ENHANCING ADAPTIVE LEARNING SYSTEMS WITH ADVANCED PERFORMANCE METRICS

MELHORAR OS SISTEMAS DE APRENDIZAGEM ADAPTATIVA COM MÉTRICAS DE DESEMPENHO AVANÇADAS

Ikram Amzil
ORCID 0009-0003-8167-9384
Information Technology and Systems Modeling Laboratory (TIMS)
FS, Abdelmalek Essaadi University
Tetouan, Morocco
ikram.amzil@etu.uae.ac.ma

Souhaib Aammou
ORCID 0000-0002-6224-0929
Information Technology and Systems Modeling Laboratory (TIMS)
FS, Abdelmalek Essaadi University
Tetouan, Morocco
s.aammou@uae.ac.ma

Youssef Jdidou
ORCID 0000-0002-7391-0940
Laboratory of Intelligent Systems and Applications (LSIA), Ecole Marocaine des sciences de l'Ingénieur.
Tangier, Morocco
y.jdidou@emsi.ma

Abstract: Adaptive learning systems are integral to contemporary educational technology, offering tailored educational content to meet individual student needs. The effectiveness of these systems significantly depends on accurately assessing learner performance and adaptability. This research is centered on implementing and evaluating sophisticated performance metrics for multi-class classification in adaptive learning systems to enhance their functionality in educational settings. The study aims to explore and validate various performance metrics that can critically enhance the functionality of adaptive learning systems. By integrating advanced multi-class classification techniques, it seeks to provide a nuanced understanding of learner interactions and outcomes, facilitating more personalized and effective learning experiences. The methodological approach of this study involves constructing theoretical models tailored to educational data, utilizing advanced statistical tools such as Cohen's Kappa, accuracy, precision, recall, and F1-Score to measure model performance, implementing these models in simulated environments to gather data on learning outcomes, and applying cross-validation techniques to ensure reliability and generalizability across different educational datasets. Initial findings suggest that the integration of refined performance metrics significantly improves the prediction accuracy and adaptability of learning systems. Employing a stratified k-fold cross-validation method has shown potential in enhancing the system's ability to dynamically tailor content based on learner performance. The efficacy of metrics like the F1-Score and Cohen's Kappa is highlighted, particularly in dealing with the imbalanced class distributions typical of personalized learning paths. The study highlights the importance of selecting suitable performance metrics in designing and enhancing adaptive learning systems. It discusses how these metrics affect the decision-making processes of adaptive algorithms and their implications for educational pedagogy. It also examines the scalability of the methods proposed and their real-world applicability. This research contributes to the field of educational technology by showing how advanced performance metrics can enhance the efficacy and personalization of adaptive learning systems. It opens pathways for creating more responsive educational environments that effectively meet diverse learner needs.

Keywords: Adaptive Learning Systems; Performance Metrics; Educational Technology; Learner Adaptability

Resumo: Os sistemas de aprendizagem adaptativa fazem parte integrante da tecnologia educativa contemporânea, oferecendo conteúdos educativos adaptados às necessidades individuais dos alunos. A eficácia destes sistemas depende significativamente da avaliação exacta do desempenho e da adaptabilidade do aluno. Esta investigação centra-se na implementação e avaliação de métricas de desempenho sofisticadas para a classificação multi-classe em sistemas de aprendizagem adaptativa, a fim de melhorar a sua funcionalidade em contextos educativos. O estudo tem como objetivo explorar e validar várias métricas de desempenho que podem melhorar de forma crítica a funcionalidade dos sistemas de aprendizagem adaptativa. Ao integrar técnicas avançadas de classificação multi-classe, procura fornecer uma compreensão diferenciada das interações e resultados do aluno, facilitando experiências de aprendizagem mais personalizadas e eficazes. A abordagem metodológica deste estudo envolve a construção de modelos teóricos adaptados aos dados educativos, a utilização de ferramentas estatísticas avançadas como o Kappa de Cohen, a exatidão, a precisão, a recordação e o F1-Score para medir o desempenho do modelo, a implementação destes modelos em ambientes simulados para recolher dados sobre os resultados da aprendizagem e a aplicação de técnicas de validação cruzada para garantir a fiabilidade e a generalização em diferentes conjuntos de dados educativos. Os resultados iniciais sugerem que a integração de métricas de desempenho refinadas melhora significativamente a precisão da previsão e a adaptabilidade dos sistemas de aprendizagem. A utilização de um método de validação cruzada estratificada k-fold demonstrou potencial para melhorar a capacidade do sistema de adaptar dinamicamente o conteúdo com base no desempenho do aluno. A eficácia de métricas como o F1-Score e o Cohen's Kappa é destacada, particularmente ao lidar com as distribuições de classes desequilibradas típicas dos percursos de aprendizagem personalizados. O estudo realça a importância de selecionar métricas de desempenho adequadas para conceber e melhorar os sistemas de aprendizagem adaptativa. Discute a forma como estas métricas afectam os processos de tomada de decisão dos algoritmos adaptativos e as suas implicações para a pedagogia educativa. Examina também a escalabilidade dos métodos propostos e a sua aplicabilidade no mundo real. Esta investigação contribui para o domínio da tecnologia educativa ao mostrar como as métricas avançadas de desempenho podem melhorar a eficácia e a personalização dos sistemas de aprendizagem adaptativos. Abre caminhos para a criação de ambientes educativos mais reactivos que respondam eficazmente às diversas necessidades dos alunos.

