classification algorithms [19], [21] like Logistic Regression, KNN, SVM and Naïve-Bayes. This technique is widely used for its simplicity, readability, and execution speed compared to other methods. It is easily understandable and interpretable by teachers and guidance counselors [16]. However, its performance tends to decrease when dealing with numerous classes. Similarly, Random Forest outperformed other classification methods [2], [18] even more sophisticated ones like MLP [2]. It is generally more effective than simple decision trees, although it may be less interpretable and not suitable for small samples.

For Multi-criteria decision methods, the results favored TOPSIS over AHP [1]. The TOPSIS method proved to be more accurate and, notably, easier to apply in the context of the student orientation system.

TABLE I. COMPREHENSIVE OVERVIEW OF THE PERFORMANCE OF DIFFERENT METHODS USED IN STUDENT ORIENTATION SYSTEM

Authors	**Year**	**Methods**	**Results**						
			Accuracy (%)	***Execution time (s)***	***AUC***	***Time-consuming (ms)***		***F1-score (%)***	***Other Criteria***
						Training	***Prediction***		
Mimis et al. [16]	2019	Naive Bayes	58,69						Cohen's Kappa
		Neural networks (MLP)	56,10						
		Decision tree (C4.5)	54,71						
Kamal et al. [2]	2020	Random Forest	96,10						Precision, Recall and Kappa
		MLP	92,20						
Tarik and Farhaoui [18]	2020	Random Forest	44,2						
		Decision tree	29,87						
		Linear regression	26,80						
Ouatik et al. [17]	2021	Naïve Bayes	92,10	36					
		Neural Network	90,37	42					
		SVM	88,13	41					
		Random Forest	86,22	28					
El Mrabet and Ait Moussa [19]	2023	Decision tree			0,80	410,493	174		
		Logistic regression			0,73	185,000	10		
		SVM			0,72	480,000	10		
		KNN			0,65	N/A	730,000		
Ouatik et al. [1]	2022	TOPSIS	91,35						Complexity
		AHP	90,83						
Mystere et al. [20]	2023	SVM	70					0,66	Precision and Recall
		Naïve Bayes	65					1	
		Neural Network	64					0,37	
		Decision tree	52					0,46	
Badrani et al. [21]	2023	Decision tree						83,30	Precision and Recall
		Naïve-Bayes						23,25	
		KNN						58,30	

In a nutshell, Naive Bayes, SVM, Random Forest, Decision Tree, and TOPSIS emerged as the best-selected methods for the student orientation system. These approaches consistently delivered high-performance and accurate results, albeit with recognized limitations. In response to these limitations, we propose the use of hybrid approaches, combining two or more algorithms. Hybridization can potentially overcome individual method constraints and enhance overall system performance in student orientation systems. The same observation was described by Mikrat et al. [22], who proposed a hybrid approach that combines KNN, Fuzzy Logic, or FAHP, and the Internet of Things to obtain a more precise and reliable result.

It is essential to acknowledge that the precision of each model is contingent upon the introduced variables (criteria) and the specific technique employed. The complexity of the task and characteristics of the dataset play pivotal roles in determining the efficacy of the chosen method.

IV. Conclusion and Future Directions

This study sheds light on the efficacy of various machine learning and Multi-Criteria Decision Making (MCDM) methods in student orientation systems. By analyzing a comprehensive set of algorithms, including Naive Bayes, SVM, Random Forest, Decision Tree, TOPSIS, and AHP, it offers a comparative assessment based on accuracy, execution time, and other relevant criteria.

The study's novelty lies in its two-fold approach: firstly, exploring a diverse range of algorithms specifically suited for student orientation, and secondly, providing a performance comparison across these methods. It further identifies inherent limitations in each approach and proposes the development of hybrid models to address these shortcomings. While the study acknowledges the potential of hybrid approaches, future research is warranted to explore specific combinations that can maximize effectiveness.

Recognizing the limitations of model dependence on training data and the need for tailoring solutions to individual systems, the study emphasizes the importance of further research. This includes refining existing methods, delving deeper into hybrid models, and tackling specific challenges to optimize student orientation system efficacy. Overall, this research provides valuable insights for researchers and developers, paving the way for advancements in student guidance systems.

References

[1] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, “Student Orientation Recommender System using TOPSIS and AHP”, *Journal of Information and Organizational Sciences*, vol. 46, no 2, pp. 473-489, 2022. https://doi.org/10.31341/jios.46.2.12

[2] N. Kamal, F. Sarker, and K. A. Mamun, “A comparative study of machine learning approaches for recommending university faculty”, in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0* (STI), IEEE, 2020, p. 1-6, doi: 10.1109/STI50764.2020.9350461.

[3] M. Qamhieh, H. Sammaneh and M. N. Demaidi, “PCRS: Personalized Career-Path Recommender System for Engineering Students”, in *IEEE Access*, vol. 8, pp. 214039-214049, 2020, doi: 10.1109/ACCESS.2020.3040338.

[4] G. Meryem, K. Douzi and S. Chantit, “Toward an E-orientation platform: Using hybrid recommendation systems”, *2016 11th International Conference on Intelligent Systems: Theories and Applications* (SITA), Mohammedia, Morocco, 2016, pp. 1-6, doi: 10.1109/SITA.2016.7772305.

