

Early GPA Prediction: A Comparison of Fuzzy Inference, Regression, and Decision Trees

¹Husnul Hakim*, ²Lionov Wiratma, ³Mariskha Tri Adithia

^{1,2,3}Center for Data Science and Artificial Intelligence Systems, Program Studi Informatika, Fakultas Sains, Universitas Katolik Parahyangan, Indonesia

^{1,2,3}Jl. Ciumbuleuit Nomor 94 Bandung, Jawa Barat, Indonesia

*e-mail: husnulhakim@unpar.ac.id

(received: 13 July 2026, revised: 25 September 2026, accepted: 26 September 2026)

Abstract

Student dropout remains a persistent challenge in higher education, making the early identification of at-risk students essential for timely intervention. In the Informatics study program examined in this research, the first-stage academic evaluation is conducted at the end of the fourth semester, creating a need for predictions that can be made well before this milestone. This study compares a Mamdani Fuzzy Inference System (FIS), linear regression, and decision tree in predicting fourth-semester GPA based on students' grades in three foundational first-semester courses: Basic Mathematics, Discrete Mathematics, and Introduction to Programming. The dataset consists of the academic records of 254 students across four cohorts. The Mamdani FIS rule base was constructed in a data-driven manner, while linear regression and decision tree served as comparative baselines. All three models were evaluated using 5-fold cross-validation, with Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) as evaluation metrics. Linear regression achieved higher accuracy and greater stability, with an average RMSE of 0.3766 and MAE of 0.2874, compared to the Mamdani FIS which obtained an RMSE of 0.5052 and MAE of 0.3660. The decision tree (RMSE = 0.4776, MAE = 0.3558) ranked between them and was the most stable. However, the Mamdani FIS produced an interpretable linguistic rule base, of which the majority remained consistent across folds, offering transparency that supports decision-making in identifying at-risk students. These findings highlight a trade-off between predictive accuracy and interpretability, indicating that the choice of method depends on whether the priority is prediction accuracy or model transparency.

Keywords: academic performance, decision tree, fuzzy inference system, GPA prediction, linear regression, mamdani

1 Introduction

The issue of student dropout in higher education has been widely studied across countries, encompassing academic, socio-economic, and institutional contexts, and affecting both individuals and institutions [1]. This challenge is also evident in Indonesia. In 2023, 315,520 students out of 8,316,919 experienced dropout, or approximately 3.79% [2]. In 2024, the number decreased to 201,889 out of 4,580,141 students, yet the percentage rose to 4.41% [3]. Although the absolute figure declined, the increase in proportion underscores that dropout remains a pressing issue requiring serious attention.

To address this persistent challenge, researchers have increasingly turned to data-driven approaches. The availability of academic records in higher education institutions provides opportunities for predictive analysis [4]. Within this context, educational data mining plays a crucial role in identifying patterns and designing strategies to enhance educational quality [5]. Early identification of at-risk students allows institutions to intervene before academic failure occurs.

To prevent early academic failure, Study Program X implements three stages of evaluation: Stage 1, Stage 2, and Stage 3. Stage 1 evaluation is conducted at the end of the fourth semester, requiring students to have completed at least 30 credits with a minimum GPA of 2.0. The purpose of this evaluation is to provide students with the opportunity to continue their studies at a more suitable institution from an early stage, thereby avoiding unnecessary loss of time and financial resources.

This study aims to compare the performance of a Fuzzy Inference System (FIS), linear regression, and decision tree in predicting students' cumulative grade point average (GPA) at the end

of the fourth semester based on their grades in three foundational courses, namely Basic Mathematics, Discrete Mathematics, and Introduction to Programming. These courses were selected because they provide the foundational knowledge required for many subsequent courses in the Informatics curriculum. The fourth-semester GPA was chosen as the prediction target because it serves as the first-stage academic evaluation in the Study Program X, Indonesia. The dataset consists of the academic records of 254 students from four different cohorts. To compare the performance of the three models, a 5-fold cross-validation approach was employed.

2 Literature Review

The use of data to build predictive models of academic performance has rapidly advanced through educational data mining (EDM). Prior study [5] emphasized that the strength of EDM lies in its ability to identify patterns relevant for prediction. Research trends show that EDM frequently employs machine learning as a primary technique; however, the quality of prediction outcomes is strongly influenced by the timing of prediction. Several studies confirmed that first-year data can already provide valuable insights [6] and [7], while the study in [8] pushed prediction even earlier, only four weeks after the start of classes. This analysis highlights a consistent direction toward earlier prediction, yet its alignment with institutional evaluation policies remains underexplored. Thus, although EDM has demonstrated significant potential in academic prediction, there is still room to connect these methods with formal evaluation mechanisms to enhance practical relevance.

Beyond timing, a major challenge in academic prediction is data uncertainty. The uncertain nature of academic data has been highlighted in [9], while studies in [10], [11], [12] demonstrated that fuzzy logic is effective in addressing this issue. Furthermore, the integration of fuzzy logic with hierarchical regression illustrates the potential of hybrid approaches [13]. Carrasco-Garrido et. al noted that among FIS types, Mamdani is well-suited to academic data due to its interpretability [14]. This analysis suggests that fuzzy logic remains consistently relevant in academic contexts, though its application using foundational course grades as inputs is still limited.