Palavras-chave: Sistemas de Aprendizagem Adaptativa; Métricas de Desempenho; Tecnologia Educativa; Adaptabilidade do Aluno

1. INTRODUCTION

Adaptive learning systems are now a vital component of individualized education in today's classrooms. These intelligent systems adapt dynamically to the individual needs of each learner by utilizing algorithms and data-driven insights to modify educational content, pacing, and evaluation techniques. In addition to promoting a more effective and engaging learning environment, this individualized approach takes into account the various learning styles and speeds of the pupils.[1][2]

But as these systems become more essential to learning at all levels, the need to improve their efficacy is becoming more widely acknowledged. Though helpful, traditional performance measures frequently give only a partial picture of how well these systems are working. They could overlook the subtleties of how students are really interacting with and benefiting from the adaptive technology in favor of surface-level results like completion rates or overall student happiness.

Advanced performance metrics are an essential component of adaptive learning systems in order to fully realize their potential. These metrics provide a more detailed examination of learning progress, system flexibility, and student engagement than simple data points alone. Teachers and system engineers can obtain deeper insights into the learning process by monitoring and evaluating a wider range of indications, including time spent on tasks, mastery of particular ideas, and even emotional responses. [3]

The design and use of adaptive learning systems can be greatly enhanced by including these advanced metrics. It enables more accurate modifications to be made to the course material, guaranteeing that every student gets the most applicable and efficient education possible.

Moreover, it gives teachers the ability to modify their teaching methods based on data in order to better meet the needs of each unique student.

This article will examine how adaptive learning systems can be improved and made more personalized and successful through the integration of advanced metrics. We will look at the different kinds of metrics that can be used, their advantages, and the difficulties in putting them into practice. By doing this, we hope to shed light on the next frontier of adaptive learning, which is the marriage of data and technology to produce genuinely transformative learning environments.

2. THEORETICAL FRAMEWORK

2.1 Adaptive Learning Systems

Adaptive learning systems (ALS) leverage advanced algorithms and data analytics to transform education by creating highly individualized learning experiences tailored to each student's unique needs. By analyzing learner profiles—comprising prior knowledge, learning pace, and preferences—ALS dynamically adjusts instructional content, learning pathways, and assessment techniques. These systems ensure that the learning process is personalized, allowing students to progress at their own pace while addressing their specific strengths and weaknesses (McCarthy, 2016).

The foundation of adaptive learning systems lies in constructivist learning theory, which posits that knowledge is actively constructed through interactions and experiences rather than passively absorbed. ALS operationalizes this theory by creating adaptive environments that respond to the learner's inputs and progress, fostering deeper engagement with the material (Bruner, 1996). As students interact with the content, the system refines its understanding of their abilities and adapts future content to optimize retention, mastery, and engagement. This iterative and responsive approach aligns closely with how constructivist theory advocates for learning as a dynamic and personalized process.

One of the most innovative features of ALS is the integration of real-time feedback mechanisms. These systems provide instant, actionable insights into a learner's progress, ensuring that students are constantly aware of their performance. This interactive feedback loop encourages students to take a proactive role in their learning journey, empowering them to identify and address gaps in their understanding immediately (Johnson et al., 2017). The motivational aspect of real-time feedback cannot be overstated; it transforms the learning process from a passive experience into an active, engaging dialogue between the learner and the system.

Adaptive learning systems go beyond personalization by incorporating predictive analytics, which enable the system to anticipate a learner's needs based on their current performance and historical data. By identifying potential challenges before they become significant obstacles, ALS can provide targeted interventions, such as additional resources, alternative explanations, or tailored exercises (Clow, 2013). This proactive approach reduces the likelihood of disengagement or dropout and fosters continuous participation. Predictive analytics also allow educators to make data-driven decisions about curriculum adjustments and instructional strategies, ensuring that the overall learning environment remains effective and supportive.

Incorporation of Multimodal Learning Strategies

To accommodate diverse learning styles, ALS integrates multimodal learning strategies, which include kinesthetic, auditory, and visual elements. This feature ensures that students with varying preferences and abilities are effectively engaged and supported. For instance, a visual learner might benefit from infographics and videos, while a kinesthetic learner could engage with interactive simulations or hands-on exercises (Mayer, 2009). This adaptability aligns with the principles of Universal Design for Learning (UDL), which emphasize equity, inclusivity,

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

and accessibility in education (Rose & Meyer, 2002). By embracing these strategies, ALS fosters an inclusive environment that meets the needs of all learners, regardless of their preferred modalities or potential learning barriers.

