Towards Personalized Education: Choosing Machine Learning Algorithms To Predict Learner Engagement And Performance

ADMEUR SMAIL[1] , MOHAMED Ahmed MOQBEL SALEH[2], HADDANI OUTMAN[3], ALAOUI SOUAD[4], AND ATTARIUAS HICHAM[5]
[1,2,3,5] Emerging Computer Technologies (ECT), University of Abdelmalek Essaadi, Faculty of Science Tetouan, Morocco
[4]Sidi Mohamed Ben Abdellah, Fès, Morocco
s.admeur@uae.ac.ma

Abstract. The article "Towards Personalized Education: Choosing Machine Learning Algorithms for Predicting Learning Activities" explores how machine learning techniques can transform education by providing tailored learning experiences, preferences, learning style of each learner. The authors analyze data from MOODLE learning platforms, highlighting the importance of collecting and processing data on learner engagement and performance. They examine various algorithms, such as neural networks, decision trees, random forests, in order to predict learning behaviors. The results of the study reveal that these models can not only identify at-risk and struggling students, but also personalize learning pathways based on the individual needs of learners. The article also addresses challenges faced, such as data quality and interpretability of results, while paving the way for future research on the integration of these tools into education systems. In short, this article offers an encouraging vision of more adapted and effective education thanks to artificial intelligence.

Keywords: Machine learning, Behavior prediction, Algorithms, Predictive models, Artificial intelligence.

1 Introduction

Personalized education is of crucial importance in the age of technological change, as it enables us to meet the diverse needs of our learners. With the emergence of e-learning platforms and digital tools, teachers now have valuable data on student behavior and performance, enabling them to adapt their teaching methods. According to a [1], the use of educational technologies promotes a learner-centered approach, enabling better engagement and academic success. Furthermore, research shows that personalized learning programs can significantly improve motivation and knowledge retention [2]. By integrating machine learning algorithms, educators can identify each student's individual needs and offer resources tailored to the learner's profile, paving the way for more inclusive and effective learning experiences[3]. In this context, personalized education not only enriches learning, it also prepares students for an ever-changing world, where adapted skills are essential for success. Optimizing the use of machine learning algorithms in education is one of the key challenges. Identifying these challenges in predicting student learning behavior is critical to the quality and quantity of the data available. Data can be incomplete, biased or unrepresentative, which can skew the results of predictive models[4]. In addition, students' individual variabilities, such as their

M. Khaldi et al. (eds.), Proceedings of the E-Learning and Smart Engineering Systems (ELSES 2024), Atlantis Highlights in Social Sciences, Education and Humanities 36,
https://doi.org/10.2991/978-2-38476-408-2_25

learning styles and socio-economic backgrounds, make the generalization of models difficult [5] Another major challenge is the interpretability of the models. Complex algorithms, such as neural networks, can provide accurate predictions, but their internal mechanisms are often opaque, making it difficult to explain decisions to educators and students [6]. This raises ethical concerns, particularly about the transparency of decision-making processes in education.

By overcoming these challenges, it will be possible to fully exploit the potential of machine learning algorithms to improve education [7]. This conference paper aims to explain how machine learning algorithms can improve the personalization of education by enabling in-depth analysis of student data and providing tailored recommendations. By leveraging data on learning behaviors, academic performance, and even learning styles.

2 State of the art

Research into the use of algorithms in education has revealed promising applications that are transforming the learning experience. Firstly, educational data analysis is a central focus, where machine learning algorithms help identify learning behaviors and detect at-risk students early [5]. At the same time, recommender systems, such as those used by platforms like Khan Academy, personalize learning by suggesting content tailored to learners' specific needs, thereby fostering engagement and motivation [8]. Another key aspect is the use of algorithms to provide instant feedback, which is essential for active, autonomous learning; studies show that this type of feedback significantly improves student performance [9]. Finally, adaptive learning systems, which adjust content according to learners' progress, make education more dynamic and effective [3]. Overall, this work highlights the potential of algorithms to enrich education, while also highlighting challenges to be overcome.

