

Interpretable and Adaptable Early Warning Learning Analytics Model

Shaleeza Sohail¹, Atif Alvi^{2,*} and Aasia Khanum³

¹The University of Newcastle, Sydney, 2000, Australia

²University of Management and Technology, Lahore, 54770, Pakistan

³Forman Christian College, Lahore, 54600, Pakistan

*Corresponding Author: Atif Alvi. Email: atif.alvi@umt.edu.pk

Received: 12 September 2021; Accepted: 13 October 2021

Abstract: Major issues currently restricting the use of learning analytics are the lack of interpretability and adaptability of the machine learning models used in this domain. Interpretability makes it easy for the stakeholders to understand the working of these models and adaptability makes it easy to use the same model for multiple cohorts and courses in educational institutions. Recently, some models in learning analytics are constructed with the consideration of interpretability but their interpretability is not quantified. However, adaptability is not specifically considered in this domain. This paper presents a new framework based on hybrid statistical fuzzy theory to overcome these limitations. It also provides explainability in the form of rules describing the reasoning behind a particular output. The paper also discusses the system evaluation on a benchmark dataset showing promising results. The measure of explainability, fuzzy index, shows that the model is highly interpretable. This system achieves more than 82% recall in both the classification and the context adaptation stages.

Keywords: Learning analytics; interpretable machine learning; fuzzy systems; early warning; interpretability; explainable artificial intelligence

1 Introduction

Learning analytics (LA) has been defined as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” [1], with the two most frequently used tasks being predicting students’ success and providing feedback [2]. LA is an active area of research and several studies have been done over the years to enhance its utility for various stakeholders including but not limited to teachers, students, and institutions. Current methods in LA have shown a lot of promise and are being applied in various setups with appreciable returns. One common outcome of these systems is to find at-risk students in early weeks of the course to provide extra support to facilitate them in successfully completing the course. The results of assessments and students’ learning behavior are used as indicators of their success in completing programming courses [3]. Similarly, a large amount of literature has



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

shown that with correct and timely analysis of data from learners, learning processes, and learning environments the success rate in higher education can improve significantly [4].

There are two basic limitations that are preventing across-the-board adoption of these methods with full confidence. First, these methods lack adaptability, and second there is poor understandability of the results of these models. Each course offering and the cohort has its own nuances like class size, students' demographic background, the year in which the student registers the course etc. Traditional Machine learning (ML) models are strongly dependent on the dataset used for training. Due to this reason, most LA systems are beneficial for only a specific type of context and cannot be adapted to different contexts involving other courses or other institutions. Poor understandability is a very common issue in ML approaches. Especially in this domain of learning analytics, such understanding is an important factor that limits the adaptability of these systems by different stakeholders. In recent years, interpretable and explainable ML techniques have gained a lot of popularity to design intelligent learning systems providing explanations of the predictions that are easily comprehended by common users [5]. Even though there is no agreement on the definition of interpretation in the ML research domain, one simple definition can be "the ability to explain or to present in understandable terms to a human" [6–8]. In the next section, recent literature on the use of interpretable ML for LA is reviewed, and as none of the research contributions measures the interpretability of student prediction systems, any comparison and evaluation is difficult to conduct.

In this paper, the hybrid framework proposed for facilitating the portability of models for learning analytics [9] has been extended by adding a new module, interpretability optimiser, that focuses on measuring and enhancing the interpretability of the ML approach. Earlier, a two-step hybrid approach for implementing the early warning system was used in our previous work. The first step is a data-centric algorithm that processes the data collected from the Learning management system (LMS) to learn a decision tree (CART and REPTree) giving the statistical relationship of various student-related parameters to the final course outcome. In the second step, fuzzy rules were constructed from the decision tree and inference was applied to predict a given student's likelihood of passing a course on the basis of his/her scores in the assessments so far. In the present paper, the framework has been enhanced by incorporating feature selection, context adaptation, and interpretability analysis. Feature selection gives us the best features to predict students at risk of failing the course during the early phase of the course. Context adaptation helps us to apply the learnt relationships between student data and course results across varying courses without having to be relearnt and remodeled. A comprehensive testing of the proposed approach has been performed using publicly available educational datasets. The transformative function shows the portability of the model across different courses. In addition, the interpretability of the ML model is analysed using fuzzy index [10]. The main contributions of the paper are:

- A statistical fuzzy framework enabling early intervention to help weak students.
- Adaptability of the model to new courses without retraining.
- Production of a set of linguistic fuzzy rules that are highly interpretable.
- Quantification of the interpretability of the fuzzy rules.
- High recall of both the classification and adaptation modules.
- Responsive and easy to use system.

Remainder of this paper is organized as follows. Section 2 discusses various research reports on LA, particularly addressing the interpretability and adaptability aspects. Section 3 describes the materials and methods used in the present study. Section 4 discusses the results. Lastly, Section 5 concludes the paper with recommendations for future research.