On the other hand, regression continues to serve as a strong baseline method due to its simplicity and interpretability. Falát & Piscová employed regression to compare the performance of random forest models [15], while Alnomay et al. found that regression outperformed several other machine learning approaches [16]. Falát & Piscová also included decision tree in their comparison [15], while Al-Barrak & Al-Razgan applied the J48 decision tree algorithm to predict final GPA from mandatory course grades [17]. These findings reaffirm that linear regression, despite its simplicity, remains a relevant benchmark in academic prediction studies. Consequently, comparing regression with fuzzy logic is crucial to evaluate the trade-off between simplicity and the ability to handle uncertainty.

The success of predictive models is also influenced by the choice of input variables. Studies by [9], [11], [12], [13] emphasized generic variables such as attendance and motivation, whereas [18] employed domain-specific variables, namely course grades, in their ANFIS-based model. Domain-specific variables in regression was also applied by [15]. This analysis indicates that both generic and domain-specific variables can be utilized, but academic grades as domain-specific variables are more accessible and therefore more practical for this study.

From prior literature, two research gaps can be identified. First, while ANFIS has utilized domain-specific course grades, the application of foundational courses in Mamdani FIS remains underexplored. Second, few studies have aligned prediction timing with Stage 1 institutional evaluation policies, despite its practical importance. Based on these gaps, this study focuses on applying Mamdani FIS with foundational course grades as inputs to predict fourth-semester GPA, and comparing its performance with linear regression and decision tree using cross-validation.

3 Research Method

This study utilized data from three fundamental courses in Study Program X, involving 254 students across four cohorts, namely the 2018 to 2021 cohorts. The three input variables employed in the predictive model were the grades from Basic Mathematics, Discrete Mathematics, and Introduction to Programming. Each of these input variables ranged from 0 to 100. Based on these inputs, a model was constructed to predict the GPA at the end of the fourth semester, which ranged from 0 to 4. Table 1 presents a summary of the input and output variables along with their ranges.

<http://sistemasi.ftik.unisi.ac.id>

The variables were selected because these three courses are foundational courses positioned in the early semesters of Study Program X and serve as prerequisites for advanced courses. Discrete Mathematics is a prerequisite for the Discrete Structures course, while Introduction to Programming is a prerequisite for the Algorithms and Data Structures course. Although Basic Mathematics is not a formal prerequisite, mathematical skills are required in various advanced courses such as Statistics for Computing or Matrix and Vector Spaces. Another reason for this selection is related to availability and timing of prediction. Choosing these fundamental courses aligns with the objective of producing early predictions before the end of the fourth semester, which marks the first evaluation stage in Study Program X. Furthermore, these three courses represent three essential competencies needed to study Informatics: Basic Mathematics for numerical and analytical skills, Discrete Mathematics for logical and abstract reasoning, and Introduction to Programming for the ability to build problem-solving solutions in the form of software. These reasons are supported by evidence that early-semester course grades can be used as predictors [18], [19].

Table 1 Input and output variables

Variable	Range	Type
Basic Math Grade (MATDAS)	0 – 100	Input
Discrete Math Grade (MATDIS)	0 – 100	Input
Introduction to Programming Grade (DASPRO)	0 – 100	Input
GPA at the end of 4th Semester	0 – 4	Output

The three input variables in Table 1 were processed to obtain the predicted GPA at the end of the fourth semester, which served as the output variable, using the three models. The first model applied was the Mamdani-type Fuzzy Inference System, while linear regression and decision tree were used as comparison models. The outputs of the three models were then compared. Figure 1 illustrates the general workflow of this study.