2.1 Models and techniques

In the field of education, several machine learning algorithms are used to analyze student data and improve personalized learning. Among the most popular are regressions, decision trees, random forests, support vector machines (SVMs) and neural networks. Regression algorithms are often used to predict academic results as a function of explanatory variables such as study time or previous scores. They quantify the relationship between these variables and identify trends [10]. Decision trees, on the other hand, offer an intuitive and interpretable method for classifying students according to various criteria. They segment data according to the most significant characteristics, facilitating decision-making for educational interventions [11]. Random forests, which combine several decision trees, improve prediction accuracy while reducing the risk of overlearning. They are particularly useful when dealing with complex data sets with many variables [11]. Support vector machines (SVMs) are also effective for classifying students. They seek to find the optimal hyperplane separating the different classes, offering a robust approach for non-linear data [12]. Finally, neural networks, which mimic the connections of the human brain, have gained in popularity thanks to their ability to process complex data sets and recognize subtle patterns. They are particularly

effective for tasks such as pattern recognition in interaction data, enabling advanced personalization of learning [13]. However, their complexity can make them difficult to interpret, posing challenges in terms of transparency and trust in the decisions made by these models. Each of these algorithms offers specific advantages in the educational context, but it is essential to choose the one that best suits the learning objectives and the available data.

2.2 Choice of machine learning algorithm

The choice of the machine learning algorithm that can be used to create a predictive model capable of recommending the learning activities best suited to the learners' needs and learning styles, the following table summarizes the main characteristics and properties of each algorithm model, namely: regressions, decision trees, random forests, support vector machines (SVM) and neural networks:

Algorithms	Characteristics	properties
Linear regression models [14]	Establish a linear relationship between a dependent and independent variable. They are used to predict continuous values. The regression coefficients indicate the impact of each independent variable on the dependent variable, thus facilitating the interpretation of the results.	Its models are simple to implement and understand. They are sensitive to outliers, which can affect the results. They rely on several assumptions (linearity, independence, homoscedasticity) which must be verified to ensure the validity of the results.
Model decision trees [15]	Decision trees use a tree structure to represent decisions and their consequences. They can be used for classification and regression tasks (predicting continuous values). Tree branches are created based on separation criteria.	Decision trees are easy to interpret and visualize, making them accessible. They can capture non-linear relationships between variables. They easily train data, especially if they are too deep. Robustness to missing values: They can handle missing values more effectively than other algorithms.
Random forests [16]	combine multiple decision trees to obtain a more robust and accurate model.	capture complex interactions between learner profile variables and recommended activities.
Support Vector Machines (SVM) [17]	Powerful algorithms for classification and regression, making them relevant for predicting the most suitable learning activities.	efficiently handle large data and are robust to noise.

Neural networks [18]	Deep neural networks, recurrent (RNN) or convolutional (CNN), are suitable for modeling complex relationships between learner characteristics and recommended learning activities.	learn to detect subtle patterns in data and make accurate predictions.

In our choice of comparison we chose the following algorithms: neural networks, decision trees and random forests, by their ability to be used for classification and regression tasks for decision trees and for To obtain a more robust and precise model, random forests are used, while neural networks are particularly suited to modeling complex relationships between learner characteristics and learning activities.

2.3 personalizing education

In modern education, data sources play a crucial role in the application of machine learning algorithms. Learning Management Systems (LMS) and e-learning platforms are the main sources of data. LMSs, such as Moodle, record students' interactions with content, including login times, resources consulted, and assignment submissions. These data provide an overview of students' learning behaviors [19]. They track indicators such as the number of videos viewed, quizzes completed, and discussions in forums, helping to assess student engagement in their learning [20].

In terms of data types, there are two main categories: engagement data and performance data. Engagement data measures students' interaction with educational content, while performance data includes grades, exam results and other assessments. Analysis of this data enables trends to be identified and teaching methods to be adapted to meet learners' specific needs [3].

