Quality & Quantity
https://doi.org/10.1007/s11135-025-02545-5

Modeling PISA 2022 student performance with interpretable fuzzy methods: a comparison of FPM and ANFIS

Aras Beren Tatlı[1] · Nuri Can Aksoy[2]

Received: 25 May 2025 / Accepted: 12 December 2025

Abstract

This study explores the potential of the Fuzzy Propositional Model (FPM) for predicting student achievement using process data from the PISA 2022 mathematics tasks in the Turkish sample. The model interprets students' problem-solving behaviours through rule-based reasoning and triangular membership functions, providing insights into how learning processes unfold rather than focusing solely on correctness. The results indicate that the FPM yields pedagogically meaningful interpretations of behavioural indicators such as response time, number of actions, and task revisits, linking them to varying achievement levels. Although data-driven models like ANFIS may achieve marginally higher numerical precision, the FPM stands out by offering transparent, interpretable rules that enhance educational understanding and support data-informed decision-making. These findings demonstrate that explainable fuzzy logic models can serve as practical tools in large-scale assessments, helping educators and policymakers transform process data into actionable insights about student learning.

Keywords PISA 2022 · Fuzzy logic · Fuzzy propositional model · Student performance · Process data

1 Introductıon

Interpreting process data from large-scale assessments is increasingly important because these data capture how students navigate tasks, not only whether they answer correctly. Recent technical documentation of PISA 2022 formalizes the collection and structure of

✉ Nuri Can Aksoy
ncan.aksoy@hku.edu.tr

Aras Beren Tatlı
aras.tatli@lisa.swiss

1 International School Altdorf, Altdorf, Switzerland

2 Department of Elementary Education, Faculty of Education, Hasan Kalyoncu University, Gaziantep, Turkey

Published online: 29 December 2025

timing and interaction indicators, underscoring their intended use for secondary analyses of response processes (OECD 2024).

Log- and process-data studies consistently show that behavioral indicators, such as total response time and number of actions, provide insight into cognitive strategies and engagement beyond static scores. Recent syntheses in large-scale assessment research emphasize both opportunities and methodological cautions associated with using such indicators for modeling (Anghel et al. 2024). In prior work, interpretable fuzzy systems have been applied to educational prediction with human-readable rule bases, demonstrating classroom-friendly explanations, albeit with moderate accuracy (Aziz et al. 2019; Hegazi et al. 2023). Data-driven neuro-fuzzy approaches, such as ANFIS, improve predictive fit by learning membership functions and rules from data, however they typically reduce transparency (Jang 1993; Kasabov 1996; Parkavi et al. 2024). Recent studies (Kalaycı Alas 2024; Loan et al. 2024) further support the integration of fuzzy and neuro-fuzzy systems for student performance prediction, emphasizing the need to balance accuracy with interpretability. At the assessment-system level, process/log-data research show that behavioral indicators capture aspects of strategy and engagement beyond correctness, motivating models that can connect these behaviors to interpretable inferences (Anghel et al. 2024; Greiff et al. 2016). Within this space, FPM positions itself as an explainable, low-resource alternative that trades slight losses in numerical precision for gains in pedagogical interpretability.

While the traditional understanding of education, generally defines success by exam scores and academic performance, today it is considered insufficient to limit success solely to academic criteria. Now, emotional, social, and cognitive development of the individual is also considered as a part of the concept of success (Zins et al. 2007). In the twenty-first century, prominent skills include thinking skills such as critical thinking, communication, self-regulation, problem solving and emotional awareness (Trilling and Fadel 2009). These skills enable the individual to cope with situations encountered both in school life and daily life, contributing to their multidimensional development (Sleegers et al. 2005). Therefore, education systems need to address success with a multidimensional approach. For students to be effective and productive individuals not only in academic exams but also in various areas of life, education policies and practices need to be restructured in this direction (OECD 2019). There are learning measurement systems International Mathematics and Science Study (TIMSS) and Programme for International Student Assessment (PISA), which are widely used to compare the education systems of countries globally. Large-scale international assessment programmes also adopt this approach, recording in detail not only whether students answer correctly but also how they solve a task (OECD 2024).

PISA has been conducted by the Organisation for Economic Co-operation and Development (OECD) on a triennial basis since 2000, with Turkey participating for the first time in 2003. PISA was conducted in 2015, 2018 and 2022 with the participation of 72, 79 and 81 countries and economies, respectively. This exam aims to measure the extent to which 15-year-old students can apply the knowledge and skills they acquire at school to real-life situations. PISA assessments are conducted in three main areas: reading, mathematics and science. In each exam period, one of these areas is determined as a 'weighted area' and more in-depth assessments are made in that area. In the 2022 PISA exam, mathematical literacy was chosen as the weighted area. Mathematical literacy is defined by the OECD as 'the ability to identify, analyse and find solutions to problems encountered by individuals through mathematical means' (OECD 2019, 2023).

In summary, this section establishes the rationale for examining process data in large-scale assessments and introduces the theoretical motivation for adopting interpretable fuzzy modeling approaches. The present study builds upon the growing interest in explainable, process-based modeling approaches and positions the Fuzzy Propositional Model (FPM) as a potential solution to bridge the gap between predictive precision and interpretability in large-scale datasets such as PISA.

Based on this theoretical framework, this study analyses student interaction data for the 'quantity domain' mathematics questions in the PISA 2022 Turkey sample. The primary objective of this study is to investigate the differences between process-based achievement predictions derived from FPM and traditional scoring and official PISA analytics approaches.. The PISA 2022 Turkey results strengthen the context of this study. The average score of Turkish students in mathematics was 453, which is below the OECD average. Moreover, only 5.4 per cent of students reached the highest level of achievement, indicating that most of the achievements are clustered at the basic level (Ministry of National Education [MoNE] 2023). While this situation highlights the need for a better understanding of individual differences, it also underscores the importance of explainable modelling approaches that make sense of process data. Therefore, in the Turkish context, fuzzy logic-based, low-resource and interpretable systems are of potential value in terms of education policies.

The basic assumption of this research is that a fuzzy logic-based assessment of student behaviour can provide more descriptive and nuanced achievement predictions compared to classical methods. Building on this foundation, the study's aims and research questions are articulated below to make explicit how interpretability and predictive utility are jointly evaluated in the PISA 2022 context. In this direction, the research seeks answers to the following questions:

1. To what extent can process data (response time, number of actions, etc.) differentially predict achievement levels compared to classical scoring approaches?
2. Can a fuzzy logic-based model accurately predict student achievement using PISA interaction data?
3. Can the Fuzzy Propositional Model (FPM) provide a valid alternative to traditional fuzzy inference systems (e.g. Mamdani) in terms of low resource utilisation and high explainability?
4. How do the process-based achievement predictions obtained with the FPM differ from the results of PISA's classical scoring approach and formal analyses?