2 Literature Review

The amount of research in the domain of learning analytics involving ML approaches is tremendous. Surprisingly, very few research efforts consider explainability and interpretability of the ML models in this context. A few interpretable techniques are proposed for predicting students' performance based on data collected from learning management systems, assignment marks and other enrolment systems and provide interpretability of the results. However, none of the approaches measures the interpretability of student prediction systems, and hence any comparison and evaluation is difficult to conduct. When it comes to measuring interpretability of ML models, a common approach is to use qualitative methods using human subjects to analyse the perceived interpretability of these models [11]. However, due to the difference of application domains and the nature of predictions, these methods may not be feasible and suitable. Hence, we use a quantitative method to evaluate the interpretability in the proposed system as it provides a way to objectively compare different ML models without involving human subjects. A method proposed for quantitatively measuring interpretability is discussed in Section 3.

Moreover, the literature on quantification or measurement of explainability of AI models used for learning analytics is even more limited. In this section, a comprehensive overview of literature in the domain of learning analytics is provided with a focus on interpretability.

Socio-demographic data in addition to academic and LMS activity data is used to predict student performance in order to provide early support for at-risk students [12]. The three datasets are partitioned into sub-datasets to form student subgroups at first level on the basis of gender, age, attendance type and attendance mode. On the second level the gender based sub-datasets are further partitioned into 6 sub-datasets according to student age, attendance type and attendance mode. The student performance is predicted using four classification methods using the sub-datasets: two black box methods are naïve-Bayes and SMO and the two white box methods are J48 and JRip. The black box techniques produced models that were not interpretable and hence cannot be effectively used for intervention purposes. The white box techniques produced interpretable models by identifying features that impacted performance predictions for different sub-groups. Due to the comprehensibility of the predictions in the rule and tree form, these models can be used for early intervention purposes [12].

Most early warning systems predict students' performance using large student datasets that ignore the idiosyncrasies of underrepresented students and consider general student population data only [13]. Moreover, the lack of decision making reasoning makes these systems difficult to adopt by institutions. Educators cannot use these systems for providing effective student support due the inability of these systems to provide motivation for particular predictions. An interpretable rule based genetic programming classifier by using student data from multiple sources can predict at-risk students, especially the ones belonging to underrepresented minorities. These students are generally faced with more challenges in their educational journey and hence, this interpretable prediction can assist educators to provide early intervention effectively. The performance of multi view programming approach is consistently better when compared with other white box rule and tree based traditional approaches. Moreover, the rules provide comprehensibility of predictions for at-risk and not at-risk classes for easy adaptation and understanding [13].

The interpretation of rules is provided by CN2 rule inducer and multivariate projection for the student performance prediction system that uses video learning analytics and data mining techniques. Student academic data, student activity data and student video interaction data are used to predict student performance by multiple algorithms. In addition, the effect of feature selection and transformation is also compared. The best prediction accuracy was achieved by the Random Forest

algorithm with equal width transformation method and information gain ratio selection technique. The CN2 rule inducer algorithm also performed well but it provided easy rule induction with probability for non-expert viewers like educators that require this interpretation for providing support to students [14].

High drop-out rate and poor academic performance are two main issues affecting the reputation of educational institutes [15]. Association rule algorithm with Classification Based on Algorithm (CBA) rule generation algorithm can predict student performance in advance. This provides interpretable information to educators to adjust their teaching strategies and to provide additional support for struggling students. The experiments showed medium accuracy in predicting student performance, but it provides insight into the factor affecting the performance.

The student performance prediction systems that use black box techniques can be converted to interpretable systems by employing some design practices [16]. A black box prediction model, C-parameterized margin, SVM classifier (C-SVC) was chosen due to its ability of providing high prediction accuracy, and the interpretability of the system was enhanced by leveraging a rich catalogue of data visualization methods. Three features were considered in the design of the classifier to improve impressibility: multiclass, probabilistic and progressive. Hence, a progressive multiclass probabilistic prediction model not only provided high prediction accuracy but was interpretable so the educators can use it to support their students. A set of representation tools was discussed to provide expressive graphical output.

A knowledge gap has been identified between the model creation for student performance prediction and the interpretation of that prediction for actionable decision-making process. For this pedagogical change, a model based on recursive partitioning and automatic selection of features for robust classification was developed with high interpretability. The strength of the model was the transparent characterization of student subgroups based on relevant features for easy translation into actionable processes [17].

For student performance prediction systems, an important aspect that must be considered is prediction uncertainty or confidence in addition to prediction accuracy for developing a reliable early warning system for at-risk students. Two Bayesian deep learning models and Long Short-Term Memory (LSTM) models were used for predicting students' performance in future courses based on their performance in courses already completed by the students. Prediction uncertainty associated with at-risk student prediction was considered before reaching out to provide additional support for effective and targeted utilization of resources. Also, the explainable results of the models provided information about the previous courses whose results influenced the prediction, which can be used for guiding students [18].