ALS offers several benefits:

- Enhanced Retention and Mastery: By tailoring the pace and content to individual learners, ALS promotes better retention of knowledge and mastery of skills, ensuring students build a solid foundation before moving forward (McCarthy, 2016).
- Increased Engagement: Interactive features, such as real-time feedback and multimodal content delivery, keep students engaged and motivated throughout their learning journey (Johnson et al., 2017).
- Proactive Problem-Solving: Predictive analytics and interventions prevent challenges from escalating, reducing frustration and dropout rates (Clow, 2013).
- Inclusivity and Accessibility: By incorporating UDL principles, ALS ensures that all learners, regardless of their abilities or preferences, can access and benefit from the system (Rose & Meyer, 2002).

Despite their potential, adaptive learning systems face challenges that must be addressed for optimal implementation. The cost of development and deployment can be a barrier, particularly for underfunded educational institutions. Privacy concerns related to the collection and analysis of learner data must also be carefully managed to ensure compliance with ethical and legal standards (Prinsloo & Slade, 2017). Additionally, the success of ALS depends on the quality and accuracy of the algorithms driving personalization, making continuous refinement and validation essential.

As technology evolves, adaptive learning systems are likely to become even more sophisticated, incorporating emerging technologies like artificial intelligence (AI) and natural language processing (NLP) to enhance their capabilities. The integration of virtual and augmented reality could further enrich the learning experience, providing immersive environments for skill development and exploration. Collaborative features that facilitate peer learning and teamwork within adaptive systems could also bridge the gap between personalized and group-based learning (Luckin et al., 2016).

2.2 Performance Metrics

Performance metrics are both quantitative and qualitative indicators used to evaluate the effectiveness of educational processes, particularly in technology-integrated classrooms. In the context of Adaptive Learning Systems (ALS), these metrics are essential tools for assessing the system's overall efficacy while also tracking learner progress and adaptability. Traditional performance metrics, such as time spent on tasks, completion rates, and test scores, remain relevant for understanding basic learning outcomes (Anderson, 2019). However, advanced metrics have emerged to provide deeper insights, including student engagement levels, concept mastery rates, learning path efficiency, and even emotional responses (Baker & Siemens, 2014). These advanced metrics align with the theoretical foundations of educational assessment, which emphasize the importance of diverse measures for capturing the complexity of learning outcomes (Black & Wiliam, 1998).

Advanced performance metrics in ALS extend beyond traditional indicators by leveraging data analytics and artificial intelligence to continuously assess learners' progress. These metrics not only evaluate academic performance but also provide real-time feedback to improve the system itself. For example, emotional response tracking, often measured through biometric data or sentiment analysis, can detect dissatisfaction, frustration, or disengagement in learners. This information allows the system to trigger interventions, such as adjusting the difficulty level or providing motivational prompts, to re-engage the student (D'Mello et al., 2017).

Br. J. Ed., Tech. Soc., v.18, n.se1, p.22-36, 2025
DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

Additionally, metrics such as concept mastery rates and learning path efficiency offer insights into how well learners are progressing through the material. Concept mastery rates measure the depth of understanding for individual topics, while learning path efficiency assesses how effectively learners navigate through the curriculum to achieve their goals (Desmarais & Baker, 2012). These indicators not only evaluate individual performance but also help identify areas where the system can optimize instructional strategies.

In addition to cognitive metrics, ALS increasingly incorporates measures of socio-emotional traits, such as resilience, confidence, and emotional intelligence. These traits are critical for long-term success, as they influence a learner's ability to overcome challenges, stay motivated, and apply knowledge in real-world settings (Pekrun, 2011). By integrating socio-emotional metrics into performance assessments, ALS aligns with modern educational frameworks that emphasize the holistic development of learners. These frameworks recognize that fostering emotional intelligence is as important as developing cognitive skills for success in both academic and professional environments (Rose & Meyer, 2002).

As adaptive learning technologies are adopted across diverse educational settings, the scalability of performance metrics becomes increasingly critical. Metrics must be designed to accommodate a variety of learner populations, subject areas, and educational levels. For instance, an ALS used in elementary education must capture different data points than one used in higher education or corporate training. Scalability ensures that the system remains accurate, resilient, and effective, regardless of the context in which it is implemented (Luckin et al., 2016). Furthermore, scalable metrics enable comparative analyses across different learning environments, contributing to a broader understanding of the impact of adaptive technologies on education.

Advanced performance metrics not only evaluate learner outcomes but also serve as feedback mechanisms to refine the algorithms powering ALS. By analyzing patterns of success and failure, the system can dynamically adjust instructional tactics to better meet learners' needs. For example, if engagement metrics indicate that a particular instructional approach is not effective, the system can switch to alternative methods or suggest additional resources. This iterative process ensures that ALS remains adaptive, responsive, and aligned with the evolving needs of learners (McCarthy, 2016).

