Journal of Theoretical and Applied Information Technology
15th June 2025. Vol.103. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

RECOMMENDATION SYSTEM FOR E-LEARNING STUDENT ORIENTATION BASED ON MACHINE LEARNING ALGORITHMS AND QCM

MOULAY AMZIL[1], AHMED ELGHAZI[2], MOHAMED ERRITALI [3]

[1]Data science for sustainable Earth laboratory (DataEarth), Sultan Moulay Slimane University, Beni Mellal 23000, Morocco

[2]Department of Computer Science FST Beni Mellal, Sultan Moulay Slimane University, Beni Mellal,Morocco

[3]Data science for sustainable Earth laboratory (DataEarth), Sultan Moulay Slimane University, Beni Mellal 23000, Morocco

E-mail: [1] amzilmoulay@gmail.com , [2] hmadgm@yahoo.fr, [3]m.erritali@usms.ma

ABSTRACT

This article is part of a student guidance project. It involves a recommendation and classification system (E_orientation) based on a real MCQ test. This test is administered to the student and provides random answers that are then used by our recommendation system. We use binary modelling of the answers, which produces a vector of data for each student at the end of the test. For modelling and classification, we use several machine-learning algorithms to optimize the accuracy of the recommendations. The results of the experiment show that Random Forest is the best model (85.93% accuracy), ahead of SVM (80.43%). KNN achieves 76.10%, and the Decision Tree, Logistic Regression and Naive Bayes algorithms have the lowest performance (Accuracy ≤ 67.03%). We can therefore improve the students' method of orientation by basing it on simple technical questions. This will also improve their contribution to the labor market.

Keywords: *Recommendation System, School Guidance, Machine Learning, E-Learning Platform, MCQ Test*

1. INTRODUCTION

For several years now, educational and career guidance mechanisms have been at the heart of educational, political and social debates [36]. At the same time, the worldwide success of artificial learning and intelligence techniques is undeniable. The education sector has been particularly affected by this technological revolution, arousing the interest of numerous researchers [17], [21], [22], [24]. Their work focuses mainly on the issue of predicting students' university choices....

School and career guidance is a major issue for students, teachers and guidance counsellors, particularly when it comes to making decisions about their future studies [42], [43]; however, thanks to artificial intelligence, which is playing an increasing role in this field, particularly with the application of machine learning, which can analyze large quantities of data on students' skills, interests and performance to provide more personalized and precise guidance recommendations, the future of education is becoming increasingly complex.

In this context, our research will attempt to address the issue of educational and vocational guidance at the decision-making level. To do this, we use machine-learning algorithms group students according to the results of a test administered on an e-learning platform. Each student is invited to take this online MCQ test, which is then transformed into an automatic correction and a vector of binary data. This will be the input for a classification system (E_orientation) of students, after these results, in groups for each specialty.

This system will make it easier to recommend courses of study and avoid the problems of choosing the wrong subject. However, several factors can influence a student's orientation: the family's social level, the student's attendance record, his ambitions, etc.... Most of these factors are qualitative and cannot be measured directly, so to collect them you need to conduct a survey of the student's life.

In the literature, the dynamics of educational and vocational guidance have been the subject of number of research projects over the years. On the one hand, most of these studies present good results based on artificial learning algorithms. On the other hand, the data used for decision-making is always incomplete and does not in fact reflect the various attributes linked to the student's life: his or her attendance and technical skills.

The contribution of this article is therefore to use machine-learning models to analyze students' responses and predict the educational paths that correspond to their academic skills.

The distinguishing feature of our study is the utilization of machine learning algorithms to provide students with personalised guidance, informed by their performance in an online multiple-choice question (MCQ) test. In contradistinction to conventional academic guidance systems, this system provides personalized recommendations based on students' actual performance, thereby enabling a more dynamic and individualized approach.

To develop our research topic, we decided to organize and divide our work into four sections structured as follows:

- The first section presents the fundamental concepts and the general context of academic guidance assisted by Machine Learning.
- The second section consists of presenting the state of the art in existing work on Machine Learning applied to academic guidance.
- The third section will present the methodology adopted, which will focus on the methods used and the work proposed.
- The fourth section presents the experimental results obtained.

The final section summarizes the results, followed by a discussion of future work prospects. These are aimed at improving educational support, reducing misdirection and fostering the professional development of young people.

2. BACKGROUND

2.1 Machine learning algorithm

Machine learning algorithms can process a wide range of input data and producing forecasts, thereby contributing to cost reduction, the enhancement of business procedures, and the improvement of customer service. These algorithms are employed to develop efficient and intelligent systems[1]. Understanding the math principles behind these algorithms is vital for a full comprehension of their methods and interpretations. The selection of the most appropriate algorithm depends on various factors in a study[2].

2.1.1 Supervised learning

Supervised machine learning algorithms are utilized for outcome prediction through training with labeled data. Different supervised learning algorithms include gradient boosting, random forest, K-nearest neighbors, support vector machines(SVM), decision trees, artificial neural networks, naive Bayes, and binary logistic regression. The performance of these algorithms has been assessed based on metrics such as accuracy, area under the curve, sensitivity, specificity, and kappa statistics. Random forest has been found to be the best-performing algorithm in terms of accuracy and precision in multiple studies[3].

2.1.2 Unsupervised learning

Unsupervised machine learning algorithms is a type of algorithm that is designed to create a learning paradigm for the sake of learning, its categorize and group data without labeled training data, and can identify patterns and relationships within the data, leading to the discovery of hidden structures and insights. The motivation of unsupervised learning is that although the data that goes through the unsupervised learning algorithms has a rich inherent structure, the ground truth and the metric used for training are typically sparse. This implies that the majority of what is learned by the algorithm should come from the input data structure rather than applying a specific understanding to a practical task[4].

2.1.3 Semi-Supervised learning

Semi-supervised learning is a pedagogical framework that amalgamates annotated and unannotated data to enhance supervised learning endeavors in situations where annotated data is limited or costly. It is of great interest in machine learning and data mining. Several models and algorithms have been proposed in the field, including self-training, mixture models, co-training, Multiview learning, graph-based methods, and semi-supervised support vector machines[5].Figure 1 shows the types of machine learning.