A combination of black box and white box prediction approaches was proposed for high prediction accuracy and interpretability. High prediction accuracy was achieved by using the SVM model and for interpretability the Decision Tree (DT) and Random Forest (RF) models were employed for extracting symbolic rules. In addition, an attribute dictionary was built from students' comments which was converted to attribute vectors for predicting students' grades. The combination of these techniques showed accurate prediction of students' performance based on students' comments after each lesson and the interpretable results showed the characteristics of attribute patterns for each grade [19].

The above mentioned literature provides interpretable student performance prediction models but the quantification and measurement of the interpretability is not considered in any of these [20]. For quantification of explainability models, generally user studies and qualitative approaches are

used [21]. However, the interpretability of models can be iteratively enhanced by employing relevant quantification methods which also provides means to compare these approaches. Interpretability indices have been proposed for white box approaches like rule-based systems which can effectively be used to quantify the explainability of these models for comparison and evaluation [22].

3 Proposed Approach

The main components of the proposed framework are discussed in this section and Fig. 1 provides an overview of the framework.

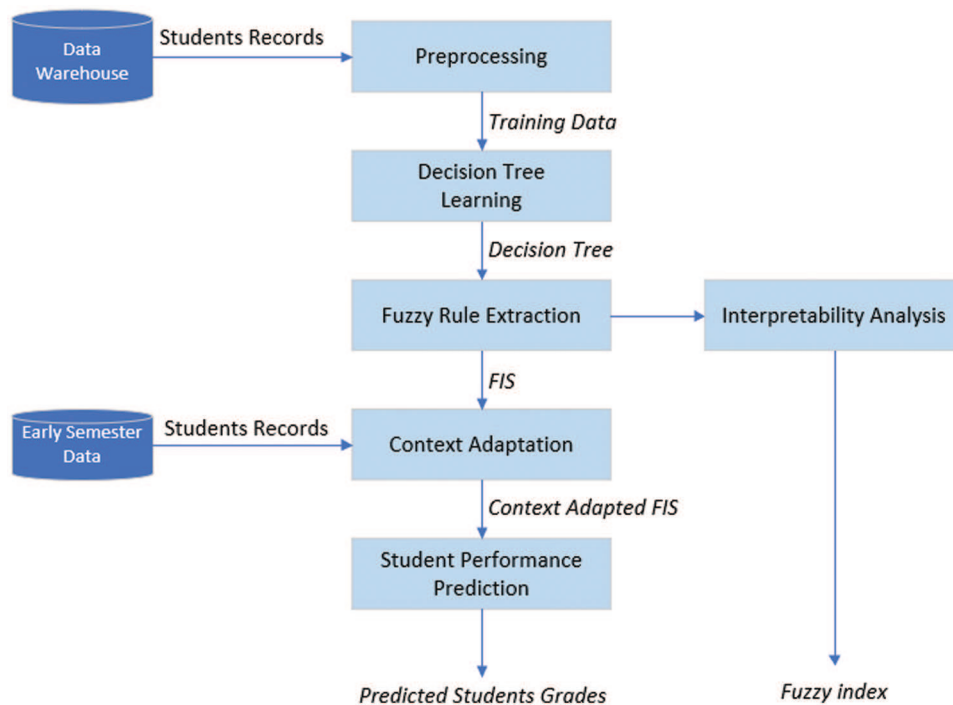


Figure 1: System overview

3.1 Decision Trees

Decision Trees are an established and efficient tool in Machine Learning that are increasingly being adopted to support explainability of algorithmic decisions of classification and regression. Various algorithms have been developed for constructing optimized decision trees under various conditions. Some examples include CART, ID3, REPTree, C4.5 etc.

Tree algorithms apply a top-down, divide-and-conquer approach to the data to construct a tree or a set of rules. In a decision tree, the inner nodes represent value sub-ranges of the input variable and the leaf nodes represent the output values. The tree is constructed by recursively splitting the dataset by applying statistical measures to the variables and selecting a split variable based on the results. Some examples of statistical measures are Entropy, Information Gain, and Gini Index. Once the tree has been constructed, various input-output mapping rules can be traced out by traversing the tree from the root to a leaf. Moreover, there are algorithms for optimizing the tree in terms of its complexity. Pruning algorithms are the most popular in this category and many tree construction

algorithms incorporate pruning in their operation based on various criteria, e.g., reduced error pruning (REP) is used by a REPTree. In the present work, the REPTree algorithm is used, which is an efficient decision tree algorithm capable of learning both classification and regression problems.