The incorporation of advanced performance metrics in adaptive learning systems represents a significant advancement in educational assessment. These metrics provide a comprehensive understanding of both academic and socio-emotional outcomes, enabling educators and technologists to optimize learning experiences. By focusing on scalability, accuracy, and adaptability, performance metrics ensure that ALS can be effectively implemented across diverse educational contexts. As educational institutions continue to adopt adaptive technologies, the role of sophisticated metrics in driving innovation and improving learning outcomes will become increasingly important.

2.3 Learner Adaptability

Learner adaptability refers to the capacity of a student to adjust to and thrive in dynamic, personalized learning environments, such as those provided by adaptive learning systems (ALS). This concept is closely tied to self-regulated learning theory, which posits that students are more likely to succeed if they can modify their learning strategies in response to feedback and changing conditions (Zimmerman, 2002). In the context of ALS, learner adaptability is critical because it reflects the student's ability to adjust to different paces, instructional methods, and content formats offered by the system.

The relationship between learner adaptability and ALS is inherently reciprocal. While adaptive systems adjust their content and instructional approaches to suit individual learners, students must also adapt to the system's evolving demands and opportunities. This dynamic

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

interaction fosters a personalized learning environment where both the system and the learner work collaboratively toward educational goals. For instance, learners who are flexible in adopting new tools or adjusting their approaches to meet the system's requirements are more likely to reap the benefits of personalized learning pathways.

Learner adaptability is deeply linked to the development of metacognitive skills, such as self-awareness, goal-setting, and reflective learning. Adaptive learning systems that emphasize these skills not only enhance immediate academic outcomes but also promote lifelong learning habits (Flavell, 1979). For example, learners who regularly reflect on their progress and adjust their strategies accordingly are better equipped to tackle real-world problems. ALS can support this process by providing structured opportunities for reflection, such as progress dashboards and prompts for goal revision.

Metacognition also empowers students to take ownership of their learning journeys. By understanding their strengths and areas for improvement, learners can make informed decisions about how to engage with the system, increasing their overall adaptability and academic success.

Motivation and emotional resilience are crucial factors in supporting learner adaptability. Adaptive systems that provide timely and relevant feedback help maintain learners' motivation, even when they face challenging tasks. This feedback not only enhances their understanding of the material but also builds confidence in their ability to overcome obstacles.

When combined with personalized learning pathways, emotional support from ALS creates an environment where students feel empowered to take charge of their education. For instance, systems that acknowledge milestones or provide encouraging messages during difficult tasks foster a sense of accomplishment and persistence. These motivational strategies are essential for sustaining engagement, especially in the face of complex or unfamiliar content.

Adaptive learning systems support not only academic growth but also personal development by fostering adaptability. By helping students develop resilience, self-efficacy, and a willingness to embrace challenges, these systems prepare learners for success beyond the classroom. For example, a student who learns to adapt to a system's changing requirements—such as shifting from video-based content to hands-on simulations—may transfer this skill to professional or everyday problem-solving contexts.

Moreover, the ability to adapt promotes critical thinking and flexibility, which are increasingly valued in today's fast-changing world. By addressing both cognitive and emotional dimensions, ALS cultivates well-rounded individuals equipped to thrive in diverse environments.

Learner adaptability is a cornerstone of effective engagement with adaptive learning systems. Rooted in self-regulated learning and metacognitive development, adaptability enables students to navigate personalized, dynamic learning environments successfully. By fostering skills such as self-awareness, goal-setting, and emotional resilience, ALS not only enhances academic outcomes but also promotes lifelong learning and holistic development. Adaptive systems that prioritize motivation, provide timely feedback, and support individualized pathways create an empowering atmosphere where students are encouraged to take charge of their education. This interplay between system adaptability and learner adaptability forms the foundation for meaningful and transformative educational experiences.

2.5 Cohen's Kappa

Cohen's Kappa is a robust statistical metric used to evaluate the extent of agreement between two raters or classifiers, accounting for the possibility of agreement occurring by chance. In the context of adaptive learning systems (ALS), Cohen's Kappa serves as a valuable tool for assessing the alignment between expected learning outcomes and actual learner performance. Unlike simple accuracy, which merely measures the proportion of correct

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

predictions, Cohen's Kappa adjusts for chance agreement, providing a more reliable and nuanced measure of actual agreement (Cohen, 1960).

Adaptive learning systems often rely on machine learning models to predict learner outcomes, such as whether a student will successfully master a topic or require additional support. Cohen's Kappa is particularly suited for evaluating the performance of these models, as it considers the imbalance that often exists in class distributions. For instance, in an individualized Python course, most students may achieve success while only a minority may struggle. In such scenarios, accuracy alone may overestimate the model's effectiveness because it is disproportionately influenced by the majority class.