[5] M. S. Khan and M. Darbari, “Myers-Briggs Type Indicator and Social Media Mining Based Career Counselling System”, *Partners Universal International Research Journal*, vol. 1, no 2, pp. 35-42, Jun. 2022, https://doi.org/10.5281/zenodo.6726646.

[6] A. Pantoja-Vallejo and B. Berrios-Aguayo, “TIMONEL: Recommendation System Applied to the Educational Orientation of Higher Education Students”, In: Serrhini, M., Silva, C., Aljahdali, S. (eds) Innovation in Information Systems and Technologies to Support Learning Research. EMENA-ISTL 2019. Learning and Analytics in Intelligent Systems, vol 7. Springer, Cham., pp. 14–26, 2020. Available: https://doi.org/10.1007/978-3-030-36778-7_2

[7] B. E. V. Comendador, W. F. C. Becbec, and J. R. P. de Guzman, “Implementation of fuzzy logic technique in a decision support tool: Basis for choosing appropriate career path”, *International Journal of Machine Learning and Computing*, vol. 10, no. 2, pp. 339-345, 2020. doi: 10.18178/ijmlc.2020.10.2.940

[8] M. C. B. Natividad, B. D. Gerardo, and R. P. Medina, “A fuzzy-based career recommender system for senior high school students in K to 12 education”, in *IOP Conference Series: Materials Science and Engineering*, *IOP Publishing*, 2019, vol. 482, no. 1, p. 012025. doi:10.1088/1757-899X/482/1/012025

[9] E. EL HAJI and A. Azmani, “Proposal of a Digital Ecosystem Based on Big Data and Artificial Intelligence to Support Educational and Vocational Guidance”, *International Journal of Modern Education and Computer Science(IJMECS)*, vol. 12, no 4, pp. 1-11, 2020. doi: 10.5815/ijmecs.2020.04.01

[10] S. Myla, S. Marella, S. Goud, H. Shaik, S. Inthiyaz, and G. Santosh Kumar, “Design Decision Taking System For Student Career Selection For Accurate Academic System”, *International Journal of Scientific & Technology Research*, vol. 8, pp. 2199-2206, sept. 2019.

[11] 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 Computer Science*, vol. 175, pp. 554-559, 2020. Available: https://doi.org/10.1016/j.procs.2020.07.079

[12] L. H. Su, T. Dang-Huy, T. Thi-Yen-Linh, N. Thi-Duyen-Ngoc, L. Bao-Tuyen, and N. Ha-Phuong-Truc, “Development of an AI Chatbot to support admissions and career guidance for universities”, *International Journal of Emerging Multidisciplinary Research*, vol. 4, no 2, pp. 11-17, 2020. Available: https://doi.org/10.22662/IJEMR.2020.4.2.013

[13] C. T. Quang, “Classification automatique des textes vietnamiens Hanoi”, Institut de la Francophonie pour l’informatique, 2005.

[14] M. Hasan and F. Boris, “Svm: Machines à vecteurs de support ou séparateurs à vastes marges”, *Rapport technique*, Versailles St Quentin, France. Cité, vol. 64, 2006.

[15] B. R. Patel and K. K. Rana, “A survey on decision tree algorithm for classification”, *International Journal of Engineering Development and Research*, vol. 2, no 1, pp. 1-5, 2014.

[16] 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”, *Education and Information Technologies*, vol. 24, no 2, pp. 1379-1393, 2019. Available: https://doi.org/10.1007/s10639-018-9838-8

[17] 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. 01, pp. 111, Jan. 2021, doi: 10.3991/ijoe.v17i01.18037. Available: http://dx.doi.org/10.3991/ijoe.v17i01.18037

[18] A. Tarik and Y. Farhaoui, “Recommender System for Orientation Student”, In: Farhaoui, Y. (eds) Big Data and Networks Technologies. BDNT 2019. Lecture Notes in Networks and Systems, vol 81, Cham: Springer International Publishing, 2020, pp. 367-370. doi: 10.1007/978-3-030-23672-4_27. Available: https://doi.org/10.1007/978-3-030-23672-4_27

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

[20] K. K. Mystere, H. N. Mpia, and M. B. Vingi, “Prédiction de l’orientation des étudiants dans des filières d’études appropriées en utilisant les techniques de Data Mining [Predicting students’ orientation in suitable fields of study using data mining techniques]”, *International Journal of Innovation and Applied Studies*, vol. 39, no 1, pp. 193-208, 2023.

[21] M. Badrani, A. Marouan, N. Kannouf, and A. Chetouani, “Smart Techniques for Moroccan Students’ Orientation”. In: Al-Sharafi, M.A., Al-Emran, M., Al-Kabi, M.N., Shaalan, K. (eds) Proceedings of the 2nd International Conference on Emerging Technologies and Intelligent Systems. ICETIS 2022. Lecture Notes in Networks and Systems, vol 573, Cham: Springer International Publishing, 2023, pp. 361-366. https://doi.org/10.1007/978-3-031-20429-6_33

[22] H. Mikrat, M. Fakhouri Amr, A. Bahnasse, and M. Talea, “Review of the Student’s Orientation in the Age of New Technologies”, *Procedia Computer Science*, vol. 203, pp. 683-687, 2022, doi: 10.1016/j.procs.2022.07.101. Available: https://doi.org/10.1016/j.procs.2022.07.101