2 Conceptual framework

Today, educational assessment approaches are gradually shifting away from classical systems that focus solely on fixed test scores towards process-oriented and multidimensional analyses. With this transformation in education, multidimensional approaches have started to be adopted in the evaluation of achievement. In education, artificial intelligence (AI) and fuzzy logic systems, which aim to model complex learning behaviours, have become essential tools. Working in place of classical binary logic, fuzzy logic provides a powerful framework for modelling the uncertainties and transitions inherent in human learning

(Zadeh 1999). Mamdani and Sugeno fuzzy inference systems, which are widely used in student achievement modelling studies, offer explainability but require high computational power and extensive rule definitions (Algshat 2024). In this context, the Fuzzy Propositional Model (FPM) proposed by Hegazi et al. (2023) offers an explainable and feasible alternative to classical systems with lower resource consumption and a simpler rule structure. In addition, despite the explainability advantage of classical fuzzy systems, the prediction success may be limited (Liu 2024). In this study, in addition to the FPM approach, a comparative test with ANFIS—a data-driven fuzzy model—was also performed, as further detailed in the methodology section. In this way, the performance of the model was evaluated in terms of both explainability and accuracy. In this study, the concepts of process data, interpretability, and fuzzy modeling are treated as interconnected dimensions: process data represent observable actions, interpretability ensures that inferences remain transparent to educators, and fuzzy modeling provides the computational bridge linking the two.

The present study theorizes three linked constructs. Achievement is the outcome dimension operationalised by item-level performance; mathematical literacy frames the domain in which problem-solving competence manifests; and prediction models provide the mapping from observed process data to expected achievement outcomes. Process data index behaviors during task navigation (e.g., timing, actions, revisits); fuzzy modeling encodes these behaviors as graded linguistic states; and interpretability ensures the mapping remains auditable for educators. Under this framing, FPM functions as a transparent bridge from process signals to achievement in mathematical literacy, while neuro-fuzzy baselines (e.g., ANFIS) test the accuracy–explainability trade-off.

Within educational modeling, classical fuzzy inference systems such as Mamdani and Sugeno offer human-readable rules but often require extensive rule bases and parameter tuning when tasks and indicators proliferate (Mamdani and Assilian 1975; Sugeno 1985). Data-driven neuro-fuzzy systems (e.g., ANFIS) increase predictive fit by learning membership functions and rules, yet they tend to reduce transparency and auditability for pedagogical interpretation (Jang 1993; Kasabov 1996). Positioned between these families, the Fuzzy Propositional Model (FPM) maintains linguistic interpretability with a comparatively compact rule structure while remaining computationally light, which is advantageous for process-data settings like PISA where indicators are multi-source and task-dependent (Hegazi et al. 2023). Recent studies have also highlighted the growing use of fuzzy decision-support and hybrid neuro-fuzzy approaches in educational prediction contexts (Li et al. 2025), reinforcing the relevance of interpretable and efficient modeling solutions. In this design, FPM operationalizes an explainable mapping from observed process indicators to achievement, whereas ANFIS tests the accuracy–explainability trade-off. Accordingly, FPM is adopted here to privilege explainability and rule-level insight, while a neuro-fuzzy baseline is retained to benchmark predictive performance.

3 Method

3.1 Research model and approach

This study analyses students' operation-based process data on three tasks in the ‘multiplicity’ domain of mathematics in the PISA 2022 Türkiye sample. The study aims to predict students' achievement levels using the Fuzzy Propositional Model (FPM) and to provide an explainable approach that is different from classical accuracy-based assessments. The model consists of the steps of fuzzification of process data with triangular membership functions, rule-based inference and sharpening the results. The general flow of this process is shown schematically in Fig. 1.

The Fuzzy Propositional Model (FPM), although relatively new to educational data analysis, was initially designed as an interpretable alternative to conventional fuzzy inference systems. It differs from classical Mamdani and Sugeno models by employing a propositional rule structure and a Fuzzy Set Transformer (FST) matrix that explicitly links input–output relationships. This structure reduces computational complexity and enhances rule-level transparency, making it particularly suitable for process-data contexts where interpretability is essential. Previous research demonstrated the feasibility and robustness of FPM for academic performance prediction (Hegazi et al. 2023), and its theoretical foundation has been validated against other fuzzy systems through comparative analyses (Aziz et al. 2019; Yadav et al. 2014). In the present study, FPM was therefore adopted as a model that balances computational efficiency and interpretability, aligning with educational research needs for transparent, low-resource modeling approaches.

3.2 Data set and participants

PISA 2022 Mathematical Literacy Türkiye data were used in the study. The three questions included in the assessment belong to the 'quantity' sub-field of mathematics. Both scored responses and five process variables (Total Response Time—TT, Initial Response Time—TFA, Number of Actions Taken—NA, Number of Visits—NV, Number of Short Visits—NSV) of these questions were analysed.

In the first stage, data belonging to a total of 919 students were analysed. Missing observations were excluded through the data cleaning process. Students with missing responses in the process-based process data were excluded from the analysis. For each of the three

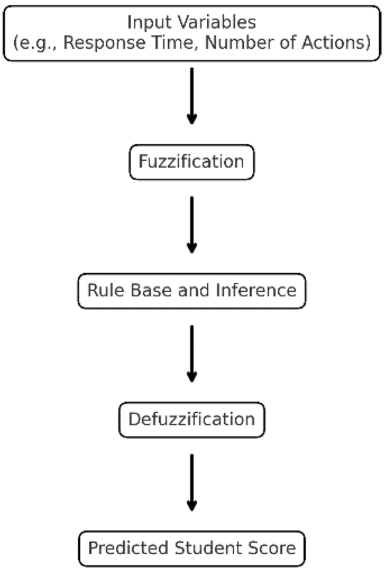

Fig. 1 Prediction process based on the fuzzy propositional model (FPM) structure

tasks (CMA150Q01, CMA150Q02, CMA123Q01), the analysis set was formed by selecting the students with all the required data. For both models (FPM and ANFIS), missing data were removed from the process-based process data and a common analysis set consisting of n=612 students was used.

The three questions in question present students with tasks that measure cognitive processes such as making logical inferences from quantitative data, ordering operations, and applying arithmetic. Three main criteria were considered in the selection of the questions: (1) containing scored responses, (2) being rich in process data, and (3) requiring both basic and advanced operations in terms of content.

3.3 Preliminary analysis of data

In the data set, the process-based process data were normalised using the z-score method. The distributions were modelled by means of membership functions. Outlier analysis was not performed separately. In addition, Pearson correlation coefficient was calculated to test the relationship between the variables (Table 1).