Cohen's Kappa mitigates this issue by weighting agreement based on class distributions, offering a fairer evaluation of the model's predictive performance. A Kappa value of 1 indicates perfect agreement, while a value of 0 signifies no better agreement than chance. Values less than 0 suggest disagreement. By using Cohen's Kappa, adaptive systems can ensure that their predictions are effective for all students, not just the majority, thus enhancing equity in personalized learning environments (McHugh, 2012).

Unbalanced datasets, where one outcome is significantly more prevalent than others, are a common challenge in adaptive learning systems. For example, in a scenario where most students perform well but a smaller group requires additional help, a model could achieve high accuracy by focusing predominantly on the majority class. However, this would fail to address the needs of struggling learners.

Cohen's Kappa effectively addresses this issue by accounting for the expected agreement due to class imbalance. By doing so, it prevents the metric from being unduly influenced by the majority class, ensuring a more accurate evaluation of the system's ability to support all students. This is especially critical in adaptive learning, where the goal is to provide tailored support to individual learners, including those who face challenges.

Cohen's Kappa is invaluable for system designers and educators during the iterative development of adaptive learning systems. By analyzing Kappa values, system developers can identify specific areas where predictions routinely fail, such as incorrectly classifying struggling students as successful. This focused insight allows for targeted improvements to the algorithms and instructional strategies, enhancing the overall effectiveness of the system.

For example, if a low Kappa value is observed in predicting the need for additional support in a math module, system designers can investigate whether the issue stems from insufficient data, flawed model features, or an imbalance in the training dataset. This iterative refinement process is crucial for developing reliable and adaptive systems that cater to the diverse needs of learners (Artstein & Poesio, 2008).

Cohen's Kappa is widely used in educational settings to evaluate the consistency and accuracy of machine learning models. For instance, in a programming course, it might measure how well the system predicts whether a student will pass a unit or require supplementary material. A high Kappa value indicates that the system's predictions align closely with actual outcomes, affirming the reliability of the system's recommendations. Conversely, a low Kappa value signals areas for improvement, such as refining prediction models or incorporating additional learning data.

Furthermore, the interpretability of Cohen's Kappa makes it a practical tool for educators. By pinpointing specific areas of disagreement, educators can tailor interventions to address gaps in learning, ensuring that students receive the support they need to succeed.

1. **Reliability:** Cohen's Kappa adjusts for chance agreement, providing a more trustworthy measure of model performance compared to simple accuracy.
2. **Fairness:** The metric accounts for unbalanced datasets, ensuring that predictions are evaluated equitably across all learner groups.

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

3. **Actionable Insights:** By highlighting areas of disagreement, Cohen's Kappa facilitates targeted improvements in adaptive learning systems.
4. **Scalability:** Its applicability to diverse educational settings and data distributions makes it a versatile tool for evaluating ALS performance.

Cohen's Kappa is a powerful metric for assessing the reliability and effectiveness of adaptive learning systems. By accounting for chance agreement and addressing class imbalances, it provides a nuanced evaluation of predictive performance. Its interpretability enables system designers and educators to identify and address weaknesses, supporting iterative development and continuous improvement. In the context of ALS, Cohen's Kappa not only enhances the accuracy and consistency of system recommendations but also ensures that personalized learning experiences are equitable and effective for all learners.

2.6 F1-Score :

The F1-Score is an important performance metric that is particularly effective in addressing unbalanced class distributions, a common challenge in adaptive learning systems (ALS). As the harmonic mean of precision and recall, the F1-Score balances the trade-offs between these two critical components of classification performance. Precision refers to the proportion of correctly identified positive cases out of all instances labeled as positive, while recall measures the proportion of correctly identified positive cases out of all actual positives (Powers, 2011). This balance makes the F1-Score an ideal metric for evaluating ALS, where ensuring accurate identification of learner needs is essential.

In adaptive learning systems, students are often categorized into groups such as "mastered content" or "requires additional practice." The F1-Score ensures that the system effectively identifies students needing assistance while minimizing false positives (incorrectly classifying students as needing help) and false negatives (failing to identify students who actually need help). Unlike accuracy, which can be misleading in cases of imbalanced datasets, the F1-Score provides a more comprehensive measure of model performance, especially in scenarios where the costs of misclassification are significant (Saito & Rehmsmeier, 2015).

For instance, in an ALS for mathematics, a false positive could result in a student being unnecessarily assigned remedial content, leading to disengagement. Conversely, a false negative could leave a struggling student without the necessary support. By optimizing for the F1-Score, ALS can strike a balance between these outcomes, ensuring accurate and equitable support for all learners.