Journal of Theoretical and Applied Information Technology
15th June 2025. Vol.103. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

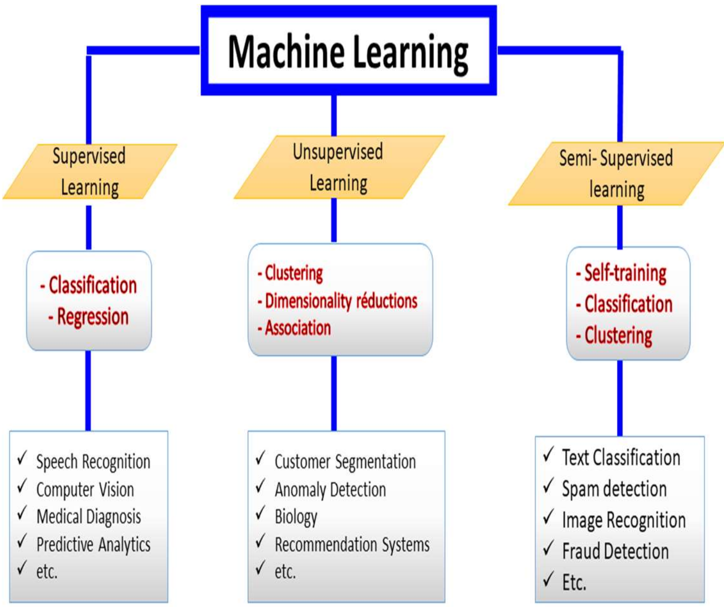


Figure 1 : Types of machine learning

2.2 Classification algorithms

In machine learning, classification is a key aspect of supervised learning, where systems learn from input data to classify new observations. Its importance grows with the complexity and unpredictability of real-world data [6], [26], [27], [28]. This study aims to identify learners' personalities based on their specific traits. Classification typically involves two phases: a learning phase, where a model is built from training data, and a testing phase, where the model classifies new data. To improve accuracy, fuzzy sets [7] are used to handle uncertainty. The study compares five major algorithms Logistic Regression, Random Forest, KNN, Decision Tree, Naive Bayes, and SVM to determine the most effective approach.

2.2.1 Naive Bayes Algorithm

Naive Bayes is a classification algorithm based on Bayes' theorem of conditional probabilities. This theorem describes the probability of an event occurring based on prior knowledge of conditions related to that event. It states that if we have a hypothesis H and evidence E, it is possible to calculate the probability of H being true given E. This algorithm is intuitive, simple, and robust, making it widely used for predictive modeling, with the main assumption that each feature is treated independently [8], [13]. Figure 2 illustrates a black triangle representing a new sample instance that needs to be classified into either the "star" class or the "square" class, depending on which category achieves the highest posterior probability. To ...object, a circle is drawn around the ...nclude several neighboring points, ...lgorithm to evaluate the surrounding ...d determine the most appropriate (Figure 2).

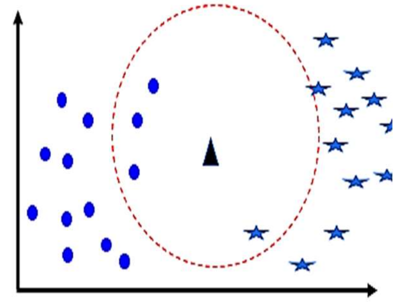


Figure 2 : Naive Bayes classification

2.2.2 Decision tree

The decision tree constitutes one of the most prevalent classification methods, and also one of the most antiquated. It emerged in the 1960s in the domains of psychology and sociology research [25], [9], [13]. This model is presented in the form of an intuitive tree structure that categorizes data items according to a hierarchical structure. Decision tree nodes generally have several levels, the first of which is called the root node that contains all the data (figure 3).

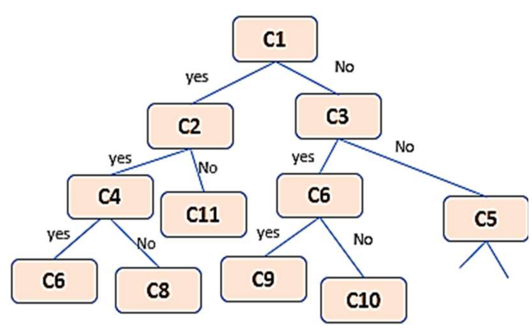


Figure 3. Decision tree classification

2.2.3 K-nearest neighbors

The k-nearest neighbors (KNN) algorithm is a classification algorithm that is both simple and fundamental in nature. It is one of the oldest

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

algorithms of its kind and functions by retaining a set of available cases for the purpose of classifying new examples based on a similarity measure [10],[35]. This non-parametric classification system is notable for the fact that it completely avoids the problem of probability densities. The term "K" in KNN refers to the number of nearest neighbors considered during the classification process, signifying that decisions are made based on the labels of the K nearest samples [11],[12],[13]. As illustrated in Figure 4, the KNN algorithm functions by classifying new objects. For instance, when K = 1, a new object (e.g., a triangle) is classified as a "circle". Conversely, when K = 4, the same object is classified as a "star" (figure 4).

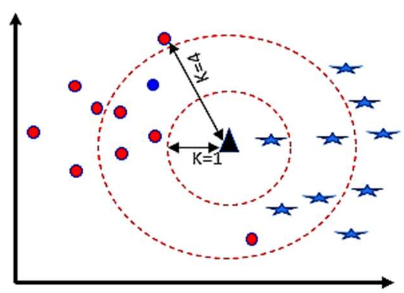


Figure 4: KNN classification

2.2.4 Support vector machine

The Support Vector Machine (SVM) is a simple yet powerful algorithm [25], [29],[30],[31] capable of producing accurate results with efficient computing power. It is one of the best-known classification algorithms, relying on linear or non-linear methods to separate data into different classes. The fundamental principle of SVM is to utilize a hyperplane (or line in the case of two-dimensional space) to differentiate between classes while maximizing the marginal distance between them, thereby minimizing classification errors. As illustrated in Figure 5, this marginal distance corresponds to the space between the hyperplane and the closest instance of each class, termed the support vector [12]. (figure 5).