3.4 Setup of the fuzzy logic model

In defining the membership functions; triangular membership functions were defined for each input variable (e.g. Short, Medium, Long). The functions were defined based on the quartiles of the data (Q1, Q2, Q3) and expressed by the following formula:

$$\mu_A(x) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

Min–Max inference method was used in fuzzy rule evaluation:

$$\mu_{\text{output}}(z) = \max[\min(\mu_A(x), \mu_B(y))]$$

The output variable is defuzzified with the Centre of Gravity (CoG) method:

$$Z^* = \frac{\sum z_i \cdot \mu(z_i)}{\sum \mu(z_i)}$$

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were calculated as Performance Evaluation Metrics for the prediction success of the model.

Table 1 Pearson correlations among process variables and actual score

Variables	1	2	3	4	5	6
1. Total Response Time	–	−0.13	0.59	0.18	0.04	−0.02
2. Time to First Action	−0.13	–	−0.07	−0.05	0.01	0.63**
3. Number of Actions	0.59	−0.07	–	0.08	−0.03	0.03
4. Number of Visits	0.18	−0.05	0.08	–	0.29	−0.03
5. Number of Short Visits	0.04	0.01	−0.03	0.29	–	0.01
6. Actual Score	−0.02	0.63**	0.03	−0.03	0.01	–

n=612. **p<0.01

A total of 27 if–then rules were created in the Rule Base Generation phase. The rules were determined based on both patterns in the literature (Aziz et al. 2019; Parkavi et al. 2024; Yadav et al. 2014) and expert judgements. For example: 'Response Time = Short ∧ Number of Transactions = Low ∧ Success = High if Visit = 1'. The relationships between the membership values of the input variables and the success level are modelled by means of the Fuzzy Set Transformer (FST) matrix.

3.4.1 Comparative ANFIS model implementation

In order to evaluate the performance of the FPM model and compare it with alternative data-driven methods, an ANFIS (Adaptive Neuro-Fuzzy Inference System) model was implemented using the same transaction data. The ANFIS model is built with 9 Takagi–Sugeno rules based on classical triangular membership functions and trained on linear output functions. The learning parameters of the model were optimised by least squares regression method. Bootstrap analysis with 1000 iterations was performed to test the statistical significance of the difference in model performances.

3.5 Software and infrastructure

The modelling process was carried out in Python 3.11 environment, pandas and numpy were used for data processing, matplotlib and seaborn were used for graphical analysis, skfuzzy and specially defined functions were used for fuzzy logic operations. To validate the robustness of the model, a stratified 80/20 train-test split was used. Additionally, a 5-fold cross-validation was conducted, and the average RMSE remained consistent across folds. Pearson correlation was also calculated to assess the linear relationship between predicted and actual scores.

4 Findings

In this section, the process data obtained from the PISA 2022 Türkiye sample are analysed with the Fuzzy Propositional Model (FPM) approach. The findings are structured under five sub-headings; starting with descriptive statistics of the process data, visualisation of the membership functions used in the model, evaluation of the predictive power of the model, presentation of error metrics and finally interpretation of rule patterns based on student interactions. In each subsection, the applicability and explainability of the model to educational data are evaluated through both numerical and visual outputs. The results obtained show that achievement prediction based on process data provides not only technical but also pedagogical clues.3.1. Descriptive Statistics for Item CMA150Q01.

Table 2 presents descriptive statistics for the process-related variables associated with item CMA150Q01, a task in the Quantity domain of the PISA 2022 assessment. The variables include total response time, time to first action, number of actions, number of visits, and number of short visits. Descriptive statistics were calculated after excluding extreme values, based on data from 911 Turkish students. Descriptive statistics for the process variables for question CMA150Q01 were calculated on a larger subset of students (n = 911).

Table 2 Summary statistics for process variables—item CMA150Q01 (Quantity Domain, PISA 2022)

Variable Name	Description	Min	Mean	Max
CMA150Q01TT	Total Response Time (ms)	3875	78,497.5	872,615
CMA150Q01F	Time to First Action (ms)	103	44,405.13	275,377
CMA150Q01A	Number of Actions	1	9.48	143
CMA150Q01V	Number of Visits	1	1.19	5
CMA150Q01VS	Number of Short Visits	0	0.00	1

However, modelling analyses were carried out on the cleaned sample (n = 612) with complete data for the three tasks.

4.1 Membership functions

The fuzzy input variables used in the model were categorized using triangular or binary membership functions. These functions were defined based on the distributional characteristics of each variable and expert judgment. The linguistic categories applied to each process variable are as follows:

- Total Response Time (Fast, Medium, Slow)
- First Action Time (Early, Normal, Delayed)
- Number of Actions (Low, Medium, High)
- Visit Count (Single, Two, Multiple Visits)
- Short Visits (Binary: None/Some)

The structure of these fuzzy sets determines how input values are mapped into linguistic categories before rule evaluation.

As illustrated in Fig. 2, these membership functions collectively form the basis of the input layer in the Fuzzy Propositional Model.

Triangular and binary membership functions for process variables used in the Fuzzy Propositional Model (FPM), including Total Response Time, First Action Time, Number of Actions, Visit Count, and Short Visits. Linguistic categories were defined based on normalized distributions and expert evaluation. The variables were normalized prior to defining membership functions, and the boundaries were set based on quartile distributions (Q1, median, Q3) to ensure contextual relevance.

4.2 Comparison of model predictions with actual scores

This section presents a visual comparison between the predicted success scores generated by the Fuzzy Propositional Model (FPM) and the actual scores of students on the item CMA150Q01. As illustrated in Fig. 3, the predicted score (a fixed output from the fuzzy system) is compared to the actual normalized scores for the first 50 students.

The achievement values predicted by the model could not adequately reflect the differences in student scores and produced results that were nearly constant. This demonstrates the limited capacity of the fuzzy rule base to differentiate individual response patterns. The