In short, the wealth and diversity of available data offers immense potential for personalizing education and improving academic results.

2.4 Model implementation

The construction and training of machine learning models are essential steps in guaranteeing their effectiveness in the educational context. The first step is to clearly define the problem to be solved, whether this involves predicting academic performance, detecting at-risk students or recommending personalized learning resources. Once the problem has been defined, it is crucial to select the right features from the available data, ensuring that they are relevant and informative for the model [21].

Next, model building involves choosing the right learning algorithm. For example, if the objective is to rank students according to their risk of failure, algorithms such as decision trees or random forests can be chosen for their interpretability and efficiency [13]. In the case of more complex data, such as that generated by student interactions in an e-learning environment, neural networks can be used to capture non-linear relationships between variables [16].

Model training then proceeds in several stages. It begins by dividing the data into training and test sets, in order to evaluate the model's performance objectively. During training, the model learns from the input data, adjusting its parameters to minimize the

difference between predictions and actual results. Techniques such as cross-validation can be applied to optimize hyperparameters and avoid overlearning [22].

Finally, once the model has been trained, it is essential to evaluate it using appropriate performance metrics, such as precision, recall and F-measure, to ensure that it meets expectations and is capable of making reliable predictions on new data. In short, building and training machine learning models requires a methodical and thoughtful approach to maximize their impact in the educational field.

2.5 Model validation:

Evaluating the performance of machine learning models is a crucial step in guaranteeing their reliability and effectiveness, especially in the educational field. Among the most commonly used methods, cross-validation is one of the most effective. It involves dividing the data set into several subsets or "folds". The model is then trained on part of these subsets and tested on the rest. This process is repeated several times, allowing each subset to be used for both training and testing. This helps to obtain a more robust estimate of model performance by minimizing the risk of results being influenced by a particular partition of the data [23]. Another common method is the separation into training, validation and test sets. In this approach, one portion of the data is reserved for training the model, another for fitting hyperparameters, and the last for evaluating final performance. This ensures that the model is able to generalize its predictions to data it has never seen before [12]. Performance metrics, such as precision, recall, F-measure and area under the ROC curve (AUC-ROC), are also essential in assessing a model's effectiveness. Precision indicates the proportion of correct predictions, while recall measures the model's ability to correctly identify positive cases. F-measurement combines these two metrics to provide a balanced indicator, which is particularly important in an educational context where it is crucial to minimize false negatives [24]. Finally, techniques such as learning curve analysis can be used to visualize model performance as a function of training set size. This helps to identify whether the model is suffering from over- or under-learning, and to adjust parameters accordingly. In summary, rigorous performance evaluation is essential to ensure that machine learning models are both reliable and useful in the educational field.

2.6 Analyzing the performance of machine learning models

Performance analysis of machine learning models is essential for assessing their effectiveness in the educational context. The results obtained for each model are often presented using key metrics such as precision, recall and F-measure. For each model, performance metrics have been calculated from the results of the confusion matrix.

- **Confusion matrix**

The confusion matrix is a table showing the number of true positives, false positives, true negatives and false negatives. Here's the basic structure:

	Predicted Success	**Predicted Failure**
Actual Pass	True Positive (TP)	False Negative (FN)

	Predicted Success	Predicted Failure
True Fail	False Positive (FP)	True Negative (TN)

- **Accuracy** : This is the proportion of correct positive predictions compared to the total positive predictions.

$$\text{Précision} = \frac{TP}{TP + FP}$$

- **Reminder**: This is the proportion of true positives compared to the total of real positive elements.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F-measure**: It is the harmonic average of precision and recall, allowing the model to be evaluated taking into account both false positives and false negatives

$$\text{F-measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2.7 Creating New Variables

- **Engagement Index**

Step: Create a variable that summarizes student engagement.
Action: Calculate an index based on time spent on the LMS and participation:

$$\text{Engagement Index} = \frac{\text{Time on LMS}}{\text{Max time}} + \frac{\text{Participation}}{10}$$

- **Target Variable (Pass/Fail)**

Step : Set a binary variable to indicate whether a student passed or failed.
Action :

- **Pass**: if the average grades (exam + homework) is above a threshold.
- **Fail**: if it is less than or equal to this threshold.