The F1-Score is particularly valuable for monitoring the effectiveness of interventions within adaptive learning systems. For example, when the system adjusts content to meet a student's needs, the F1-Score provides immediate feedback on the success of those changes. A high F1-Score would indicate that the system is accurately identifying students' difficulties and providing effective interventions, while a lower score would suggest the need for adjustments in the system's algorithms or instructional strategies. This continuous monitoring enables teachers and system developers to make dynamic improvements, ensuring that the system remains responsive to learners' evolving needs.

Another strength of the F1-Score is its ability to detect subtle patterns in learner performance. By balancing precision and recall, the F1-Score ensures that even rare instances of student difficulties are identified. This level of granularity allows for the development of highly personalized learning pathways that address the unique strengths and weaknesses of each student. For example, if a small subset of students consistently struggles with a specific concept, the F1-Score ensures that these cases are neither overlooked nor disproportionately emphasized, enabling targeted interventions.

The F1-Score is also robust in multi-class classification problems, which are common in adaptive systems that handle diverse learning objectives. In such scenarios, the metric evaluates performance across multiple categories, ensuring that all aspects of learner development are effectively addressed. For instance, in a language-learning ALS, the F1-Score can assess the system's ability to classify learners' proficiency in reading, writing, speaking, and listening, providing a holistic view of their progress.

By offering a detailed analysis of performance across multiple dimensions, the F1-Score supports a more inclusive approach to education, where all learners receive the tailored assistance they need to succeed. This comprehensive assessment ensures that no aspect of a student's development is overlooked, fostering equitable outcomes for diverse learner populations.

3. METHODOLOGY

Using a mixed-methods approach, this study examines how advanced performance metrics might be integrated into Adaptive Learning Systems (ALS) to improve their effectiveness. The study focuses on a sample of 60 master's degree students who are enrolled in a Python programming course. In particular, the study looks at how well metrics like Cohen's Kappa and the F1-Score work to resolve the unequal class distributions that are common in personalized learning paths.

In a Python course, if an adaptive learning system predicts whether students have understood a concept based on their interaction data, the F1-Score will ensure that the system maintains a high standard of both precision and recall. This balance is essential in providing effective and targeted interventions, which are crucial for student success in complex subjects like programming.

3.1 Participants

The study involved 60 master's degree students in Pedagogical engineering Multimedia, divided into two groups: a control group (n=30) and an experimental group (n=30). Participants were randomly assigned to each group.

- **Control Group:** This group used the ALS that relied on traditional accuracy-based performance metrics.
- **Experimental Group:** This group used the ALS enhanced with advanced performance metrics, namely the F1-Score and Cohen's Kappa.
- **Demographic Information:**
    - Participants ranged in age from 22 to 30 years old, both male and female students.
    - Although the participants' competence with Python varied, all had prior experience with basic programming.

3.2 Procedure

The process was created to evaluate how the advanced measures affected the ALS's capacity to forecast student performance with accuracy and modify the learning path as necessary.

Pre-Assessment: Each participant filled out a pre-assessment to determine their baseline Python programming expertise before the intervention. This test made sure that the baseline knowledge of the two groups was similar.

Random Assignment: Participants were randomly assigned to either the control or experimental group to minimize biases.

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

Course Delivery: Both groups completed the same Python programming course over four weeks. The course covered key topics, including Basic Syntax, Control Structures, Functions, Object-Oriented Programming (OOP), Data Structures, and Recursion.

Based on the interactions and performance of the learners, the ALS offered individualized learning paths.

Performance Metrics:

Control Group: The ALS employed traditional accuracy as the primary metric for adapting learning paths and predicting student success.

Experimental Group: The ALS included standard accuracy as well as Cohen's Kappa and the F1-Score. By using these measurements, the system's adaptation processes were improved, especially when it came to managing imbalanced class distributions (students who found it difficult to study advanced material, for example).

Post-Assessment: Each participant completed a post-assessment at the end of the course to gauge their level of understanding and the ALS's efficacy.

The students' performance on the post-assessment and their interactions with the ALS were used to construct the performance measures (accuracy, F1-Score, Cohen's Kappa).

Data Collection: The data were analyzed to compare the effectiveness of the ALS between the control and experimental groups.

3.3 Data Analysis

The goal of the data analysis was to compare the ALS performance in the two groups by applying the following statistical techniques:

Descriptive Statistics: For the pre- and post-assessment scores in both groups, means and standard deviations were calculated.

Descriptive statistics were employed to provide an overview of each topic's accuracy, F1-Score, and Cohen's Kappa performance measures.

Inferential Statistics: A paired t-test was used to discover if there was a significant difference in pre-assessment and post-assessment scores within each group.

The post-assessment scores of the control and experimental groups were compared using an independent t-test.

To determine the extent of the group differences, effect sizes were calculated.

Comparative Analysis: The F1-Score and Cohen's Kappa were compared with traditional accuracy metrics to assess their effectiveness in improving the ALS's predictive accuracy, particularly in topics with imbalanced class distributions.