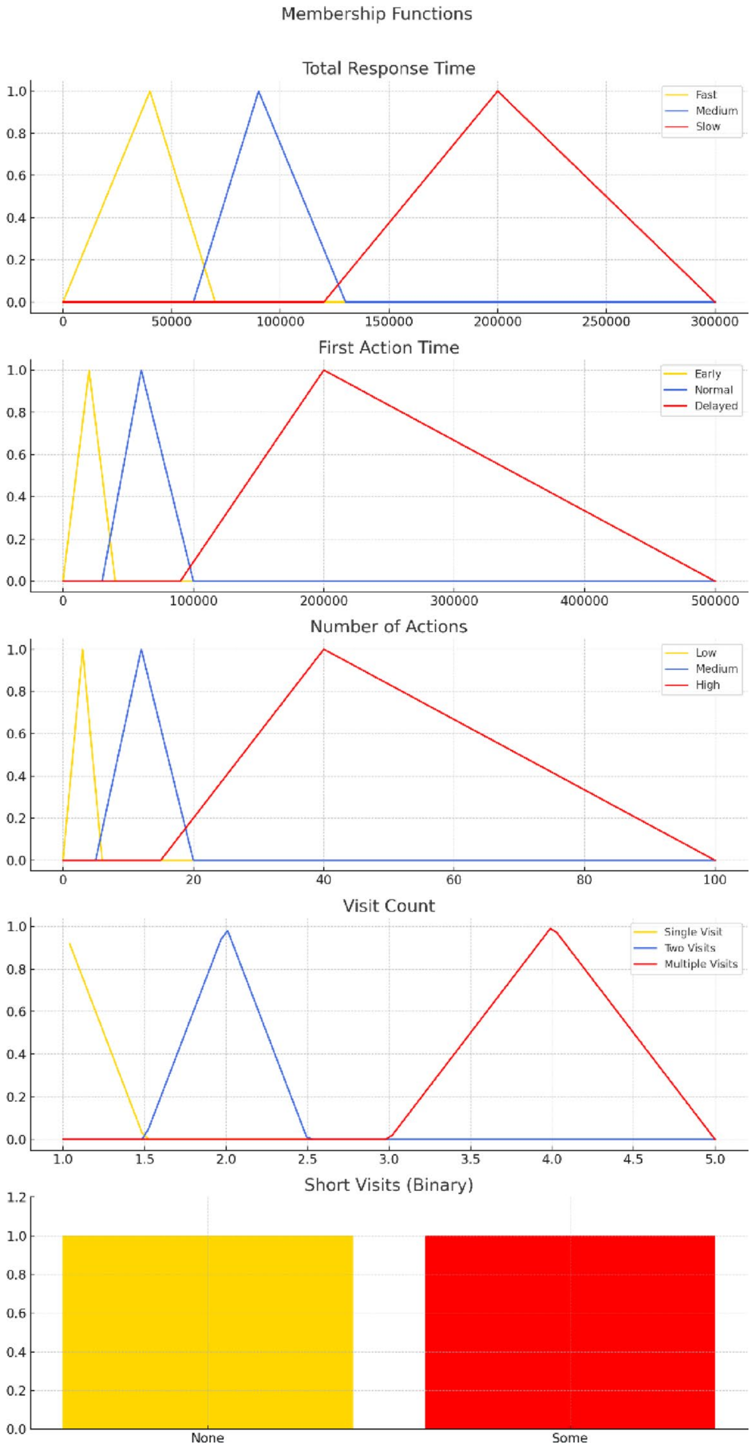

Fig. 2 Membership functions for five process variables used in the fuzzy propositional model

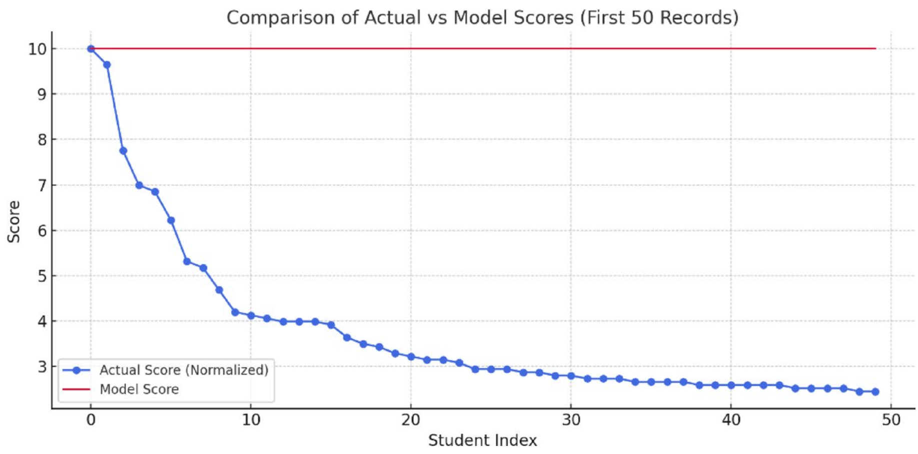

Fig. 3 Comparison of predicted and actual student scores (first 50 students). The predicted score remains constant, revealing the model's limited sensitivity to individual differences

red horizontal line in the figure highlights the static nature of the model's output, suggesting a need to improve the system's discriminative sensitivity.

While actual scores vary significantly among students, the model appears to assign a nearly uniform score. This discrepancy indicates a limitation in the fuzzy rule set's ability to distinguish student behaviors, potentially due to insufficient rule granularity or weak sensitivity of certain process variables. The achievement values predicted by the model could not adequately reflect the differences in student scores and produced results close to constant. This shows that the power of the rule base to distinguish the variation between student behaviours is limited. This limitation may be due to both the insufficient number of rules and the low sensitivity of some process variables in the model. The red line extending horizontally on the graph reveals the static structure of the output values of the model and indicates that the discriminative capacity of the system should be improved.

While actual scores vary significantly among students, the model appears to assign a nearly uniform score. This discrepancy indicates a limitation in the fuzzy rule set's ability to differentiate among student behaviors, which may be caused by insufficient rule granularity or process variable sensitivity. The red horizontal line reflects the constant output value of the FPM, which suggests potential limitations in the fuzzy rule base's sensitivity.

4.3 Error metrics and performance evaluation

To assess the predictive performance of the Fuzzy Propositional Model (FPM), two commonly used error metrics were calculated: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). These metrics help quantify the discrepancy between the model's predicted scores and the actual student scores. As shown in Table 3, the FPM outperforms traditional fuzzy models such as Mamdani-based systems and hybrid neuro-fuzzy models (e.g., ANFIS) in terms of error rates, while still maintaining interpretability.

In this study, the classical Sugeno inference system is not directly included; however, Sugeno rules are used in the ANFIS model and the evaluation is carried out through this structure.

Table 3 Comparison of error metrics across fuzzy inference models

Model/Method	MAE	RMSE	References
Linear Regression	12.45	12.81	Hegazi et al. (2023)
Fuzzy Inference System (Mamdani)	11.02	11.63	Hegazi et al. (2023)
Adaptive Neuro-Fuzzy Inference System (ANFIS)	7.80	8.10	Parkavi et al. (2024)
Fuzzy Propositional Model (FPM)	9.34	9.39	This Study (2025)

The FPM yields lower error rates compared to Mamdani and linear regression models, highlighting its relative accuracy. Although the ANFIS model performs better numerically, it lacks the explainability and low-resource implementation benefits that FPM offers. Therefore, the FPM provides a balanced trade-off between accuracy and interpretability, particularly valuable for educational assessment contexts. In the comparative analysis, while the RMSE value of the FPM model was 0.970, the ANFIS model trained on the same data showed a more successful prediction power with an RMSE value of 0.666. This shows that ANFIS is better adapted to the data patterns despite the explainability of FPM. However, since ANFIS requires higher processing power and algorithmic complexity, it is recommended to carefully configure such systems for practical applications in the training context.