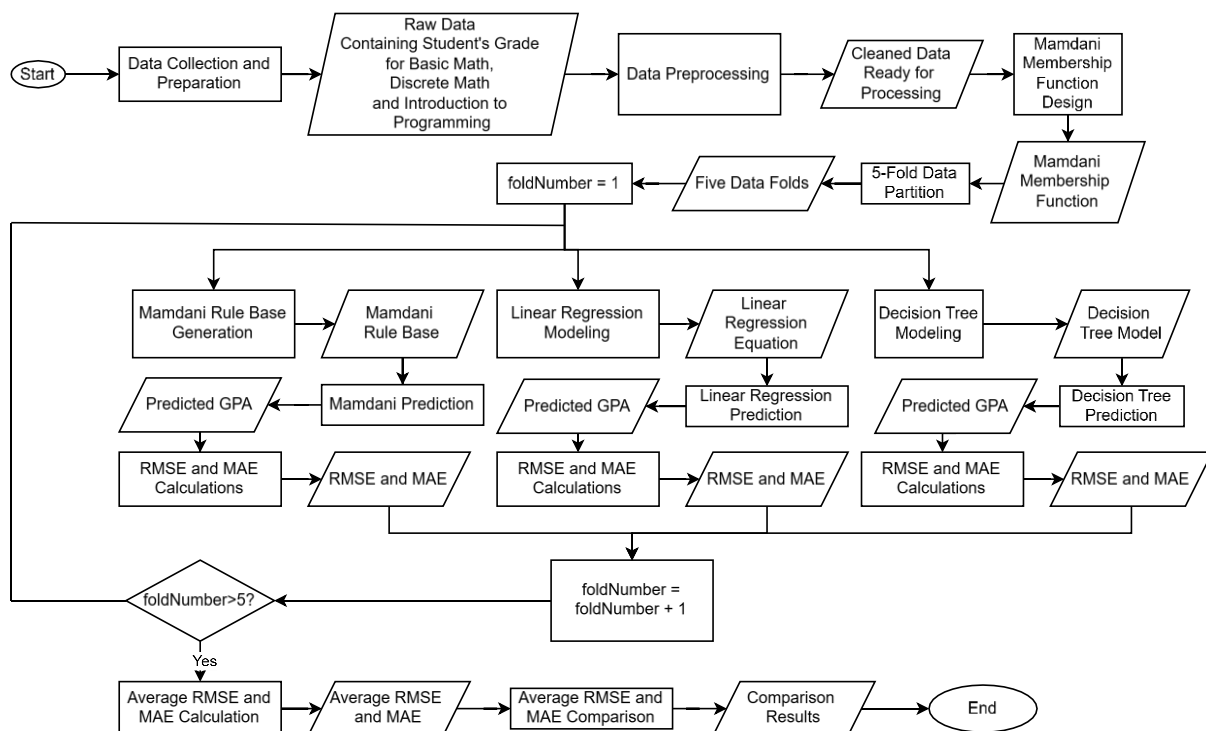


Figure 1 Research method

As shown in Figure 1, after the data were collected, the next stage was data preprocessing. The purpose of this process was to remove values that could not be processed in subsequent stages. From the initial 284 student records, 30 were removed because those students had withdrawn, resulting in

zero grades in one or more of the courses used as input variables. After this process, 254 student records remained and were ready for further analysis.

The following stage involved designing membership functions for each input and output variable listed in Table 1. As illustrated in Figure 2, each input value was divided into three different fuzzy sets: low, medium, and high. The boundary points of each set were determined based on the academic grading system commonly applied in Indonesia. Scores below 50 were categorized as low because they fall below the passing threshold (grade D or E). The range of 50 to 65 represented a transition toward medium, covering grade C. The peak of the medium fuzzy set was 65, representing the zone from grade C to B. Meanwhile, scores of 80 and above were categorized as fully high, representing grade A. This approach was chosen so that the fuzzy set boundaries reflect meaningful academic interpretations rather than uniform numerical divisions. It can also be seen that trapezoidal membership functions were used for the low and high sets, while triangular membership functions were applied for the medium fuzzy set.

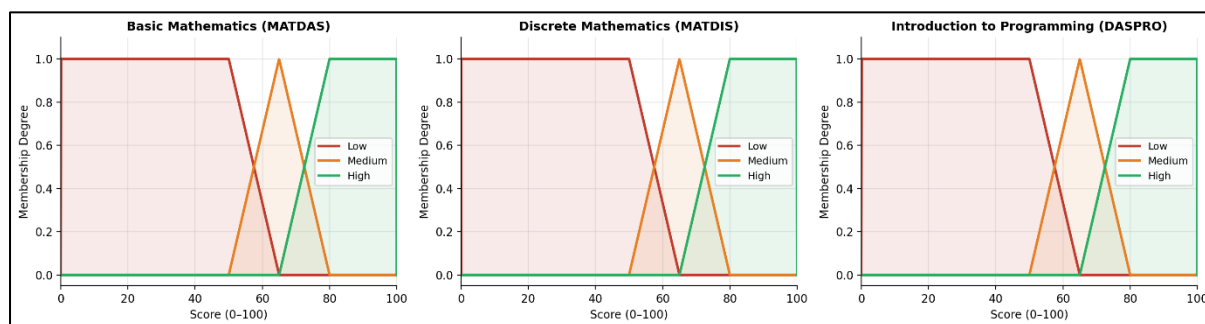


Figure 2 Membership function for input variables

For the output variable, five membership sets were used: very low, low, medium, high, and very high, as shown in Figure 3. This division was chosen to capture GPA variations in greater detail, particularly at the extreme ranges (very low and very high). As illustrated in Figure 3, the very low fuzzy set employed a trapezoidal membership function. The low, medium, and high fuzzy sets used triangular membership functions. Finally, the very high fuzzy set applied a linear membership function.