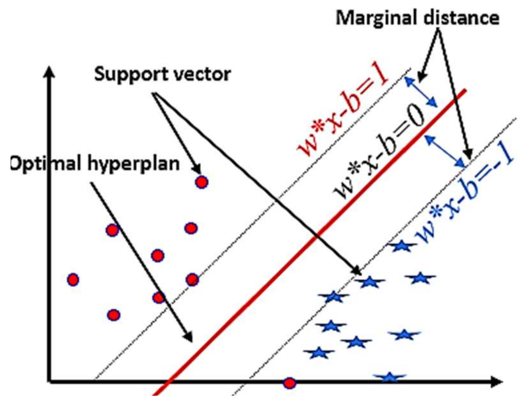


Figure 5 : SVM classification

2.2.5 Logistic regression

Logistic regression is a robust and well-established mathematical model of supervised classification, which is utilized extensively in the field of statistics for the estimation of the probability of an event occurring as a function of a set of input variables [12], [32], [33], [34]. This model operates primarily with binary data, that is to say, it evaluates the probability of an event occurring (P = 1) or not occurring (P = 0). To illustrate this application, consider the context of educational guidance, where the available data comprises two possible classes. In this scenario, logistic regression can be employed to ascertain the probability of receiving favorable guidance (P = 1) or unfavorable guidance (P = 0). (figure 6).

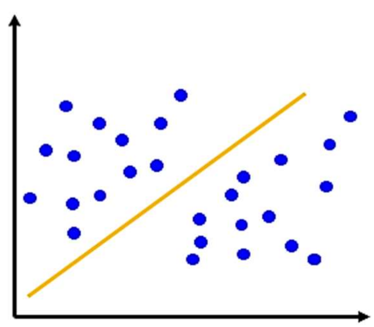


Figure 6 :Logistic regression classification

2.2.6 Random Forest (RF)

Introduced by Breiman [37], is an ensemble-learning algorithm that combines several decision trees via bagging to reduce overfitting and improve generalization. Its robustness is based on the random selection of data subsets (bootstrap) and variables at each split [36], optimizing predictor diversity. In classification, RF excels in a variety of domains (e.g. medical imaging), thanks to its ability to handle non-linear interactions. Its partial interpretability via variable importance [38] makes it a versatile tool for research and industry (figure 7).

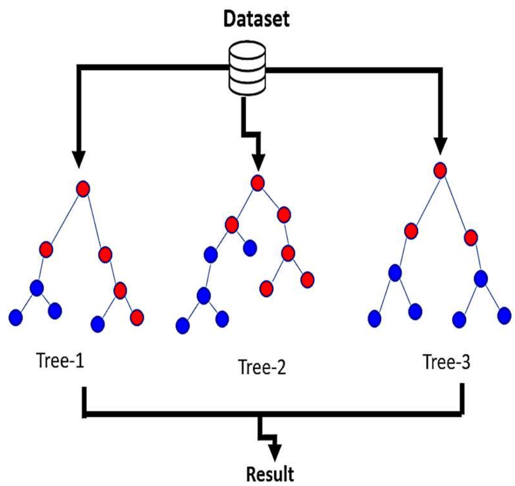


Figure 7: Random Forest classification

3. RELATED WORKS

Machine learning, otherwise termed artificial learning, is currently one of the most extensively practised fields. It has a profound impact on almost every aspect of life, and human beings derive considerable benefit from the advantages offered by this technique. Within the field of education, a significant amount of research has been dedicated to optimising the school learning process, as evidenced by the references cited in [19] and [28]. These references include a number of studies that have explored methods for guiding pupils using artificial learning approaches.In educational research, many researchers have implemented the integration of artificial learning in conjunction with other methodologies, with the aim of generating conclusions that are beneficial for both students and schools.

Articles [14], [15], [16], [18] and [13] describe the design of a framework for predicting academic orientation, based on artificial learning. The system developed is based on proven artificial intelligence algorithms, such as the Bayes network and the decision tree. The results obtained by the decision tree are particularly promising.

In his article, K. Kivuyirwa [14] presents a system for guiding students towards appropriate studies using the usual machine learning techniques (SVM, Bayes, RNN, decision tree) [20]. The data used is based on questionnaires given to students in order to derive a vector of characteristics for each one. The results obtained are satisfactory, particularly for the SVM technique.

In his article [16], F. Ouatik presents a student guidance system based on attendance and the number of absences during lessons. To model the data, the author uses the Big Data technique [18],[23]. The results obtained by neural networks, Bayesian networks and KNNs are then compared. This analysis leads to the conclusion that the best results are obtained by Naive Bayes.

In his article [17], F. Ouatik proposes an automatic student guidance system based on machine learning and the processing of large quantities of data [18]. This system uses three parameters relating to the student: the mark obtained in each subject, the number of absences in each subject and the general trend. The Naïve Bayes, SVM, Random Forest and RNN algorithms are then implemented in order to obtain results. A comparison of these algorithms led to the conclusion that Naïve Bayes gives good results.

Existing literature has extensively explored the use of artificial intelligence (AI) in educational guidance, but few studies have incorporated a dynamic MCQ test to personalize recommendations. Previous studies, such as those in [13], [14], [15], [16], [17], [18] and [20], have shown the benefits of applying machine learning algorithms in education, but have not used an interactive test to refine recommendations. This study aims to fill this gap by proposing a system based on MCQ answers, enabling academic orientations to be refined in a more personalized way.

4. THE WORK PROPOSES

4.1 Research methodology and contribution

Our study proposes an innovative academic guidance system. The system is based on the analysis of answers to a multiple-choice questionnaire (MCQ). The questionnaire covers six subjects and is integrated into an online learning platform. We use the scores obtained as predictive indicators. We combine machine learning classification algorithms with the principles of professional decision-making. Our aim is to optimize students' academic careers.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Our approach differs from traditional methodologies, which are based on stable criteria such as average academic grades, in that it is resolutely dynamic. This is based on interactive data from ongoing assessments, enabling the system's predictions to be constantly personalized and optimized. In addition, the integration of an example of online learning means that the analysis of learners' skills and profiles can be progressively refined, guaranteeing more personalized and progressive recommendations.