Interpretation: The objective of the analysis was to ascertain whether the ALS's capacity to deliver individualized learning experiences and raise student outcomes was considerably enhanced by the implementation of advanced metrics, such as Cohen's Kappa and F1-Score.

4. RESULTS

4.1 Experimental Design

The study was conducted with 60 master's degree students, randomly assigned to:

Control Group (n = 30): Utilized the ALS with traditional accuracy-based metrics.

Experimental Group (n = 30): Utilized the ALS with advanced metrics, including the F1-Score and Cohen's Kappa.

Both groups completed pre- and post-assessment tests on six course topics.

4.2 Assessment and Performance Metrics

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

The pre- and post-assessment scores were compared to measure improvements.

Table 1. Control Group Results - Pre and post-assessment scores

Topic	**Pre-Assessment (%)**	**Post-Assessment (%)**
Basic Syntax	50.0	70.0
Control Structures	46.7	63.3
Functions	40.0	56.7
OOP	36.7	50.0
Data Structures	30.0	43.3
Recursion	26.7	36.7

Source:

Table 2. Experimental Group Results - Pre and post-assessment scores

Topic	**Pre-Assessment (%)**	**Post-Assessment (%)**
Basic Syntax	50.0	80.0
Control Structures	46.7	76.7
Functions	40.0	70
OOP	36.7	63.3
Data Structures	30.0	56.7
Recursion	26.7	50

Source:

4.3 Statistical Analysis

The following t-tests were conducted to measure the improvements:

Paired t-test (for pre- and post-assessment within each group):

$$\mathbf{t} = \frac{\underline{d}}{S_d \,/\, \sqrt{n}}$$

Where:

$\underline{d}$ = Difference between pre- and post-assessment scores
S_d = Standard deviation of the differences
n = Number of students
Compute differences:
$d_i = Post - assessment_i$ - $Pre - assessment_i$
Compute mean difference ($\underline{d}$) and standard deviation (S_d) of differences:

$$\underline{d} = \frac{\Sigma\, d_i}{n}$$

$$S_d = \sqrt{\frac{\Sigma(d_i - \underline{d}\,)^{\;2}}{n-1}}$$

Suppose:
Mean difference $\underline{d}$ = 20
Standard deviation S_d = 5
Number of students n = 30

$$\mathbf{t} = \frac{20}{5 \,/\, \sqrt{30}} \approx 21.8$$

Compare with critical value from t-distribution tables for df=29 at α=0.05
If **t** exceeds the critical value, the improvement is statistically significant.

Br. J. Ed., Tech. Soc., v.18, n.se1, p.22-36, 2025
DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

Independent t-test (for comparing improvements between control and experimental groups):

$$\mathbf{t} = \frac{\underline{X_1} - \underline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$\underline{X_1}$, $\underline{X_2}$ = **Mean improvements of the experimental and control groups, respectively**
s_1, s_2 = **Standard deviations of the improvements**
n_1, n_2 = **Number of students in each group**

Example Calculation:
Mean improvement (Control Group) = 30.0%
Mean improvement (Experimental Group) = 60.0%
Standard deviation (Control Group) = 10
Standard deviation (Experimental Group) = 15

$$\mathrm{t} = \frac{60.0 - 30.0}{\sqrt{\frac{10^2}{30} + \frac{15^2}{30}}} \approx 4.5$$

Compare with critical value from t-distribution tables for df=58 at α=0.05.
If t exceeds the critical value, the difference between groups is statistically significant.

4.4 Interpretation of Results:

The t-tests confirm :

Greater Improvement in Experimental Group: Compared to the Control group, the Experimental group greatly outperformed it in all themes, especially in the more difficult ones.

Effectiveness of Advanced Metrics: More accurate and dependable evaluations of student performance were made possible by the use of advanced metrics, such as Cohen's Kappa and F1-Score.

Statistical Significance: The improved ALS's efficacy was validated by the t-tests, which showed statistically significant differences in improvements.

These findings highlight the benefits of ALS's integration of sophisticated performance metrics, which also improve educational outcomes and tailored learning environments.

5. DISCUSSION

The results of this study offer strong proof that incorporating sophisticated performance measurements, like the F1-Score and Cohen's Kappa, greatly improves the effectiveness of adaptive learning systems. A noteworthy finding is that these measurements are better equipped to capture the nuances of individualized education than traditional accuracy metrics, offering a more nuanced view of learner behavior and outcomes.

The improved ability of the experimental group's ALS to recognize and address learning difficulties, especially in complicated subjects like data structures and recursion, is a particularly noteworthy finding. By eliminating false negatives and guaranteeing that struggling students received timely interventions, the enhanced metrics made it possible to identify kids who needed extra support more precisely. This research shows how sophisticated

Br. J. Ed., Tech. Soc., v.18, n.se1, p.22-36, 2025
DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

measurements can improve the equity of adaptive learning systems and better meet the demands of a wide range of learners.