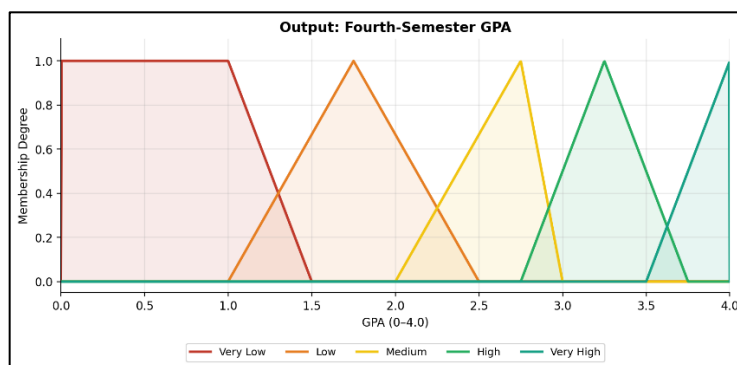


Figure 3 Membership function for output variable

After the membership functions for the input and output variables were designed, the next stage was data partitioning into 5-fold partitions. From the 254 cleaned records, the data order was randomized so that each fold contained a diverse set of students from different cohorts, making it more representative of the overall population. For each fold, the process continued with model construction, namely the Mamdani FIS model, linear regression, and decision tree models.

In the Mamdani FIS model, rule construction was performed in a data-driven manner using the training data in each fold. Since there were three input variables, each divided into three fuzzy sets (low, medium, and high), there were 27 possible rule combinations. These rules took the form of IF (combination of fuzzy input sets) THEN (fuzzy output set for GPA). The consequent part of the rule was not determined by experts but was objectively derived from the data. This was done by

calculating the average actual GPA of students within the group corresponding to the input label combination, as shown in Equation (1). The average GPA value was then mapped to the fuzzy set for GPA as the output.

$$c_k = \frac{1}{|S_k|} \sum_{i \in S_k} y_i \quad (1)$$

Explanation:

- c_k = consequent value (average GPA) for the k -th rule
- S_k = set of students belonging to the k -th input combination
- $|S_k|$ = number of students in k -th combination
- y_i = actual GPA of i -th student
- k = rule index, with $k = 1, 2, \dots, 27$

If a particular input combination group could not be assigned a GPA value because no students were present in that group, interpolation was applied. Each fuzzy set of the input variables was assigned a score: low = 1, medium = 2, and high = 3. For each combination without GPA data, the three scores of the fuzzy input sets were averaged using Equation (2). This average value was then mapped into the output variable's fuzzy set, referring to Table 2.

Table 2 Interpolation mapping for rule combinations without data

c_k	Fuzzy Set
$1.00 \leq c_k < 1.50$	Very Low
$1.50 \leq c_k < 2.00$	Low
$2.00 \leq c_k < 2.50$	Medium
$2.50 \leq c_k < 3$	High
$c_k = 3.0$	Very High

This approach was based on the assumption that the three course grades used as input variables were positively correlated with GPA. Thus, higher input combinations generally produced higher GPA values. In this way, all 27 rules could be fully populated, ensuring that the system was able to provide outputs even when facing input combinations not represented in the training data. Table 3 provides an example of a subset of the generated rules.

$$c_k = f\left(\frac{s_1 + s_2 + s_3}{3}\right), \text{ if } |S_k| = 0 \quad (2)$$

Explanation:

- s_1, s_2, s_3 = scores of fuzzy sets for input variables
- $f(\cdot)$ = mapping function from average scores to output variable fuzzy sets

Table 3 Sample of rule set

Rule#	MATDAS	MATDIS	DASPRO	GPA
1	Low	Low	Low	Low
2	Medium	Low	Low	Low
...	...	...	...	...
27	High	High	High	Very High

After the rules were constructed, the GPA prediction process could be carried out. The first step in this process was fuzzification. Fuzzification is the process of transforming the three input variable values—MATDAS, MATDIS, and DASPRO—into fuzzy membership degrees using the membership functions shown in Figure 2. As a result, each value had membership degrees in the low, medium, and high sets within the range [0, 1]. Unlike rule construction, which only considered the dominant membership set, prediction required the next step of rule aggregation. To enable aggregation, all rules were retained and their firing strengths were calculated. The firing strength of each rule was computed using Equation (3).

$$\alpha_k = \min(\mu_{A_1^k}(x_1), \mu_{A_2^k}(x_2), \mu_{A_3^k}(x_3)) \quad (3)$$

Explanation:

- α_k = firing strength of the k -th rule
- $\mu_{A_j^k}(x_j)$ = membership degree of input x_j in the fuzzy set of the k -th rule

After all firing strengths were calculated, the implication step was performed. In the implication step, the consequent fuzzy set of each rule was clipped at the height of its firing strength using the minimum operator. This produced a truncated output fuzzy set for each active rule. The next stage was aggregation. In aggregation, the consequents of the active rules were combined using the maximum operator to aggregate the fuzzy output sets. The result was a single combined fuzzy set in the GPA domain. The final stage was defuzzification. Defuzzification was performed to obtain a single predicted GPA value from the Mamdani FIS. The method used was the centroid or center of gravity method, calculated using Equation (4).

$$GPA^* = \frac{\int y \cdot \mu_{agg}(y) dy}{\int \mu_{agg}(y) dy} \quad (4)$$

Explanation:

- GPA^* = predicted GPA
- y = point in GPA domain
- $\mu_{agg}(y)$ = aggregated membership degree at y

For comparison, linear regression and decision tree were used. Linear regression models the linear relationship between the three input variables—MATDAS, MATDIS, and DASPRO—and GPA as the output. The model was trained on the training data of each fold, similar to the Mamdani FIS. The linear regression equation with three variables was calculated using Equation (5).