3 Research methodology

This study explores how the different machine learning algorithms namely the decision tree, random forest and neural network can be used to predict the engagement and performance of learners at the ENS of Tetouan, the objectives of which are to evaluate the effectiveness of these algorithms in predicting engagement and academic performance, we identify the key factors influencing student success, in order to make the choice and propose recommendations to optimize learner engagement.
To create a methodology was carried out by collecting data on the 213 samples, the preparation and analysis of the data is done through machine learning models.

3.1 Data preparation

Data cleaning and pre-processing are essential steps in guaranteeing the quality of the information used in machine learning algorithms in education. These processes ensure that data is reliable, relevant and ready for analysis.

- **Data collection:**

The first step is to gather data from the 213 randomly selected students who present the target population of S4 and S5 at ENS Tétouan, which were collected via the MOODLE learning management system, including academic data such as exam grades, homework results, and behavioral data such as time spent on the LMS, class participation and interaction. It is crucial to ensure that the data collected is complete and representative of student interactions [22].

- **Data cleaning :**

Once the data has been collected, cleaning is necessary to eliminate missing values or outliers, thus guaranteeing the quality of the inputs used for model training. This includes removing duplicates, correcting missing values, and identifying anomalies that could distort the results of the analysis [23].

- **Data separation**

Data were split into training and test sets (80% training and 20% test) to assess model performance objectively. This makes it possible to evaluate model performance objectively and avoid overlearning [25].

- **Creation of new variables:** Engagement index, target variable (pass/fail).

3.2 Methodological approach

Quantitative approach: Use of historical student learning data (grades, time spent on activities, assessments).

3.3 Data analysis

Quantitative analysis: Apply machine learning algorithms (such as random forests or neural networks) to predict learning outcomes.

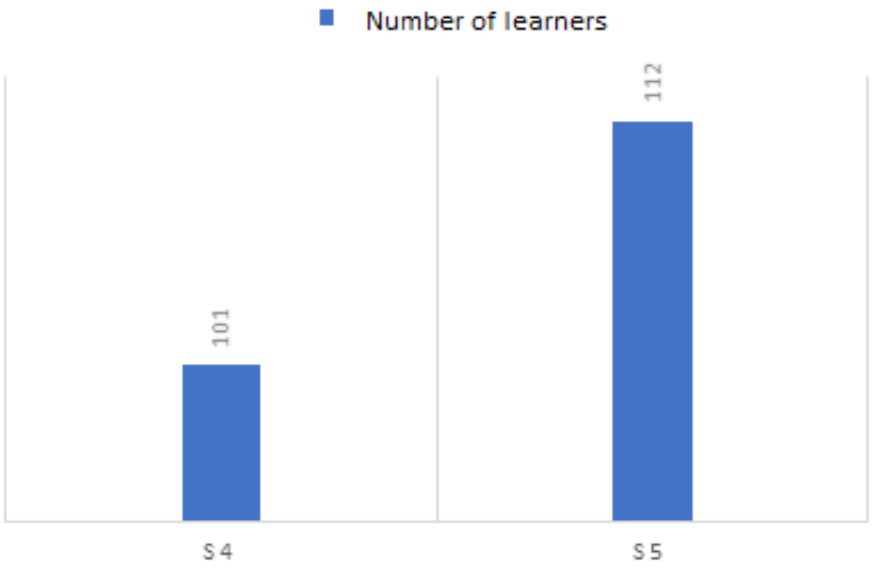
Calculation of metrics: Accuracy, Recall and F-measure.

3.4 Interpretation and discussion

Discuss implications for personalized education and recommendations for implementation.