This study has two aims: to create a recommendation system to help students choose their educational path and to find models to suggest the best educational path for each student. To do this, an online test was set up on an e-learning platform to collect data on students. The second objective is to create and evaluate an educational path recommendation system using machine-learning algorithms. To do this, algorithms were used on the data collected on the platform. This makes it possible to create profiles of students and offer them personalized recommendations for their orientation.

The conceptual model of this study is predicated on the hypothesis that students' answers to a multiple-choice (MCQ) test can be used to predict their academic aptitudes and preferences. The machine-learning model was designed to classify students into different majors based on their performance on the test. The model under discussion is predicated on the hypothesis that academic aptitude is manifest in the results of the test, thus enabling the generation of recommendations that are tailored to the individual.

4.2 classification

In this work, we propose an innovative approach for classifying secondary schools into specialty groups using supervised machine learning. More specifically, we have exploited six algorithms - Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes and Random Forest - which have demonstrated their effectiveness in this field. To implement and evaluate these models, we used Python, as well as the Anaconda and Jupyter Notebook tools. Anaconda provides an integrated environment for scientific computing and data science, while Jupyter Notebook makes it easy to write, run and share code using a user-friendly interface. The dataset studied comes from a qualifying high school, where each student (Ei, (i=1,..., 6000) is represented by a feature vector (Vi) derived from their performance on an online test. Figure 9 provides a simplified illustration of the classification process and the work carried out.

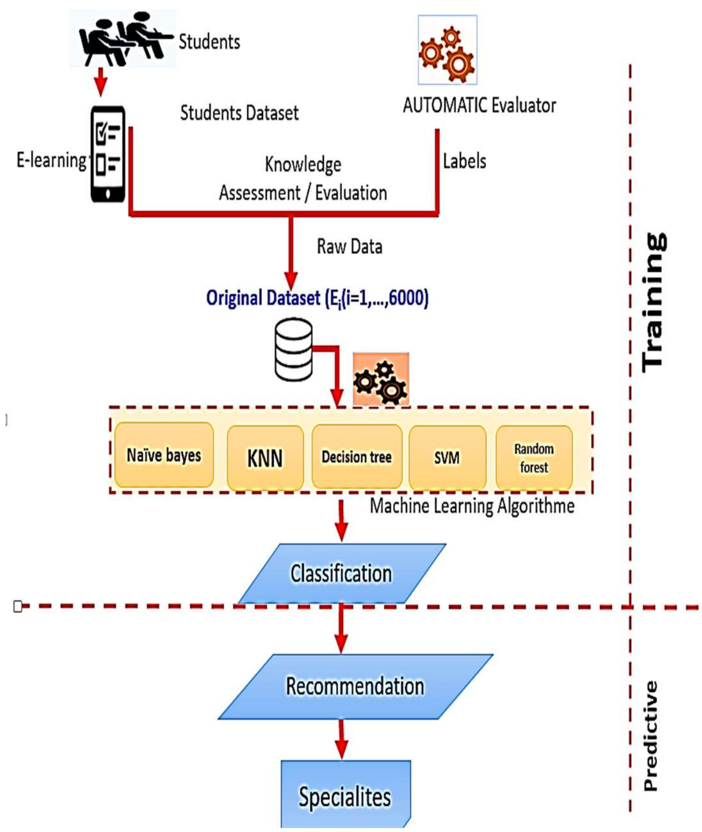


Figure 9. The proposed system

4.3 Data description :

The dataset used was collected from a real test administered to teachers in four secondary schools, using an online learning platform. The sample selected was that of the baccalaureate aimed at classifying students for university careers. A database of 6,000 baccalaureate holders was compiled, divided into the different streams shown in table 1 below:

Journal of Theoretical and Applied Information Technology
15th June 2025. Vol.103. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Specialities of the baccalaureate

	Baccalaureate specialties			
	SH	L	SVT	SP
Girls	515	420	1105	1005
Boys	485	580	895	995

SP: Physical Sciences SH: Human Sciences
SVT: Life and Earth Sciences L: Literary

The test proposed for students is given in the form of technical MCQ questions, with each specialty linked to a group of questions. Our test is made up of 60 questions, divided into 10 questions per specialty according to the following table:

Table 2: Distribution of questions in the test

Geology	Q5, Q10,Q16, Q19,Q29, Q36,Q42, Q45,Q52, Q49
Biology	Q2, Q11,Q18, Q21,Q26, Q33,Q39, Q48,Q55, Q58
Mathematics	Q1, Q7,Q14, Q22,Q30, Q32,Q37, Q43,Q56, Q57
Physics	Q3, Q8,Q17, Q23,Q27, Q31, Q41, Q47,Q54, Q59
Chemistry	Q4, Q12, Q13, Q20, Q28, Q34, Q40, 46, Q53, Q60
Economics	Q6, Q9,Q15, Q24,Q25, Q35,Q38, Q44,Q50, Q51

We used a database containing the answers of 6000 students (boys and girls) from various specialties in the baccalaureate. Each student was invited to take an online test consisting of technical questions on an e-learning platform. The results obtained were then analysed using Machine Learning algorithms to recommend the most appropriate university course for each student. Figures 10 and 11 (see appendix) show respectively the number of correct answers per materials and Correlation matrix between materials

5. RESULTS AND DISCUSSION

After collecting the results of the tests carried out on the e-learning platform for six subjects: Mathematics, Economics, Geology, Biology, Chemistry and Physics, an initial classification was made. The results obtained are presented in Table 4. Machine learning algorithms were used to generate personalized recommendations, either for a single subject or for a combination of subjects, according to the following categories:

- **A = {MIP}**: Mathematics, Computer Science, Physics
- **B = {Economy}**
- **C = {BCG} : Biology**, Chemistry, Geology
- **D = {A, B}**: Mathematics, Computer Science, Physics, Economics
- **E = {B, C}** : Economics, Biology, Chemistry, Geology
- **F = {A, C}**: Mathematics, Computer Science, Physics, Biology, Chemistry, Geology
- **G = {A, B, C}**: Mathematics, Computer Science, Physics, Economics, Biology, Chemistry, Geology
- **Non oriented**