These findings have wider ramifications for adaptive learning technologies' scalability. Strong, scalable metrics are becoming more and more necessary as educational institutions use data-driven solutions more and more. By guaranteeing consistent performance across various educational contexts and student groups, advanced metrics such as the F1-Score provide a solution to this problem by balancing precision and recall across numerous categories.

The study also emphasizes how sophisticated metrics can help create a more welcoming learning environment. Metrics like Cohen's Kappa make ensuring that adaptive learning systems don't unfairly favor high-performing students by resolving imbalanced datasets and giving priority to identifying underperforming pupils. Promoting equitable and easily accessible learning opportunities requires this alignment with educational equity goals.

The importance of these measures in iterative system development is another important realization. Adaptive learning algorithms can be continuously improved by using metrics like Cohen's Kappa, which offer detailed feedback on system performance. In addition to improving the system's existing effectiveness, this iterative method lays the groundwork for upcoming developments in personalized education technologies.

The results also highlight the significance of a comprehensive framework for evaluation that takes into account both socioemotional and cognitive aspects. Although academic achievements were the main focus of the study, adding emotional and engagement metrics could improve ALS's customisation skills even more. The increasing focus on holistic education, which aims to create well-rounded students ready for both academic and real-world difficulties, is consistent with this multifaceted approach.

In summary, the incorporation of sophisticated performance indicators is a revolutionary development for adaptive learning systems. These measures not only improve educational outcomes but also open the door to more inclusive, equitable, and responsive learning environments by providing a more thorough assessment of student development and system efficacy.

6. CONCLUSION

The study shows how Adaptive Learning Systems (ALS) in a master's degree Python programming course can greatly benefit from the integration of advanced performance metrics, namely Cohen's Kappa and the F1-Score. These advanced measures offer a more nuanced assessment of system performance by resolving the shortcomings of traditional accuracy-based metrics, especially when considering the unbalanced class distributions common of personalized learning pathways.

The experimental findings clearly show that the ALS improved with these advanced metrics produces better results, as shown by increased predictability, consistency, and accuracy in anticipating and adjusting to the specific demands of each learner. In every major area, the experimental group performed better than the control group; considerable gains were made in difficult subjects like data structures and recursion.

These findings underscore the significance of integrating advanced assessment instruments into instructional technology to augment their flexibility and efficacy. Adopting such sophisticated metrics in ALS is essential for improving learning outcomes and learning experiences as educational environments continue to change. This method not only increases the precision of learner evaluations but also creates a more successful and individualized learning experience, opening the door for further advancements in educational technology

Furthermore, these metrics' scalability and adaptability make them perfect for a variety of learning situations, guaranteeing that adaptive learning systems continue to be inclusive and egalitarian. They support learner engagement, confidence, and long-term success in addition to better academic results.

To further customize learning experiences, future studies could investigate including other variables, such as socio-emotional factors. By doing this, adaptive systems may keep developing to satisfy the many and ever-changing demands of students around the globe, opening the door for more inclusive, efficient, and responsive educational institutions. In the end, this study marks a significant advancement in using data-driven insights to revolutionize contemporary education.

REFERENCES

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Anderson, T. (2019). *The theory and practice of online learning* (3rd ed.). Athabasca University Press.

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. *Journal of Educational Research, 1*(3), 89–102. https://doi.org/10.1016/j.edurev.2013.12.001

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction, 22*(1-2), 9–38. https://doi.org/10.1007/s11257-011-9106-8

D'Mello, S. K., Dieterle, E., & Duckworth, A. L. (2017). Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104–123. https://doi.org/10.1080/00461520.2017.1281747

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An argument for AI in education*. Pearson Education.

McCarthy, B. (2016). *The 4MAT System: Teaching to Learning Styles with Right/Left Mode Techniques*. About Learning, Inc.

Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In R. A. Calvo & S. K. D'Mello (Eds.), *New perspectives on affect and learning technologies* (pp. 23–39). Springer.

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning.* ASCD.

Bruner, J. S. (1996). *The culture of education*. Harvard University Press.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2017). *NMC Horizon Report: 2017 Higher Education Edition.* EDUCAUSE.

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An argument for AI in education*. Pearson Education.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

McCarthy, B. (2016). *The 4MAT System: Teaching to Learning Styles with Right/Left Mode Techniques*. About Learning, Inc.

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907

Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: Charting the (un)charted. *Australasian Journal of Educational Technology, 33*(3). https://doi.org/10.14742/ajet.2912

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal Design for Learning.* ASCD.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics, 34*(4), 555–596. https://doi.org/10.1162/coli.07-034-R2

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness, and correlation. *Journal of Machine Learning Technologies, 2*(1), 37–63.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Van Rijsbergen, C. J. (1979). *Information retrieval* (2nd ed.). Butterworth-Heinemann.

DOI http://dx.doi.org/10.14571/brajets.v18.nse1.22-36
ISSN 2316-9907