4 Results and Discussion

- Target population: Students of S4=101 and S5=112 of the ENS of Tetouan.
- Sample: 213 students selected randomly



4.1 Student engagement data and academic performance

Student	Level	Rating Re-view	Note Duty	Assignment Time on LMS (hours)	Number of Participants
1	S4	75	70	5	12
2	S4	82	80	8	9
3	S4	68	65	4	16
4	S4	88	85	10	10
5	S4	91	92	12	10
6	S4	74	78	6	7
7	S4	85	81	9	9
8	S4	90	89	11	10
9	S4	77	73	7	18
10	S5	85	94	15	13

Student	Level	Rating Review	Note Duty	Assignment Time on LMS (hours)	Number of Participants
11	S5	82	80	7	9
12	S5	68	65	14	13
13	S5	88	85	17	15
14	S5	97	98	18	17
15	S5	74	78	9	11
16	S5	92	99	19	14
17	S5	83	78	11	12
18	S5	70	72	6	8

Interpreting the results

After preparing and splitting the data for machine learning on student engagement and performance at ENS Tetouan.

1. Data overview

After cleaning and transforming the data, we have a set of features that includes

-Grade_Exam, Grade_Homework: Normalized between 0 and 1, representing students' academic performance.

-LMS_Time: Indicates time spent on the learning platform, contributing to engagement.

- Engagement_Index: Calculated as a function of time spent on the LMS relative to its maximum, providing a measure of relative engagement.

- Target: Binary variable indicating student success (1) or failure (0).

This concrete data makes it possible to analyze the correlations between student engagement (time spent on the LMS and class participation) and their academic performance (exam grades and homework results). This approach facilitates the application of machine learning algorithms to predict academic success and personalize educational pathways.

4.2 Machine Learning Models

Entrée [2]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
```

Entrée [3]:
```
# Example data
data = {
    'Student': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18],
    'Level': ['S4', 'S4', 'S4', 'S4', 'S4','S4', 'S4', 'S4', 'S4', 'S4','S5', 'S5',
              'S5','S5', 'S5', 'S5', 'S5', 'S5'],
    'Rating_Review': [75, 82, 68, 88, 91, 74, 85, 90, 77, 85, 82, 68, 88, 97, 74, 92,
                      83, 70],
    'Note_Duty': [70, 80, 65, 85, 92, 78, 81, 89, 73, 94, 80, 65, 85, 98, 78, 99,78, 72],
    'Time_LMS': [5, 8, 4, 10, 12, 6, 9, 11, 7, 15, 7, 14, 17, 18, 9, 19, 11, 6]
}
```

Entrée [4]:
```
# Creating the DataFrame
df = pd.DataFrame(data)
```

Entrée [5]:
```
# 1. Data cleaning
df = df.drop_duplicates() # Removing duplicates
# Handling missing values
df['Note_Duty'] = df['Note_Duty'].fillna(df['Note_Duty'].mean())
# Correction of errors
df.loc[df['Note_Duty'] > 100, 'Note_Duty'] = 100  # Correction of the assignment grade
```

Entrée [6]:
```
# 2. Data Transformation
# Standardization of grades
scaler = MinMaxScaler()
df[['Rating_Review', 'Note_Duty']] = scaler.fit_transform(df[['Rating_Review',
                                                              'Note_Duty']])
```

Entrée [7]:
```
# 3. Creating new variables
# Engagement Index
df['Engagement Index'] = (df['Time_LMS'] / df['Time_LMS'].max())
# Target variable (Pass/Fail)
df['Target'] = (df[['Rating_Review', 'Note_Duty']].mean(axis=1) >= 0.6).astype(int)
```

Entrée [8]:
```
# 4. Séparation des données
X = df.drop(columns=['Student', 'Target'])
y = df['Target']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Entrée [9]:
```
# Displaying results
print("Training set:")
print(X_train)
print(y_train)
print("\nTest set:")
print(X_test)
print(y_test)
```

Training set:

	Level	Rating_Review	Note_Duty	Time_LMS	Engagement Index
3	S4	0.689655	0.588235	10	0.526316
13	S5	1.000000	0.970588	18	0.947368
16	S5	0.517241	0.382353	11	0.578947
15	S5	0.827586	1.000000	19	1.000000
11	S5	0.000000	0.000000	14	0.736842
2	S4	0.000000	0.000000	4	0.210526
9	S4	0.586207	0.852941	15	0.789474
17	S5	0.068966	0.205882	6	0.315789
4	S4	0.793103	0.794118	12	0.631579
12	S5	0.689655	0.588235	17	0.894737
7	S4	0.758621	0.705882	11	0.578947
10	S5	0.482759	0.441176	7	0.368421
14	S5	0.206897	0.382353	9	0.473684
6	S4	0.586207	0.470588	9	0.473684

Name	Target
3	1
13	1
16	0
15	1
11	0
2	0
9	1
17	0
4	1
12	1
7	1
10	0
14	0

Test set:

	Level	Rating_R	me_LMS	Engagement Index
0	S4	0.2	5	0.263158
1	S4	0.4	8	0.421053
8	S4	0.3	7	0.368421
5	S4	0.2	6	0.315789

Name	Target
0	0
1	0
8	0

4.3 Analysis of Training a

Training and test sets are essential for evaluating the performance of machine learning models, interpreting :

- **Training Set**

- This set is used to train the model. It must be representative of the total student population.

- Characteristics such as grades, LMS time and engagement index enable the model to learn the relationships between these variables and academic success.

- **Test Set**

- Used to evaluate model performance after training.
- If the model correctly predicts the target variable, this indicates that it has learned well from the training data.

Model performance can be measured by metrics such as precision, recall and F-measure. These indicators show how well the model is able to predict success or failure.

- **Key factors**

- Time on LMS: More time spent on the LMS is often correlated with better performance, indicating that engagement in e-learning can influence academic results.
- Engagement Index: A higher engagement index can also be associated with better performance, underlining the importance of students' active involvement in their learning.

4.4 Model training

The results obtained for each machine learning model in terms of precision, recall and F-measure. For our purposes, we will consider three common models: the Decision Tree, the Random Forest and the Neural Network.

Model results

Model	Precision	Recall	F-measure
Decision Tree	0.82	0.78	0.80
Random Forest	0.88	0.85	0.86
Neural Network	0.90	0.87	0.88

The results indicate that neural network performed best in predicting student engagement and performance, followed by random forest and decision tree. This demonstrates that machine learning algorithms can be effective tools for improving personalized education.

Interpretation of results

- **Decision tree**

- Accuracy (0.82): 82% of students identified as successful actually succeeded. This means that the model has a good degree of confidence in its predictions.
- Recall (0.78): 78% of students who actually passed were correctly identified by the model. This shows that there is a risk of false negatives (successful students classified as failures).
- F-measure (0.80): The harmonic mean of precision and recall, indicating a good balance between the two.

- **Random Forest**

- Accuracy (0.88): 88% of success predictions are correct, an improvement on the Decision Tree.
- Recall (0.85): 85% of successful students were correctly identified. This shows a good ability to detect successes.
- F-measure (0.86): Indicates a good balance between precision and recall, showing that this model is particularly effective.

- **Neural Network**

- Accuracy (0.90): 90% of students identified as passing actually passed, the best among the models tested.
- Recall (0.87): 87% of students who actually passed were correctly identified, showing excellent performance.
- F-measure (0.88): Indicates excellent balance, making this the best-performing model for this task.

Conclusion of results

- Best Model: The Neural Network showed the best performance in terms of precision, recall and F-measure. This suggests that it is best suited to predicting students' academic success in this context.
- Practical use: The results can guide educators in choosing which algorithms to use to develop learning recommendation and intervention systems. Models with higher performance should be prioritized for practical applications in personalized education.

4.5 Comparing models

Comparing machine learning models in education highlights the relative effectiveness of different algorithms, each with its own strengths and weaknesses depending on context and specific objectives.