Figure 11 shows the results of classifying the top five students according to their scores in different subjects (Mathematics, Physics, Biology, Chemistry, Geology, and Economics). To predict their specialty (A, B, C, E, G),

```
   Mathematics_Score  Physics_Score  Biology_Score  Chemistry_Score  \
0                  4              5              3                4
1                  5              7              8                8
2                  2              6              4                8
3                  4              6              4                3
4                  3              6              7                7

   Geology_Score  Economics_Score Specialty  \
0              3                6         B
1              4                3         C
2              3                9         E
3              5                3         C
4              2                8         G
```

Figure 11: Classification results for some students

The performance of the different Machine Learning models is summarized in Table 3 below and in Figure 12. They have been evaluated using four key metrics: accuracy, precision, recall and F1-score. These measures are used to determine the most effective model for our recommendation system.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: The performance of the different Machine Learning models

Algorithm	*Accuracy*	*Precision*	*Recall*	*F1-Score*
Logistic Regression	*0.670275*	*0.636421*	*0.670275*	*0.639315*
Decision Tree	*0.677769*	*0.664356*	*0.677769*	*0.667957*
Random Forest	*0.859284*	*0.856568*	*0.859284*	*0.855616*
SVM	*0.804330*	*0.788851*	*0.804330*	*0.791704*
KNN	*0.761032*	*0.756425*	*0.761032*	*0.756410*
Naive Bayes	*0.661948*	*0.633411*	*0.661948*	*0.618458*

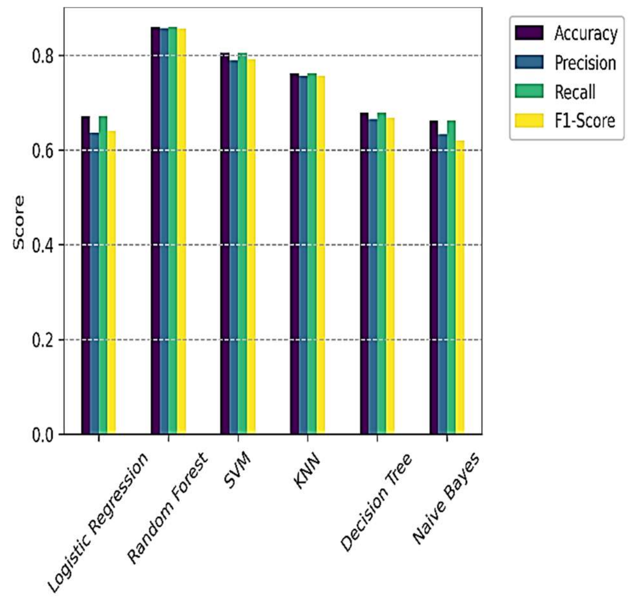
The results show that the Random Forest model performs best, followed by the SVM with a score of 80.43%. KNN scored 76.10%, showing that these two models performed acceptably well. On the other hand, the Decision Tree, Logistic Regression and Naïve Bayes algorithms performed the worst, with an accuracy of 67.03% or less.

After applying the Random Forest algorithm, Table 4 below shows the results of our guidance system, which recommends a specialization to baccalaureate students after assessing their skills via a test. The students are classified into different branches: A, B, C, D, E, F, G and “Not oriented”, corresponding to the various categories of specialization. Each column shows the number of students referred to each specialism. These recommendations are generated by analyzing the students’ performances in order to propose the most suitable orientation.

Table 4: Results of our proposed guidance system

Specialties	Number of students recommended
A	298
B	1102
C	2691
D	701
E	800
F	503
G	0
Non –oriented	197
Total	6000

Figure 12: Comparison of models



In percentage terms, Table 5. Below are the results.

Table 5. Results of our proposed guidance system in percentage

specialties	% Number of students recommended
A	4,97%
B	18,37%
C	44,85%
D	11,68%
E	13,33%
F	8,38%
G	0,00%
Non - oriented	3,28%
Total	100%

Figures 13 to 18 show a comparison of the classification results obtained with different machine learning models.