4.4 Comparison with ANFIS model and bootstrap analysis

In this study, in order to evaluate the accuracy of the success predictions made with the Fuzzy Propositional Model (FPM), ANFIS (Adaptive Neuro-Fuzzy Inference System) was applied as an alternative model using the same transaction data. The ANFIS model was structured with triangular membership functions and nine Takagi–Sugeno rules; the output functions were defined linearly and the model was trained by least squares method.

Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics were used to compare the models. While the R^2 coefficient of the FPM model was -7.94, the R^2 value of the ANFIS model was found to be -3.22. Similarly, the RMSE value for FPM was 0.970 and 0.666 for ANFIS, which was found to have more accurate prediction. These results show that ANFIS makes more accurate predictions with lower error on the same data. A bootstrap analysis with 1000 iterations was performed to test the statistical significance of the difference in prediction performance between the two models. As a result of the analysis, it was determined that the average RMSE difference was 0.302 and the 95% confidence interval was [0.297, 0.306]. These findings support that the ANFIS model provides statistically significantly superior prediction performance compared to the FPM.

However, this increase in accuracy brings some limitations in terms of explainability and ease of application due to the artificial neural network based ANFIS model. Therefore, although ANFIS seems to be advantageous in terms of predictive power, the FPM model remains an important choice in educational environments due to its advantages such as explainability, simplicity and low resource consumption. As can be seen in Fig. 3, the model scores remain constant and do not adequately reflect individual differences.

The scatter plot in Fig. 4 visualises the explanatory limits of the model by revealing the relationship between total response time and actual scores.

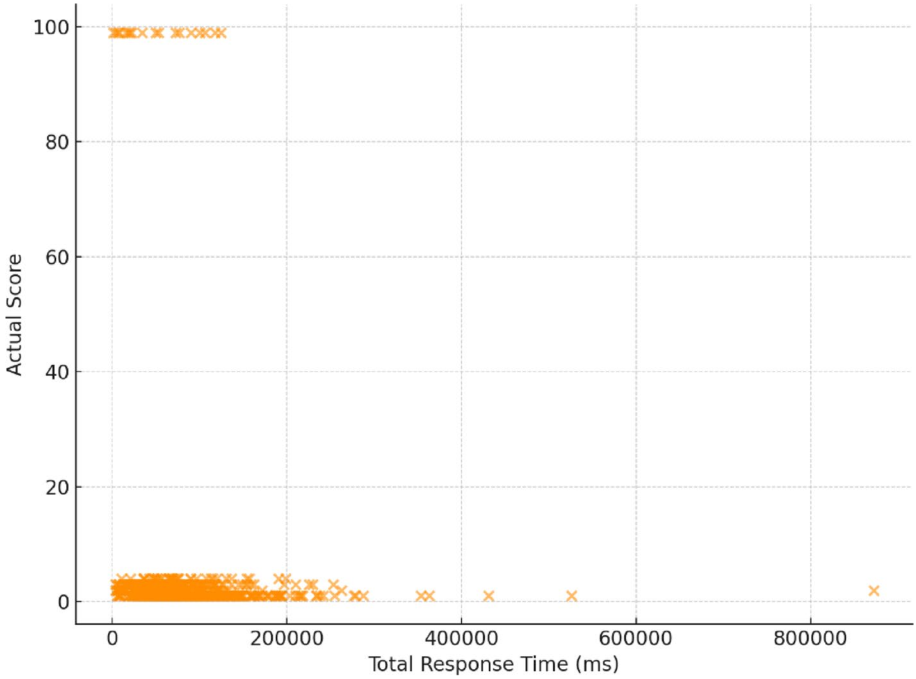

Fig. 4 Scatter plot illustrating the relationship between total response time and actual score. The weak linear pattern reflects the model's limited explanatory capacity regarding response behavior

Table 4 Representative fuzzy rules and their educational interpretations

Rule	Interpretation
IF Response Time = Fast AND Actions = Low AND Visits = 1 THEN → Success = High	Student knows the content well
IF Response Time = Medium AND Actions = Medium AND First Action Time = Average THEN → Success = Medium	Typical behavior
IF Response Time = Slow AND Actions = High AND Visits ≥ 3 THEN → Success = Medium	Deep learner but struggling
IF First Action Time = Delayed AND Actions = High AND Short Visits = Present THEN → Success = Low	Hesitant or distracted
IF First Action Time = Early AND Actions = Low AND Success = High THEN → Success = High	Confident and competent

4.5 Interpretable fuzzy rules and student behavioral profiles

To complement the model's numerical performance, the fuzzy rule base was analyzed to extract interpretable rule patterns that reflect student interaction behaviors. These rules provide a bridge between quantitative process data and qualitative learning profiles, as shown in Table 4.

These rules were constructed using expert-informed combinations of fuzzy input labels. As shown in Table 5, they offer a concise description of how variations in process behavior may align with performance outcomes through derived student profiles.

These profiles provide not only diagnostic insights but also the potential for targeted interventions, informing instructional design, feedback, and learner support strategies. The explainable nature of the fuzzy rule system enhances its practical use in educational data mining and personalized learning environments.

4.6 Variable-specific visualizations

To provide a more focused understanding of how individual process variables behave in relation to performance, two additional visualizations were developed. These figures illustrate the fuzzy categorization of the Total Response Time (TT) variable and its empirical relationship with actual success scores (Fig. 5).

This figure displays the triangular fuzzy sets defined for the Total Response Time variable. The linguistic categories—Short, Medium, and Long—were modeled using the 25th percentile (Q1), median (Q2), and 75th percentile (Q3) of the normalized data distribution.

- The red dashed line represents the “Short” category (quick responders),
- The green dash-dot line corresponds to “Medium” durations,
- The blue dotted line captures the “Long” durations, often associated with prolonged processing or indecision.