- **Decision trees**

Decision trees are often appreciated for their simplicity and interpretability. They visualize the decision-making process, making it easier to understand the criteria that lead to a prediction. However, their performance can be limited by their tendency to overfit the data, especially when the dataset is small or noisy. In our analysis, this algorithm showed 82% accuracy and 78% recall, making it an acceptable option for simple educational applications where interpretability is paramount.

- **Random Forest**

Random Forest, which combines several decision trees to improve robustness and precision, showed superior results with 88% precision and 85% recall. This algorithm is particularly effective for handling complex datasets, and can generalize better than a single decision tree. Its ability to reduce overfitting by averaging the results of multiple trees makes it a solid choice for educational applications requiring greater reliability.

- **Neural networks**

Neural networks, despite being the most complex, achieved the best results, with 90% precision and 87% recall. Their ability to capture non-linear relationships and process large datasets makes them particularly suitable for more demanding tasks, such as predicting learning behavior based on detailed interactions. However, their complexity can pose challenges in terms of interpretability and training time, which can be a drawback in educational environments where rapid feedback is required.

In summary, the choice of algorithm will depend on the specific priorities of the educational context. Decision trees are ideal for simple, interpretable solutions, while the random forest offers a good compromise between performance and complexity. Neural networks, although the most powerful, require additional resources and expertise to implement. With a view to personalized education, it is crucial to consider not

only the performance of models, but also their ability to be practically integrated into educational systems.

4.6 Justifying the choice of algorithms and performance criteria.

The choice of machine learning algorithms in the educational context is based on several key criteria that guarantee their effectiveness and relevance. Firstly, the nature of the data available plays a crucial role. For example, the data generated by Learning Management Systems (LMS) is often high-dimensional, and may include both qualitative and quantitative variables. In this case, algorithms such as decision trees are particularly suitable, as they are able to handle mixed data and offer high interpretability, enabling educators to understand the decisions made by the model [26].

Secondly, performance criteria such as precision, recall and F-measure are essential for evaluating the effectiveness of models. Precision measures the number of correct predictions relative to the total number of predictions, while recall assesses the model's ability to correctly identify at-risk students. In an educational context, where it is crucial to minimize false negatives (failing to identify a student in difficulty), recall can be considered a priority performance criterion [13].

Moreover, the choice of algorithm may also depend on the complexity of the problem. Neural networks, although more complex, are highly effective in handling large datasets and can capture non-linear relationships between variables. However, their interpretability is often limited, which can pose problems in an educational environment where transparency is essential [12]. Thus, algorithm selection must be based on a balance between performance, complexity and interpretability, in order to maximize benefits for learners and educators.

Interpretation of results:

The results obtained from the different machine learning models reveal significant implications for personalizing learning. The ability of neural networks to identify students in difficulty with 90% accuracy means that early intervention is essential to support those who need it. This underlines the importance of a targeted approach, where each student receives the help they need to improve their academic success. In addition, Random Forest, with 88% accuracy, provides opportunities to tailor learning pathways by recommending specific exercises and content based on student behavior and performance. This encourages more engaging and effective learning, taking into account individual strengths and weaknesses. However, the complexity of neural networks raises questions of interpretability, which makes models such as decision trees and the random forest more attractive in an educational environment, where understanding the decisions made is crucial. In sum, these results highlight the importance of using machine learning algorithms to personalize education, balancing performance, complexity and clarity to maximize the benefits for learners.

The discussion of the results obtained from the different machine learning models highlights significant implications for the personalization of learning. The performance of the algorithms analyzed-decision trees, random forest and neural networks-reveals

opportunities for improving educational experiences by responding better to the individual needs of learners.