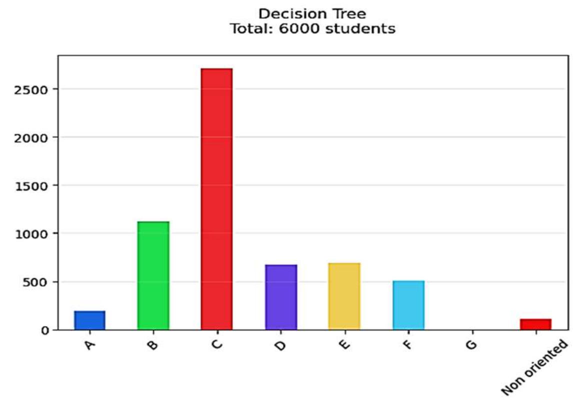


Figure13: Classification recommended by Decision Tree

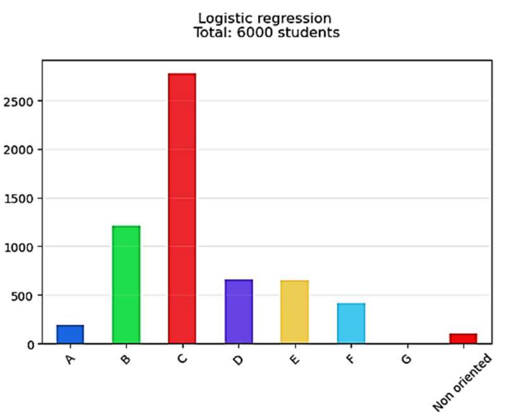


Figure14: Classification recommended by logistic Regression

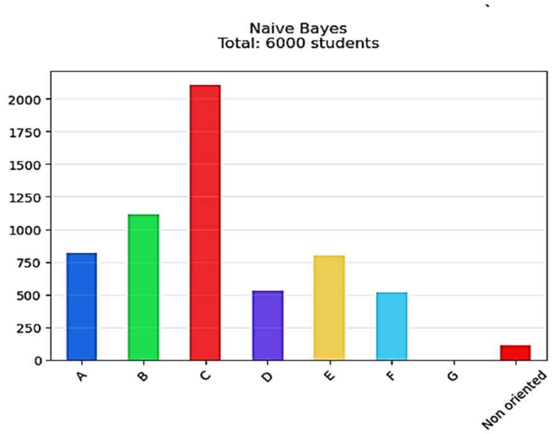


Figure15: Classification recommended by naive Bayes

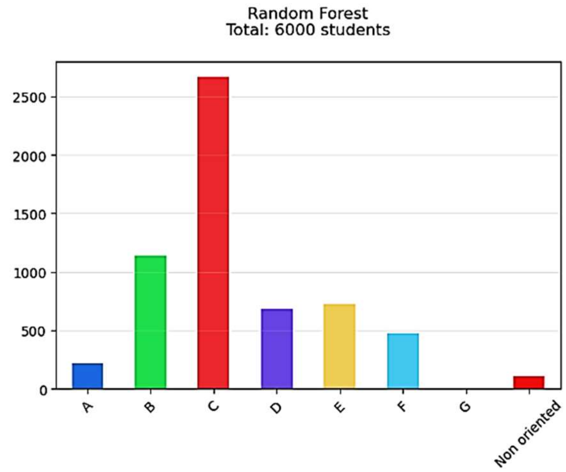


Figure16: Classification recommended by Rondom forest

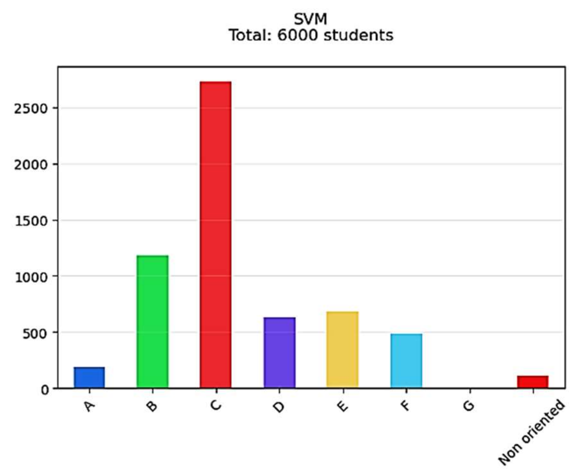


Figure17: Classification recommended by SVM

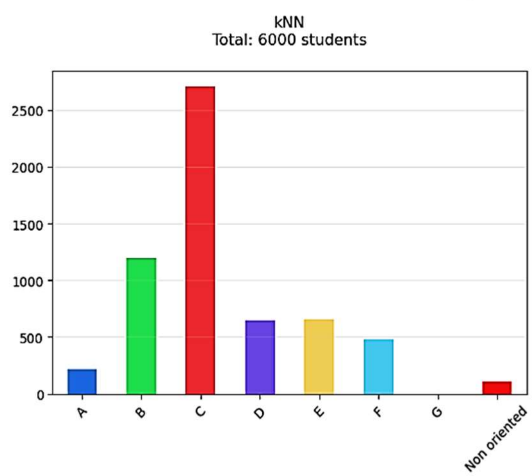


Figure18: Classification recommended by KNN

The graphs in the Figures 13 to 18 and table 3 show a dominance of class C, which accounts for about 42% of the predictions for each model, suggesting a bias in the data. Random Forest achieves 85.93% accuracy, followed by SVM (80.43%) and KNN (76.10%), while Decision Tree (67.78%), Logistic Regression (67.03%) and Naïve Bayes (66.19%) show more errors. Classes A (≈5%) and G (≈4%) are underrepresented, confirming an imbalance in the data. Less than 3% of the students are classified as 'non-orientated', confirming the reliability of the system. Naïve Bayes classifies about 10% of the students as A and B, but with an unbalanced classification. Its lowest F1 score of 0.6185 reflects its errors. Overall, the system gives promising results despite some biases.

The classification of students by machine learning algorithms gives very satisfactory results. The system set up in this work has produced promising results and confirms the effectiveness of setting up a guidance and classification system available to Moroccan school-leavers.

The following conclusions can be drawn from the results obtained:

Students in the physical sciences are directed straight into the MIP in the classic cases, while there are some who can give good results in other specialties (in economics, for example), other literary students also have a scientific bent, you just have to give them a chance.

The proposed guidance system, combined with machine learning models, means that certain pupils are not automatically directed towards traditional specialties, but can make choices that are better suited to their potential.

Adapting a traditional guidance system can considerably reduce the risk of disorientation at school by offering pupils more suitable pathways.

Our guidance system is powerful and makes an effective contribution to reducing the number of pupils dropping out of school, as effective guidance helps pupils to make better choices based on their aptitudes and interests, thereby reducing the risk of disengagement and failure at school.

The originality of this study lies in the integration of a MCQ test combined with machine learning algorithms to predict students' academic orientations. Unlike traditional systems, which are often based on fixed criteria such as grades, our approach proposes a personalized recommendation based on each student's specific results.

Although this recommendation system shows promising results, it is essential to stress that it has certain limitations. Firstly, the system relies heavily on multiple-choice test responses, which could make it difficult to take into account students' non-academic skills, such as creativity or social skills, in a comprehensive way. In addition, it should be noted that biases in the learning data may influence the recommendations, especially if certain specializations are under-represented.

6. CONCLUSION AND PERSPECTIVES

In this work, we have presented an overview of a guidance system for classifying and grouping high school students with a view to guiding them towards a university specialty appropriate to their tested skills. The system developed tends to integrated into Moroccan school guidance platforms in parallel with other guidance survey techniques. The results obtained are very satisfactory and the technique used is effective for this type of classification.