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (5)$$

Explanation:

- $\hat{y}$ = predicted GPA
- x_1, x_2, x_3 = values MATDAS, MATDIS, DASPRO
- β_0 = intercept
- $\beta_1, \beta_2, \beta_3$ = regression coefficients

Decision tree is a non-parametric model that recursively partitions the input space based on feature values, assigning prediction to each resulting region. To prevent overfitting, the maximum depth of the tree was determined through an inner 5-fold cross-validation on the training data of each fold, selecting the depth that yielded the lowest RMSE among candidate values ranging from 2 to 8, as well as an unrestricted depth. The model was trained on the training data of each fold, similar to the Mamdani FIS and linear regression.

The prediction results from all three models were then evaluated by calculating the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). These two metrics were used to provide a more comprehensive evaluation. MAE calculates the average prediction error and treats all errors equally. Meanwhile, RMSE penalizes larger errors more heavily, making it more sensitive to outliers. Both metrics were calculated on the testing data of each fold and then averaged across all five folds. In addition to the average values, the standard deviation of RMSE and MAE across the five folds was also calculated to assess the stability of each model. Equation (6) shows the formula for RMSE, while Equation (7) shows the formula for MAE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

Explanation:

- y_i = actual GPA
- $\hat{y}_i$ = predicted GPA
- n = number of test data points

4 Results and Analysis

This section presents the outcomes of applying the Mamdani FIS, linear regression and decision tree models to predict GPA at the end of the fourth semester. The results are organized according to

the methodological stages, beginning with rule construction, followed by evaluation of the performance of all three models in predicting GPA.

As described in the methodology, the rule base was constructed for each fold. Since there were three input variables, each divided into three fuzzy sets, a total of 27 rule combinations were formed. The results of rule construction across folds are shown in Table 4. Because the rules were data driven, the consequents reflected the actual patterns in the data. Table 4 shows that of the 27 rules generated, 18 remained consistent across all five folds, while the other 9 varied. The rules that changed were numbers 2, 4, 5, 10, 13, 16, 19, 22, and 24. Moreover, a pattern was observed in the consequent changes across folds: most variations occurred between adjacent fuzzy output sets (e.g., low shifting to medium), with one exception in Rule 4, which shifted two levels (from very low to medium). This majority consistency indicates that the resulting rule base was relatively independent of specific data partitions, aligning with the purpose of cross validation to assess model generalization.

Table 4 Rule consequents across five folds

Rule#	MATDAS	MATDIS	DASPRO	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
1	Low	Low	Low	Low	Low	Low	Low	Low
2*	Medium	Low	Low	Low	Medium	Low	Low	Medium
3	High	Low	Low	Medium	Medium	Medium	Medium	Medium
4*	Low	Medium	Low	Medium	Medium	Medium	Low	Medium
5*	Medium	Medium	Low	Medium	Medium	Low	Low	Medium
6	High	Medium	Low	Medium	Medium	Medium	Medium	Medium
7	Low	High	Low	Low	Low	Low	Low	Low
8	Medium	High	Low	Medium	Medium	Medium	Medium	Medium
9	High	High	Low	High	High	High	High	High
10*	Low	Low	Medium	Low	Medium	Medium	Medium	Medium
11	Medium	Low	Medium	Medium	Medium	Medium	Medium	Medium
12	High	Low	Medium	Medium	Medium	Medium	Medium	Medium
13*	Low	Medium	Medium	Medium	Low	Medium	Medium	Medium
14	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium
15	High	Medium	Medium	High	High	High	High	High
16*	Low	High	Medium	Medium	High	High	High	High
17	Medium	High	Medium	High	High	High	High	High
18	High	High	Medium	High	High	High	High	High
19*	Low	Low	High	Medium	High	Medium	Medium	Medium
20	Medium	Low	High	High	High	High	High	High
21	High	Low	High	High	High	High	High	High
22*	Low	Medium	High	High	High	High	High	Medium
23	Medium	Medium	High	High	High	High	High	High
24*	High	Medium	High	High	High	High	High	High
25	Low	High	High	Medium	Medium	Medium	Medium	Medium
26	Medium	High	High	High	High	High	High	High
27	High	High	High	High	High	High	High	High

*Rules marked with * have consequents that vary across folds.

Beyond stability, the generated rule base was interpretable. Each rule could be expressed linguistically in IF-THEN form. For example, the rule “IF MATDAS = low AND MATDIS = low AND DASPRO = low THEN GPA = low” can be readily understood by academic advisors or

institutions as an indication that students with low grades in all three foundational courses are at risk of achieving a very low GPA. This characteristic reflects the interpretability of Mamdani FIS, which is well suited to educational contexts [14].