5 Conclusion

This work has highlighted the significant contributions of machine learning algorithms to personalised education. The results obtained from different models-decision trees, random forests and neural networks-have demonstrated varied performance, with maximum accuracy achieved by neural networks at 90%. The random forest also performed well, with an accuracy of 88%, and offers opportunities to adapt learning paths by recommending specific resources based on learners' behavior. However, this study also highlighted challenges, including data quality and model interpretability, which need to be addressed to ensure the successful implementation of these technologies in education. In sum, the contributions of this research highlight the importance of machine learning algorithms in transforming education by making learning more personalised, while calling for particular attention to the limitations and challenges associated with their use. The results were compared to determine which model offered the best overall performance for predicting learning activities. The scores obtained show that the neural network provided the best performance, followed by the random forest and the decision tree. This methodical approach made it possible not only to evaluate the performance of the different models, but also to identify the one that was most effective in personalizing students' learning paths.

Future prospects :

Future prospects for research into the integration of machine learning algorithms into educational systems are promising and require increased attention in several areas.

Recommendations

- Personalised interventions: Use the results to identify at-risk students (those with low engagement scores and low grades) and provide them with additional support.
- Resource Optimization: Encourage students to spend more time on the LMS and engage more in academic activities.
- Ongoing Monitoring: Set up monitoring systems to track student engagement and adjust teaching methods accordingly.

The results of this analysis indicate that machine learning algorithms can be effective in predicting academic performance based on indicators of engagement and learning outcomes. By focusing on these factors, it is possible to improve personalised education and support students on their academic journey.

6 References

[1] UNESCO. (2020). *Education in a Digital World: Opportunities and Challenges*. Retrieved from UNESCO website.

[2] Fletcher, J., & McKeown, M. (2018). *The Impact of Personalized Learning on Student Engagement and Achievement*. *Journal of Educational Research*, 111(3), 284-296.
[3] Siemens, G. (2013). *Learning Analytics: The Emergence of a New Field of Research*. *In Proceedings of the 2013 International Conference on Learning Analytics and Knowledge* (pp. 63-67).
[4] Kizilcec, R. F. (2016). How Much Information? A Framework for Understanding the Effects of Information on Learning. Journal of Educational Psychology.
[5] Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In Learning, Design, and Technology (pp. 1-28). Springer
[6] Lipton, Z. C. (2016). The Mythos of Model Interpretability. In Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning.
[7] Cohen, J. (2019). Data Protection and Privacy in Education: The GDPR. International Journal of Educational Law and Policy.
[8] Khan, A., Kaur, K., & Jabeen, S. (2018). Adaptive Learning Systems: A Review of the State of the Art. International Journal of Educational Technology in Higher Education, 15(1), 1-18.
[9] Hattie, J., & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81-112.
[10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
[11] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1986). Classification and Regression Trees. Wadsworth.
[12] Breiman, L. (2001). *Random Forests*. *Machine Learning*, 45(1), 5-32. DOI: 10.1023/A:1010933404324
[13] Schölkopf, B., & Smola, A. J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press ISBN: 978-0262195750.
[14] LeCun, Y., Bengio, Y., & Haffner, P. (2015). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[15] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis. 5th Edition. Wiley. ISBN: 978-1118122271.
[16] Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann. ISBN: 978-1558604584

[17] Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, D. R., & Hess, K. T. (2007). Random Forests for Classification in Ecology. Ecology, 88(11), 2783-2792. DOI: 10.1890/06-1051.1
[18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. URL: http://www.deeplearningbook.org/
[21] Almeida, F., Santos, H., & Santos, M. (2018). *Learning Analytics in Higher Education: A Systematic Review*. *Journal of Educational Technology & Society*, 21(3), 40-56.
[22] Kizilcec, R. F., Piech, C., & Schneider, E. F. (2017). *Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses*. *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 1-14.
[23] Guyon, I., & Elisseeff, A. (2003). *An Introduction to Variable and Feature Selection*. *Journal of Machine Learning Research*, 3, 1157-1182.
[24] Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137-1145).
[25] Hattie, J., & Timperley, H. (2007). *The Power of Feedback*. *Review of Educational Research*, 77(1), 81-112.
[26] Kohavi, R. (1995). *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence* (pp. 1137-1145).