In pursuing this line of work on student guidance, we hope to continue research in this direction in the future. The results are as follows:

- ✓ Integration of a guidance system (E_Orientation) into the Massar platform (digital platform set up by the Moroccan Ministry of Education for school management and pupil monitoring in schools).
- ✓ Implementation of a guidance recommendation system based on Massar and machine learning, integrating the RIASEC test (Realistic, Investigative, Artistic, Social, Enterprising and Conventional) (developed by John Holland, is a guidance test that identifies the professional interests and personality types of individuals) in order to better guide educational and professional choices.
- ✓ Developing another guidance system based on other classification techniques using CNN 1D and LSTM deep learning algorithms.
- ✓ Expand the database used by adding other specialties and increasing the number of students tested

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES

[1] N. Dhanda, S. S. Datta, and M. Dhanda,"Machine Learning Algorithms," in Research Anthology on Machine Learning Techniques, Methods, and Applications, IGI Global, 2022, pp. 849-869.

[2] Z. Somogyi and Z. Somogyi, "Machine Learning Algorithms," Appl. Artif. Intell. Step-by-Step Guid. from Begin. to Expert, pp. 17-86, 2021.

[3] R. Raman, R. Shamim, S. V. Akram, L. Thakur, B. G. Pillai, and R. Ponnusamy, "Classification and Contrast of Supervised Machine Learning Algorithms," in 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), 2023, pp. 629-633.

[4] K. Tyagi, C. Rane, R. Sriram, and M. Manry, "Unsupervised learning," in Artificial Intelligence and Machine Learning for EDGE Computing, Elsevier, 2022, pp. 33-52. X. J. Zhu, "Semi-supervised learning literature survey," 2005.

[5] Erkan, Uğur. "A precise and stable machine learning algorithm: eigenvalue classification (EigenClass)." Neural Computing and Applications 33.10 (2021): 5381-5392.

[6] Memiş S, Enginoğlu S, Erkan U (2022b) A new classifcation method using soft decision-making based on an aggregation operator offuzzy parameterized fuzzy soft matrices. Turk J Electr Eng Comput Sci. https://doi.org/10.3906/elk-2106-28.

[7] Sedkaoui, Soraya, and Mounia Khelfaoui. Sharing economy and big data analytics. John Wiley & Sons, 2020.

[8] Kasperczuk, Anna, and Agnieszka Dardzinska. "Comprehensive review of classification algorithms for medical information system." Future Data and Security Engineering: 5th International Conference, FDSE 2018, Ho Chi Minh City, Vietnam, November 28–30, 2018, Proceedings 5. Springer International Publishing, 2018.

[9] Pandey S, Sharma V, Agrawal G (2019) Modifcation of KNN Algorithm. Int J Eng Comput Sci 8(11):24869-24877. https://doi.org/ 10.18535/ijecs/v8i11.4383

[10] Jiang, Shengyi, et al. "An improved K-nearest-neighbor algorithm for text categorization." Expert Systems with Applications 39.1 (2012): 1503-1509.

[11] Cristianini, Nello, and John Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods. Cambridge university press, 2000.

[12] Uddin, Shahadat, et al. "Comparing different supervised machine learning algorithms for disease prediction." BMC medical informatics and decision making 19.1 (2019): 1-16.

[13] Hicham El Mrabet. "A framework for predicting academic orientation using supervised machine learning", Journal of Ambient Intelligence and Humanized Computing (2023) 14:16539-16549.

[14] K. Kivuyirwa, "Prediction of student orientation in appropriate study streams using DataMining techniques", ISSR Journals.

[15] A. Tripathi, "naive Bayes classification model for the student performance prediction", IEEE 2019.

[16] F. Ouatik, "Student orientation using machine learning under Mapreduce with Hadoop", Journal of Ubiquitous and Prevasive Network, V13, 2020.

[17] F. Ouatik, "Students' orientation using machine learning and BigData", iJOE V17,2021.

[18] N. ELgendy, "Big Data Analytics: A literature review Paper", Springer 2014.

[19] A. Boucherou, "Usages de l'apprentissage artificial pour l'éducation", openEdition Journals, 2022 ;

[20] Sadhana Kodali, Madhavi Dabbiru, B Thirumala Rao, U Kartheek Chandra Patnaik," A k-NN-Based Approach Using MapReduce for Meta-path Classification in Heterogeneous Information Networks" 2019.

[21] Qasem A. Al-Radaideh (2011), "a classification model for predicting the suitable study track for school students", IJRRAS 8 (2) August 2011.

[22] H. Nawang (2021), "classification model and analysis on students' performance", J Fundam Appl Sci. 2017, 9(6S), 869-885.

[23] Morad Badrani(2024), Big data and ensemble learning for effective student guidance in Morocco, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 36, No. 3, December 2024

[24] Moulay A, Ghazi A (2021) Online Students' Classification Based on the Formal Concepts Analysis and Multiple Choice Questions. pp 119-129

[25] Sangodiah, R. Ahmad, and W. F. W. Ahmad, "A review in feature extraction approach in question classification using Support Vector Machine," in Proceedings - 4th IEEE International Conference on Control System,

Computing and Engineering, ICCSCE 2014, 2014

[26] Benarchid, Omar, and Naoufal Raissouni. "Support vector machines for object based building extraction in suburban area using very high resolution satellite images, a case study: Tetuan, Morocco." IAES International Journal of Artificial Intelligence 2.1 (2013).

[27] H. Al-Shehri et al, "Student performance prediction using Support Vector Machine and K-Nearest Neighbor," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, Canada, 2017, pp. 1-4, doi: 10.1109/CCECE.2017.7946847.

[28] Burman, I., & Som, S. (2019, February). Predicting students academic performance using support vector machine. In 2019 Amity international conference on artificial intelligence (AICAI) (pp. 756-759). IEEE

[29] S. N. Liao, D. Zingaro, K. Thai, C. Alvarado, W. G. Griswold, and L. Porter, "A robust machine learning technique to predict low-performing students," ACM Trans. Comput. Educ, vol. 19, no. 3, pp. 1-19, Sep. 2019, doi: 10.1145/3277569

[30] S. G. Essa, T. Celik, and N. E. Human-Hendricks, "Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review," IEEE Access, vol. 11, pp. 48392-48409, 2023, doi: 10.1109/ACCESS.2023.3276439.