After the rule base was established, the predictive models, Mamdani FIS, linear regression, and decision tree, were evaluated on the test data of each fold using RMSE and MAE metrics. Table 5 presents the evaluation results. Overall, linear regression achieved an average RMSE of 0.3766 and MAE of 0.2874, which were lower than those of Mamdani FIS (RMSE = 0.5052, MAE = 0.3660). The decision tree obtained an average RMSE of 0.4776 and MAE of 0.3558, placing accuracy between the other two models. These findings indicate that linear regression was more accurate in predicting GPA. This result is consistent with prior studies reporting that linear regression can compete with, and even outperform, several machine learning methods in academic prediction tasks [15], [16].

Table 5 Evaluation of mamdani FIS, linear regression, and decision tree

Metric	Mamdani FIS	Linear Regression	Decision Tree
Average RMSE	0.5052	0.3766	0.4776
Std. Dev. RMSE	0.0664	0.0249	0.0213
CV RMSE (%)	13.1%	6.6%	4.5%
Average MAE	0.3660	0.2874	0.3558
Std. Dev. MAE	0.0459	0.0214	0.0183

In addition to higher accuracy, linear regression demonstrated greater stability than Mamdani FIS. The coefficient of variation for RMSE in linear regression was 6.6%, significantly lower than the 13.1% observed in Mamdani FIS. This suggests that regression performance was more consistent across variations in training data composition. In contrast, Mamdani FIS was more sensitive because its rule consequents depended on the distribution of input combinations in the training data. Consequently, changes in fold composition had a greater impact on prediction outcomes. The decision tree, however, showed the lowest variability of the three models, with coefficient of variation of 4.5% and the smallest standard deviation for both RMSE (0.0213) and MAE (0.0183). This stability can be attributed to its limited depth: the inner cross-validation selected a shallow tree with maximum depth of four, which reduced its sensitivity to changes in the training data.

A deeper analysis of the difference between RMSE and MAE provides insight into the error distribution. For Mamdani FIS, the difference between the average RMSE and MAE was 0.14 (0.5052 – 0.3660), larger than the 0.09 (0.3766 – 0.2874) observed in linear regression. The decision tree fell in between, with difference of 0.12 (0.4776 – 0.3558). Since RMSE penalizes larger errors more heavily than MAE, this greater difference indicates that Mamdani FIS produced more predictions with large errors, reflecting its lower stability compared to linear regression and decision tree.

The accuracy gap between Mamdani FIS and the other models could, in principle, be narrowed by optimizing the FIS. In this study, the membership function parameters were deliberately set according to the grading scale so that each fuzzy set retained an academic meaning: the medium set, for example, peaks at 65, which lies in the C – B grades zone. Automatic optimization, such as tuning the membership functions with evolutionary algorithms or adaptive neuro-fuzzy learning, typically shifts these parameters to minimize prediction error, which may cause linguistic terms such as low or medium to lose the meaning understood by academic advisors. This tension reflects the well-known conflict between accuracy and interpretability in linguistic fuzzy modeling, in which interpretability depends on factors such as the number of rules, the number of input variables, and the shape of fuzzy sets [20]. Optimization approaches that constrain the tuning process to preserve the semantic interpretability of the fuzzy partition have been proposed [21] and represent a promising direction for improving the accuracy of Mamdani FIS in this context without sacrificing its transparency.

Although linear regression excelled in accuracy and decision tree in stability, Mamdani FIS offered advantages in interpretability. Linear regression produces numerical coefficients that are less intuitive for non technical practitioners, whereas Mamdani FIS generates transparent linguistic rules. Thus, the choice among the three methods depends on priorities: linear regression is more suitable

when prediction accuracy is paramount, while Mamdani FIS is preferable when transparency in decision making is required. The decision tree offers a middle ground, being more accurate than Mamdani FIS and the most stable of the three models.

Decision trees are also commonly regarded as interpretable, since each path from the root to a leaf can be read as an IF–THEN rule [17]. However, these rules rely on crisp thresholds applied to the input grades, whereas the rules of Mamdani FIS use linguistic terms such as low, medium, and high, which correspond to how academic advisors typically describe student performance. The effect of crisp thresholds is illustrated by the tree obtained in Fold 1: for a student with MATDAS = 60 and MATDIS = 50, increasing DASPRO from 63 to 64 raised the predicted GPA from 2.26 to 2.64, because, along the path $\text{MATDIS} \leq 63.0$ and $\text{MATDAS} > 40.5$, the split at $\text{DASPRO} \leq 63.5$ assigned the two cases to different leaves, whereas Mamdani FIS produced the same prediction (2.16) for both. In addition, although all five trees reached a depth of four with 16 leaves, their rule structures differed across folds. The root split used MATDIS in four folds, with thresholds ranging from 55.5 to 67.5, and DASPRO in the remaining fold. The second-level splits also varied (Table 6). By contrast, the rule base of Mamdani FIS, whose antecedents are fixed by design, retained the same linguistic rules in every fold, with 18 of its 27 consequents unchanged. Consequently, the explanations offered by the decision tree depend more strongly on the particular training data. The lower accuracy of the decision tree compared with linear regression is also in line with [15], where linear regression models outperformed decision tree models on the final test set. For these reasons, Mamdani FIS remains better suited when explanations must be communicated to non-technical stakeholders.