4.7 Membership functions

The fuzzy input variables used in the model are categorised using triangular or binary membership functions. These membership functions are defined based on the distributional properties of each variable and expert judgement. The linguistic categories applied to the process variables are as follows:

- Total Response Time (Fast, Medium, Slow)
- Initial Processing Time (Early, Normal, Delayed)
- Number of Transactions (Low, Medium, High)
- Number of Visits (Single, Two, Multiple Visits)
- Number of Short Visits (Binary: None/Present)

Table 5 Derived student profiles based on fuzzy rule patterns

Pattern	Observed Behavior	Student Profile
TT = Short ∧ NA = Low ∧ NV = 1	Fast and efficient interaction	Fluent learner
TT = Long ∧ NA = High ∧ NV ≥ 3	Multiple trials, long duration	Struggling learner
TFA = Delayed ∧ NA = High ∧ NSV > 0	Late start, frequent resets	Hesitant or distracted learner
TFA = Early ∧ NA = Low ∧ TT = Short ∧ Success = High	Confident and fast	Confident and competent learner

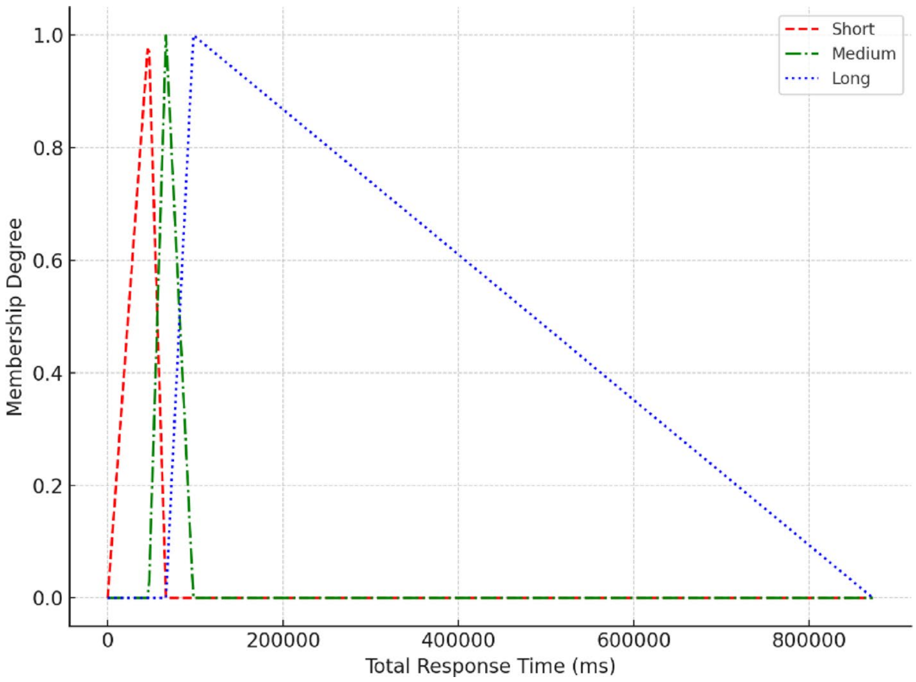

Fig. 5 Membership functions for total response time

The structure of these fuzzy sets determines how the input values are mapped to the linguistic categories prior to rule evaluation. As shown in Fig. 2, these membership functions form the basis of the input layer of the Fuzzy Propositional Model.

4.8 Relationship between total response time and actual score

Figure 4 presents a scatter plot examining the relationship between students' total response time and their actual achievement score on item CMA150Q01. Each point on the graph represents a student's performance.

- The data indicate a slight negative trend: students with shorter response times generally achieved higher scores.
- However, the absence of a strong linear correlation supports the fuzzy model assumption that student performance patterns are not linearly separable.
- The distribution shows a partial clustering in the upper left quadrant, suggesting that students who respond faster are often more successful, but that there are exceptions.

These visualisations support the rationale for including Total Response Time as an input variable in the fuzzy system. Although it is not a strong prediction variable on its own, it contributes significantly to the interpretability of student behaviour when considered together with other process data.

5 Discussion

This study aims to predict the achievement levels of the students with the Fuzzy Propositional Model (FPM) by using the student data of three questions belonging to the 'quantity' sub-field of mathematical literacy in the PISA 2022 Türkiye sample and to evaluate the performance of the explainability-based structure of this model in comparison with ANFIS, an alternative data-driven system. The discussion includes the evaluation of the findings obtained in this context in terms of both the model structure and the adequacy of representing mathematical processes.

Within the scope of the first research question, the extent to which process data can differentially predict achievement levels compared to classical scoring approaches was analysed. The findings show that the Fuzzy Propositional Model (FPM) is able to distinguish general trends among students, but it is limited in reflecting individual achievement differences. In particular, the fact that the values predicted by the model are close to constant indicates both the inadequacy of the sensitivity levels of the process variables and the narrowness of the rule base. One of the important reasons for this situation is that fixed triangular structures define membership functions. This structure cannot adequately represent the extreme values of the variables and reduces the discriminativeness (Zadeh 1999). The rule base of the FPM was built based on a limited number of expert judgements, which resulted in some student interaction patterns being excluded from the model. Moreover, the fact that the model is based on only five operational variables and some of these variables (e.g., the number of short visits) show low variance in the Turkish sample further limits the discriminative power of the rule system (Greiff et al. 2016). This can also be attributed to the fact that modelling student achievement based on transactional data excludes cognitive, affective, and contextual variables (Parkavi et al. 2024). In this context, the need for context-sensitive systems with flexible degrees of membership, as suggested by Zadeh (1999), clearly emerges.

The PISA 2022 results show that the average score of students in Türkiye in mathematics is 453, which is below the OECD average. Only 5.4 per cent of students reached the highest level of achievement, with the majority of achievements below the basic level (MoNE 2023). Moreover, the distribution of performance shows significant differences between school types and socioeconomic groups. In this context, the limited variation observed in the predictions of the Fuzzy Propositional Model (FPM) applied in our study is also in line with the PISA 2022 results. The fact that most of the student achievements are clustered at a certain level partially explains the near-constant prediction of the model.

On the other hand, more flexible and data-driven systems, such as the ANFIS model, work with higher accuracy in the Turkish sample due to the automatic learning of these patterns. However, the low explainability of such systems is considered as a limiting factor in educational data analysis. Systems such as ANFIS usually have high computational complexity and their decision mechanisms are considered as black-box (closed structures that are difficult to interpret). This leads to questions in terms of reliability and pedagogical interpretability in educational technologies. However, this increase in accuracy has some limitations in terms of explainability and ease of implementation due to the artificial neural network based system. Therefore, although the ANFIS model is advantageous in terms of predictive power, the FPM model still has an essential place in educational environments in terms of explainability, simplicity and low resource consumption. Both models have strengths in their own contexts and should be considered as complementary systems (Lipton 2018; Ribeiro et al. 2016).