[31] J. Figueroa-Cañas and T. Sancho-Vinuesa, "Early prediction of dropout and final exam performance in an online statistics course," IEEE Rev. Iberoam. Tecnol. Aprendiz. vol. 15, no. 2, pp. 86-94, May 2020, doi: 10.1109/RITA.2020.2987727. [10] A. Gupta, D. Garg, and P. Kumar, "

[32] N. Sghir, A. Adadi, and M. Lahmer, "Recent advances in predictive learning analytics: A decade systematic review (2012-2022)," Educ. Inf. Technol. vol. 28, no. 7, pp. 8299-8333, Jul. 2023, doi: 10.1007/s10639- 022-11536-0

[33] I. E. Livieris, T. Kotsilieris, V. Tampakas, and P. Pintelas, "Improving the evaluation process of students' performance utilizing decision support software, Neural Comput. Appl. vol. 31, no. 6, pp. 1683-1694, Jun. 2019, doi: 10.1007/s00521-018-3756-y.

[34] P. Jiao, F. Ouyang, Q. Zhang, and A. H. Alavi, "Artificial intelligenceenabled prediction model of student academic performancein online engineering education," Artif. Intell. Rev, vol. 55, no. 8, pp. 6321-6344, Dec. 2022, doi: 10.1007/s10462-022-10155-y. [74] L. Ramanathan, G. Parthasarathy, K. Vijayakumar, L. Laksh

[35] K. Kaur and K. Kaur, "Analyzing the effect of difficulty level of a course on students performance prediction using data mining," in Proc. 1st Int. Conf. Next Gener. Comput. Technol. (NGCT), Sep. 2015, pp. 756-761.

[36] Ho, T. K. (1998). The random subspace method for constructing decision forests. IEEE transactions on pattern analysis anmachine intelligence, 20(8), 832-844.

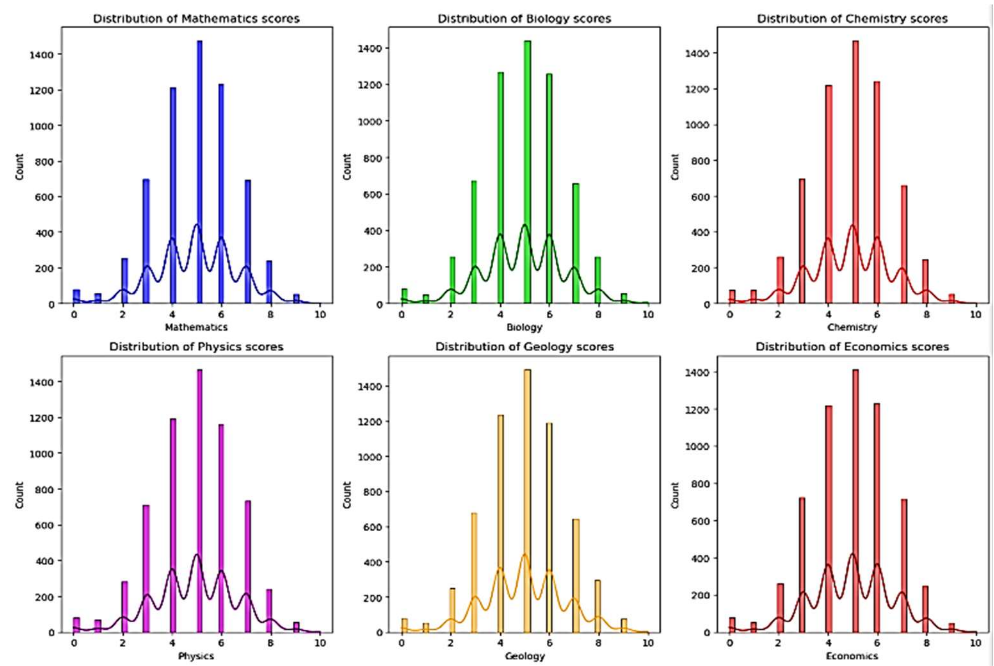
[37] Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. Expert systems with applications, 134, 93-101.

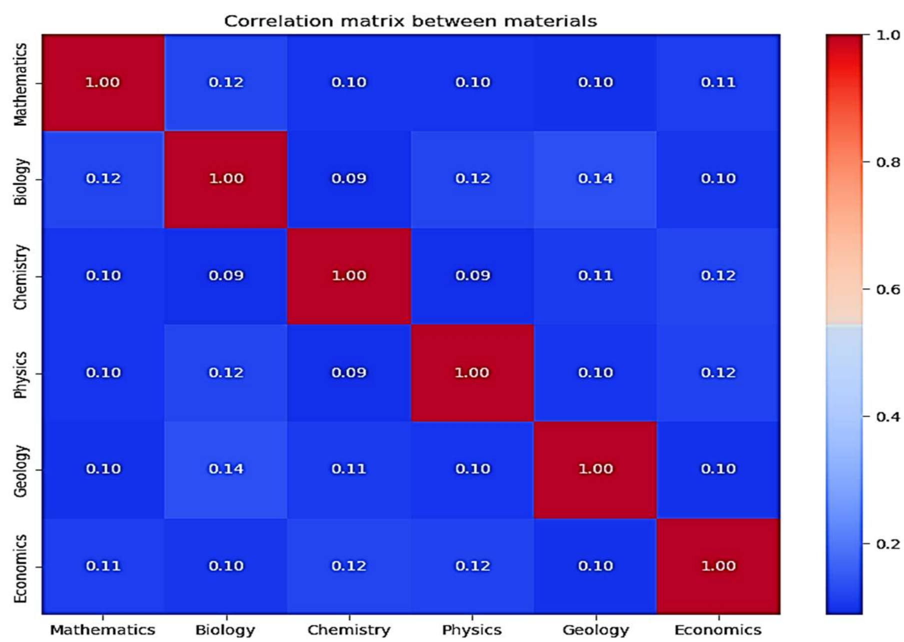
[38] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[39] Zahour, O., & El Habib Benlahmar, A. E. (2016, June). E-orientation: between prescribing theories and decision-making. In Conference TIM (Vol. 16).

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPENDIX:





Figures 11: The Correlation matrix between materials