Table 6 Decision tree split structure across five folds

Fold	Root split	Level-2 split (left)	Level-2 split (right)
1	$\text{MATDIS} \leq 63.0$	$\text{MATDAS} \leq 40.5$	$\text{DASPRO} \leq 77.5$
2	$\text{MATDIS} \leq 55.5$	$\text{DASPRO} \leq 41.5$	$\text{DASPRO} \leq 76.5$
3	$\text{DASPRO} \leq 63.5$	$\text{MATDIS} \leq 36.5$	$\text{DASPRO} \leq 78.5$
4	$\text{MATDIS} \leq 67.5$	$\text{DASPRO} \leq 44.5$	$\text{DASPRO} \leq 78.5$
5	$\text{MATDIS} \leq 66.5$	$\text{DASPRO} \leq 41.5$	$\text{DASPRO} \leq 72.5$

This interpretability advantage is particularly relevant in the context of the first evaluation stage. Since the prediction target aligns with the institutional evaluation at the end of the fourth semester, Mamdani FIS not only identifies students at risk of academic failure early but also provides explainable reasons through its rules. This enables more targeted interventions compared to models that only produce numerical predictions without explanation.

Compared with previous studies, this research offers several unique contributions. Study by [18] employed domain specific course grades with a Sugeno based ANFIS, whereas this study used Mamdani FIS, which is more interpretable. Furthermore, aligning prediction targets with institutional evaluation policies has rarely been addressed in prior literature. Therefore, this study provides practical relevance that distinguishes it from earlier work.

Despite these contributions, this study is limited to academic predictors. Only the grades of three foundational courses were used as inputs, because these grades are available at the end of the first semester, are objective, and are recorded consistently for all students. Non-academic factors such as attendance, learning engagement, and socio-economic conditions were not part of the dataset available for this study and were therefore not included. Previous studies have incorporated factors such as attendance and motivation [9], [11], [12], [13]; however, whether such factors add predictive value beyond early foundational grades in this context remains untested. Adding input variables to Mamdani FIS also involves a structural trade-off: with three fuzzy sets per variable, the number of possible rules grows exponentially, from 27 with three inputs to 81 with four and 243 with five. With 254 students, a larger share of rule combinations would contain no data and would require interpolation, and the resulting rule base would become more difficult to read, thereby weakening the interpretability that motivates the use of Mamdani FIS [20]. Future studies may therefore examine non-academic variables using strategies that keep the rule base compact, such as variable selection or hierarchical fuzzy systems.

5 Conclusion

This study compared Mamdani FIS, linear regression, and decision tree in predicting GPA at the end of the fourth semester. Predictions were based on three foundational courses: Basic Mathematics, Discrete Mathematics, and Introduction to Programming. Using five-fold cross-validation, linear regression achieved higher accuracy, with an average RMSE of 0.3766 and MAE of 0.2874, compared to Mamdani FIS (RMSE = 0.5052, MAE = 0.3660). The decision tree lay between the two in accuracy (RMSE = 0.4776, MAE = 0.3558) but showed the highest stability across folds. Linear regression also demonstrated greater stability than Mamdani FIS across folds. Nevertheless, Mamdani FIS offered superior interpretability, as evidenced by its transparent linguistic rule base, which was largely consistent across folds. Thus, there is a trade-off between prediction accuracy and interpretability. The choice of method depends on the intended priority. In the context of the first evaluation stage, the interpretability of Mamdani FIS adds value by enabling early identification of at-risk students while providing clear explanations for the predictions. Future research may explore data-driven rule construction or membership function design to improve the accuracy of Mamdani FIS without sacrificing its interpretability, and may examine whether non-academic factors, such as attendance and learning engagement, improve prediction beyond early foundational grades.