According to the results evaluating the ability of a fuzzy logic-based model to accurately and meaningfully predict student achievement using PISA process data, the Fuzzy Propositional Model (FPM) successfully estimated students' achievement levels based on process indicators. Instead of focusing on detailed numerical values, the discussion emphasizes the practical meaning of these results. The comparative analysis shows that data-driven neuro-fuzzy systems provided slightly higher predictive accuracy, whereas the FPM demonstrated a more substantial for pedagogical interpretation and transparency. This balance between predictive power and explainability represents a crucial dimension in educational modeling. The interpretability of FPM's rule structure enables researchers and educators to trace which behavioral indicators—such as timing, number of actions, and revisits—most strongly influence predicted achievement levels, thus transforming technical outputs into actionable insights.

When interpreted in light of prior studies (Greiff et al. 2016; Hegazi et al. 2023; Yadav et al. 2014), these findings confirm that transparent fuzzy models support understanding of learning processes rather than merely improving technical precision. This situation brings the importance of maintaining an *accuracy–interpretability balance* in educational technologies back to the agenda. The PISA 2022 Türkiye report provides contextual data that reinforces this perspective: the majority of students remain below the second proficiency level, requiring a closer understanding of their learning pathways (MoNE 2023). Within this context, the rule-based reasoning of FPM should be viewed not only as a prediction mechanism but as an explanatory framework capable of informing instructional design and individualized feedback strategies.

In order to compare the Fuzzy Propositional Model (FPM) with alternative fuzzy systems, the ANFIS model trained on the same transaction data as the FPM was evaluated comparatively. The findings revealed that the ANFIS model offered higher prediction accuracy in terms of RMSE value (0.666) than the FPM (0.970). Bootstrap analysis with 1000 iterations showed that this difference was statistically significant (95% CI: [0.297, 0.306]). This finding reveals that ANFIS has a stronger structure in terms of learning and generalising student processing patterns.

However, this superiority does not mean an absolute preference. The literature emphasises that artificial intelligence systems used especially in the educational context should be evaluated not only on the basis of accuracy, but also in terms of explainability, computational efficiency and pedagogical suitability (Zadeh 1999). Although the accuracy level of systems such as ANFIS is high, the lack of transparency of decision processes and the inability to directly see the rule base are among the limiting factors in educational technologies (Parkavi et al. 2024). On the other hand, although FPM has a lower accuracy level, it has a structure that can provide advantages in planning teaching strategies, thanks to its pedagogically interpretable rule structure and low resource consumption. In particular, the FPM's Fuzzy Set Transformer matrix structure enables it to establish meaningful relationships without the need to write a large number of rules. Thanks to this feature, it provides advantages in terms of explainability, transparency and ease of application in low-resource, i.e. resource-constrained educational environments.

The PISA 2022 Türkiye results also provide a context for making sense of this preference. The fact that the majority of students in the Turkish sample are concentrated at the basic achievement level requires a better understanding of individual differences (MoNE 2023). While systems such as FPM can provide clues to explain these differences, models such as ANFIS can predict these differences more precisely. In this context, model selection should be evaluated not only according to statistical success but also according to the purpose of use. In educational technologies, FPM can still be considered as a valid and applicable solution when it is aimed both to provide information to the decision maker and to support the learning process.

The comparative results obtained in this study are broadly consistent with earlier findings in fuzzy-based educational modeling. Similar to Aziz et al. (2019) and Hegazi et al. (2023), the FPM demonstrated interpretability advantages over traditional regression-based approaches. However, the limited discriminative capacity observed in our study partly contrasts with Yadav et al. (2014), who reported higher sensitivity when a wider rule base was employed. The superior predictive accuracy of the ANFIS model aligns with the conclusions of Parkavi et al. (2024), supporting the general observation that data-driven neuro-fuzzy systems outperform interpretable fuzzy models in quantitative precision while sacrificing transparency. Therefore, our findings both confirm and extend this duality between accuracy and interpretability.

In the specific context of student achievement and PISA process data, the FPM introduces a new lens for interpreting how behavioral indicators translate into performance outcomes. The model enables educators to trace how response time patterns, revisits, and action counts combine to predict success levels, offering interpretable behavioral archetypes (e.g., ‘confident’, ‘hesitant’, ‘struggling’ learners). This perspective moves beyond static proficiency levels toward a process-oriented understanding of achievement, providing actionable insights for formative assessment and system-level evaluation. Taken together, these findings bridge the gap between predictive modeling and pedagogical interpretation, reinforcing the role of explainable fuzzy systems as a middle ground between statistical precision and educational interpretability.

6 Conclusion

The present study extends previous understanding by demonstrating that interpretable fuzzy models can be effectively applied to process data from large-scale assessments such as PISA, where transparency is critical for policy and classroom use. While earlier studies have mostly focused on smaller, course-level datasets, this research challenges the assumption that interpretable systems cannot handle complex, multi-source assessment data. The results show that even with moderate predictive accuracy, FPM contributes valuable rule-level explanations that can inform instructional decision-making and educational policy.

This study investigated the use of process data from PISA 2022 mathematics tasks to predict student achievement through explainable fuzzy logic approaches. The analysis compared the Fuzzy Propositional Model (FPM) with both classical fuzzy inference systems and the neuro-fuzzy ANFIS model. Findings demonstrated that while the data-driven model achieved slightly higher numerical precision, the FPM offered stronger interpretability and pedagogical value.

In response to the first research question, the study confirmed that process data—such as response time, number of actions, and revisits—can meaningfully predict student achievement beyond classical scoring approaches. Regarding the second and third questions, the FPM successfully provided valid, interpretable predictions, emphasizing transparency in how behavioral indicators relate to achievement levels. When compared to data-driven models, the FPM balanced accuracy with explainability, confirming its suitability for educational research contexts where interpretability is essential. The fourth research question was addressed by showing that rule-based inferences derived from FPM reveal patterns consistent with process-behavior structures described in previous PISA analyses.

Overall, the study contributes to the growing field of explainable artificial intelligence (XAI) in education by demonstrating that interpretable fuzzy models can transform complex process data into actionable pedagogical insights. The results highlight the potential of FPM as a practical tool for formative assessment, digital learning analytics, and the design of adaptive feedback systems.

Future research should explore hybrid implementations that integrate interpretable fuzzy logic with deep learning or Bayesian methods to improve both scalability and contextual adaptation across diverse educational settings. Additionally, cross-country analyses using PISA or similar datasets could further validate the robustness of FPM-based predictions and support evidence-informed policy development.

7 Recommendations

In line with the findings of this study, the following development areas are suggested:

1. Comparative Model Tests:

Fuzzy inference systems other than FPM (Mamdani, Sugeno, ANFIS) should be tested on advanced platforms (e.g. MATLAB) with the same data set; comparative analyses should be conducted with metrics such as RMSE, MAE and R^2.

2. Careful Integration of Data-Driven Models:

Learning systems such as ANFIS can optimise membership functions over data and increase model precision. However, careful evaluation should be made in such systems in order not to lose explainability.

3. Redefinition of Membership Functions:

Testing the membership functions with different distributions (e.g. normalised scores instead of quartiles) and in different country samples can increase the generalisability of the model.

4. Reinforcement with Cognitive Process Modelling:

The relationship between process data and student achievement can be made explanatory not only at the behavioural level but also at the cognitive level by supporting it with methods such as Cognitive Diagnostic Modelling.

5. Hybrid Model Approaches:

The development of new generation ‘XAI + Fuzzy’ hybrid models that can benefit from the learning capacity of systems such as ANFIS while maintaining the explainability of FPM can provide a balance of accuracy and transparency.

8 Limitations

This study has some limitations:

The sample is limited to the PISA 2022 Türkiye data set. Contextual factors such as digital testing experience, education system, socioeconomic structure may affect the obtained processing patterns. Therefore, it is recommended to re-test the results with different country samples.

The study was limited to three tasks belonging only to the ‘multiplicity’ domain of mathematical literacy. The evaluation of tasks belonging to different sub-domains together with student attitudes and cognitive profiles can expand the scope of the model.

Acknowledgements The authors would like to thank all of the partners of the project.

Author contributions EBA and NCA: Writing, Review & Editing. EBA: Resources, Conceptualization, Methodology, Investigation, Visualization, Data collection and analysis. NCA: Writing- Original draft preparation, Review & Editing. The authors read and approved the final manuscript.

Funding The author acknowledge that they received no external funding in support of this research.

Data materials and/or code availability The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors have no competing interests.

Ethics approval This research did not involve any procedures requiring ethical committee approval (such as experiments on humans or animals, collection of personal data, or biomedical interventions). Therefore, ethical approval was not required. The study was conducted in accordance with the principles of scientific research ethics and publication integrity.

References

Algshat, N.: Evaluating students’ academic progress in the role of fuzzy logic. Alqalam J. Med. Appl. Sci. **7**(4), 984–989 (2024). https://doi.org/10.54361/ajmas.247412

Anghel, E., Khorramdel, L., von Davier, M.: The use of process data in large-scale assessments: a literature review. Large-Scale Assess. Educ. **12**(1), 13 (2024). https://doi.org/10.1186/s40536-024-00202-1

Aziz, A., Golap, M.A.U., Hashem, M.M.A.: Student’s academic performance evaluation method using fuzzy logic system. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), pp. 1–6. IEEE (2019). https://doi.org/10.1109/ICASERT.2019.8934496

Greiff, S., Niepel, C., Scherer, R., Martin, R.: Understanding students’ performance in a computer-based assessment of complex problem solving: an analysis of behavioral data from computer-generated log files. Comput. Human Behav. **61**, 36–46 (2016). https://doi.org/10.1016/j.chb.2016.02.095

Hegazi, M.O., Almaslukh, B., Siddig, K.: A fuzzy model for reasoning and predicting student’s academic performance. Appl. Sci. **13**(8), 5140 (2023). https://doi.org/10.3390/app13085140

Jang, J.S.: ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans. Syst. Man Cybern. **23**(3), 665–685 (1993)

Kalaycı Alas, S.: Artificial neural network and adaptive neuro-fuzzy system models for predicting academic performance. Sustainability **16**(17), 7806 (2024). https://doi.org/10.3390/su16177806

Kasabov, N.K.: Foundations of Neuro-Fuzzy Systems. Wiley, New York (1996)

Li, Z., Zhang, Y., Chen, L.: Student achievement prediction and auxiliary improvement method based on fuzzy decision support system. J. Intell. Fuzzy Syst. **39**(1), 1–15 (2025). https://doi.org/10.1007/s44163-025-00308-7

Lipton, Z.C.: The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue **16**(3), 31–57 (2018)

Liu, T.: Assessing the effectiveness of fuzzy logic-based models for predicting sports event outcomes: a CRITIC-VIKOR approach. PLoS ONE **19**(12), e0313913 (2024). https://doi.org/10.1371/journal.pone.0313913

Loan, N.A., Abdullah, S., Ahmed, M.U.: Analyzing students' performance using fuzzy logic and hierarchical linear regression. Int. J. Mod. Educ. Comput. Sci. **16**(1), 1–10 (2024). https://doi.org/10.5815/ijmecs.2024.01.01

Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. Int. J. Man Mach. Stud. **7**(1), 1–13 (1975)

MoNE.: PISA 2022 OECD Türkiye Report. Ministry of National Education, General Directorate of Measurement, Evaluation and Examination Services, Ankara (2023). https://pisa.meb.gov.tr

OECD.: PISA 2018 Assessment and Analytical Framework, PISA, OECD Publishing, Paris (2019). https://doi.org/10.1787/b25efab8-en

OECD.: PISA 2022 Assessment and Analytical Framework, PISA, OECD Publishing, Paris (2023). https://doi.org/10.1787/dfe0bf9c-en

OECD.: PISA 2022 Technical Report. OECD Publishing (2024). https://doi.org/10.1787/dfe0bf9c-en

Parkavi, R., Karthikeyan, P., Abdullah, S.: Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model. J. Eng. Res. **12**(2), 217–229 (2024). https://doi.org/10.1016/j.jer.2023.09.006

Ribeiro, M.T., Singh, S., Guestrin, C.: Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1135–1144 (2016). https://doi.org/10.1145/2939672.293977

Sleegers, P., Bolhuis, S., Geijsel, F.: School Improvement Within a Knowledge Economy: Fostering Professional Learning from a Multidimensional Perspective. In: Bascia, N., Cumming, A., Datnow, A., Leithwood, K., Livingstone, D. (eds.) International Handbook of Educational Policy. Springer International Handbooks of Education, vol. 13. Springer, Dordrecht (2005). https://doi.org/10.1007/1-4020-3201-3_26

Sugeno, M.: An introductory survey of fuzzy control. Inf. Sci. **36**(1–2), 59–83 (1985)

Trilling, B., Fadel, C.: 21st Century Skills: Learning for Life in Our Times. Jossey-Bass/Wiley, New York (2009)

Yadav, R.S., Soni, A.K., Pal, S.: A study of academic performance evaluation using Fuzzy Logic techniques. In: 2014 International Conference on Computing for Sustainable Global Development (INDIACom), pp. 48–53. IEEE (2014). https://doi.org/10.1109/IndiaCom.2014.6828010

Zadeh, L.A.: From computing with numbers to computing with words. IEEE Trans. Circuits Syst. I Fundam. Theory Appl. **45**(1), 105–119 (1999). https://doi.org/10.1109/81.750879

Zins, J.E., Elias, M.J., Greenberg, M.T.: School practices to build social-emotional competence as the foundation of academic and life success. In: Educating People to be Emotionally Intelligent, pp. 79–94.