References

- [1] R. Q. Apumayta, J. C. Cayllahua, A. C. Pari, V. I. Choque, J. C. C. Valverde, and D. Huamán Ataypoma, "University Dropout: A Systematic Review of the Main Determinant Factors (2020-2024)," *F1000Res.*, Vol. 13, p. 942, Nov. 2024, DOI: 10.12688/f1000research.154263.2.
- [2] R. dan T. Sekretariat Direktorat Jenderal Pendidikan Tinggi, "Statistik Pendidikan Tinggi 2023 Higher Education Statistic," 2023.
- [3] M. of H. E. S. and T. Pusat Data dan Teknologi Informasi Sekretariat Jendral, "Statistik Pendidikan Tinggi 2024 Higher Education Statistic," 2024.
- [4] P. Valdiviezo-Diaz and J. Chicaiza, "Prediction of Academic Outcomes using Machine Learning Techniques: A Survey of Findings on Higher Education," in *International Conference on Applied Technologies*, Springer Cham, 2024, pp. 206–218. DOI: 10.1007/978-3-031-58956-0_16.
- [5] B. Albreiki, N. Zaki, and H. Alashwal, "A Systematic Literature Review of Student' Performance Prediction using Machine Learning Techniques," *Educ. SCI. (Basel).*, Vol. 11, No. 9, p. 552, Sep. 2021, DOI: 10.3390/educsci11090552.
- [6] V. Realinho, J. Machado, L. Baptista, and M. V. Martins, "Predicting Student Dropout and Academic Success," *Data (Basel).*, Vol. 7, No. 11, p. 146, Oct. 2022, DOI: 10.3390/data7110146.
- [7] M. Ben Said, Y. H. Kacem, A. Algarni, and A. Masmoudi, "Early Prediction of Student Academic Performance based on Machine Learning Algorithms: A Case Study of Bachelor's Degree Students in KSA," *Educ. Inf. Technol. (Dordr).*, Vol. 29, No. 11, pp. 13247–13270, Aug. 2024, DOI: 10.1007/s10639-023-12370-8.
- [8] A. H. Nabizadeh, D. Goncalves, S. Gama, and J. Jorge, "Early Prediction of Students' Final Grades in a Gamified Course," *IEEE Transactions on Learning Technologies*, Vol. 15, No. 3, pp. 311–325, Jun. 2022, DOI: 10.1109/TLT.2022.3170494.
- [9] M. O. Hegazi, B. Almaslukh, and K. Siddig, "A Fuzzy Model for Reasoning and Predicting Student's Academic Performance," *Applied Sciences*, Vol. 13, No. 8, p. 5140, Apr. 2023, DOI: 10.3390/app13085140.
- [10] N. U. Jan, S. Naqvi, and Q. Ali, "Using Fuzzy Logic for Monitoring Students Academic Performance in Higher Education," in *IEEC 2023*, Basel Switzerland: MDPI, Sep. 2023, p. 21. DOI: 10.3390/engproc2023046021.
- [11] J. A. Rojas, H. E. Espitia, and L. A. Bejarano, "Design and Optimization of a Fuzzy Logic System for Academic Performance Prediction," *Symmetry (Basel).*, Vol. 13, No. 1, p. 133, Jan. 2021, DOI: 10.3390/sym13010133.

- [12] B. K. Pathak, "Assessing Student Academic Performance with Fuzzy Expert System," *International Journal of Modern Education and Computer Science*, Vol. 17, No. 2, pp. 111–122, Apr. 2025, DOI: 10.5815/ijmecs.2025.02.05.
- [13] D. T. Thanh Loan, N. Duy Tho, N. Huu Nghia, V. D. Chien, and T. Anh Tuan, "Analyzing Students' Performance using Fuzzy Logic and Hierarchical Linear Regression," *International Journal of Modern Education and Computer Science*, Vol. 16, No. 1, pp. 1–10, Feb. 2024, DOI: 10.5815/ijmecs.2024.01.01.
- [14] C. Carrasco-Garrido, B. M. Moreno-Cabezali, and A. Martínez Raya, "New Perspectives on University Quality Assessment: A Mamdani Fuzzy Inference System Approach," *PLoS One*, Vol. 20, No. 5, p. e0321013, May 2025, DOI: 10.1371/journal.pone.0321013.
- [15] L. Falát and T. Piscová, "Predicting GPA of University Students with Supervised Regression Machine Learning Models," *Applied Sciences (Switzerland)*, Vol. 12, No. 17, Sep. 2022, DOI: 10.3390/app12178403.
- [16] I. Alnomay, A. Alfadhly, and A. Alqarni, "A Comparative Analysis for GPA Prediction of Undergraduate Students using Machine and Deep Learning," *International Journal of Information and Education Technology*, Vol. 14, No. 2, pp. 287–292, 2024, DOI: 10.18178/ijiet.2024.14.2.2050.
- [17] M. A. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA using Decision Trees: A Case Study," *International Journal of Information and Education Technology*, Vol. 6, No. 7, pp. 528–533, 2016, DOI: 10.7763/IJiet.2016.V6.745.
- [18] R. Mehdi and M. Nachouki, "A Neuro-Fuzzy Model for Predicting and Analyzing Student Graduation Performance in Computing Programs," *Educ. Inf. Technol. (Dordr.)*, Vol. 28, No. 3, pp. 2455–2484, Mar. 2023, DOI: 10.1007/s10639-022-11205-2.
- [19] E. Alhazmi and A. Sheneamer, "Early Predicting of Students Performance in Higher Education," *IEEE Access*, Vol. 11, pp. 27579–27589, 2023, DOI: 10.1109/ACCESS.2023.3250702.
- [20] M. J. Gacto, R. Alcalá, and F. Herrera, "Interpretability of Linguistic Fuzzy Rule-based Systems: An Overview of Interpretability Measures," *Inf. SCI. (N. Y.)*, Vol. 181, No. 20, pp. 4340–4360, Oct. 2011, DOI: 10.1016/j.ins.2011.02.021.
- [21] M. J. Gacto, R. Alcalá, and F. Herrera, "Integration of an Index to Preserve the Semantic Interpretability in the Multiobjective Evolutionary Rule Selection and Tuning of Linguistic Fuzzy Systems," *IEEE Transactions on Fuzzy Systems*, Vol. 18, No. 3, pp. 515–531, Jun. 2010, DOI: 10.1109/TFUZZ.2010.2041008.