3.2 Construction of Fuzzy Inference System

Fuzzy inference systems (FIS) have been used effectively in various domains including pattern recognition, healthcare, robotics, and control engineering etc. Based around Fuzzy sets (FS), FIS are ideally suited to domains with a large number of complex factors and non-linear relationships which cannot be expressed by clear mathematical equations. FIS's also have the unique advantage in terms of their adaptability. Unlike other modeling approaches like neural networks, regression, etc., which need to be redeveloped and retrained for every new context, FIS can be easily adapted to the new context without having to be retrained.

A fuzzy set (FS) is an extension of the classical set where members of a set have different degrees of membership in the set. FS offer an ideal representation tool to represent imprecise and approximate concepts. FS are used to define linguistic variables to partition a Domain of discourse (DD). For instance, class size may be partitioned into three linguistic variables {low, medium, high}, with each linguistic variable expressed as a FS $\mu_A(x) \rightarrow [0, 1]$ where A is the linguistic variable, x is any variable from the DD, and μ represents the degree of membership in the FS. Membership functions (MF) can take various forms, e.g., Gaussian, trapezoidal, triangular, sigmoidal etc. In the context of current research, two types of MF's are important. Gaussian membership functions of the form Eq. (1) below are used by fuzzy sets inside the domain:

$$\mu(x, m, \sigma) = \frac{e^{-\left(\frac{x-m}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad (1)$$

where m and σ are the mean and standard of the domain respectively. Sigmoidal MF's of the form Eq. (2) below are used by fuzzy sets on domain boundaries:

$$\mu(x, a, c) = \frac{1}{1 + e^{-a(x-c)}} \quad (2)$$

where a and c are the slope and crossover point respectively. For most of the applications, normal fuzzy sets are preferred which require that at least one member $x \in X$ such that $\mu_A(x) = 1$ where X is the DD. The set of all members having a membership degree of 1 in the FS constitute the core of the FS. Correspondingly, the support of a FS is a crisp set containing all members of DD having a non-zero membership in the FS.

An FIS is a rule-based decision system making use of fuzzy sets. Various fuzzy operators are applied to aggregate the rule activations in the RB. The main components of an FIS are the rule base (RB) and the knowledge base (KB). The rule base comprises various rules of the form:

$$\text{Rule}^i : \text{IF } f_1^i \text{ IS } Fi_1 \text{ AND } f_2^i \text{ IS } Fi_2 \text{ AND } \dots \text{ AND } f_n^i \text{ IS } Fi_n \text{ THEN } p^i \text{ IS } P^i \quad (3)$$

where f_k^i are the input features, Fi_k are the input linguistic variables, p^i are the output variables, and P^i are the output linguistic variables. The definitions of fuzzy sets corresponding to the input and output linguistic variables are stored in the knowledge base. Input is provided in parallel to all the rules which are fired according to the degree of activation of the antecedent fuzzy sets. An Inference Engine maps the rule activations to output fuzzy sets which are then aggregated and optionally defuzzified to produce the output. For extracting the fuzzy rules from the decision tree, two steps are performed. First, appropriate fuzzy sets are defined to represent the feature sub-intervals represented by the tree

nodes. In the second step, various paths are traced out from the root to the leaves in the tree; each of these paths is converted to a fuzzy rule by mapping the input-output variables to the corresponding fuzzy sets.

3.3 Interpretability Analysis

Once the fuzzy rules and fuzzy sets are created the interpretability of the FIS is measured. A number of approaches have been proposed in literature for measuring interpretability of rule-based systems and tree based systems [23]. The most common indices [22] used for fuzzy logic systems are fuzzy index and Nauck index. In this paper, the interpretability of the proposed system is evaluated using the fuzzy index.

Fuzzy index is proposed as an interpretability measure for fuzzy systems and is inspired by the Nauck index [22]. The fuzzy index is assessed using a hierarchical fuzzy system with six input variables and one output variable for the four linked knowledge bases that are part of the complete systems. Four different knowledge bases calculate:

- Rule Base Dimension of the system: considering the number of rules and premises.
- Rule Base Complexity of the system: considering the number of rules with one, two and three or more variables.
- Rule Base Interpretability of the system: considering Rule Base Complexity and Rule Base Dimension of the system.
- Fuzzy index of the system (final output): considering Rule Base Interpretability and average number of labels defined by input variables.

The labels of the six inputs are identified and the four rule bases in the hierarchical fuzzy system work together to find the fuzzy index of the fuzzy system. The inputs and output of the system that calculates the interpretability of the fuzzy system are shown in [Tab. 1](#). The readers are encouraged to read [22] for further details on the calculation of the fuzzy index.

Table 1: Input and output variables for the fuzzy index [11]

Type	Name of the variable	Description	Labels
Input	Total number of rules	Number of rules in the fuzzy system	3 (Low, medium, high)
Input	Total number of premises	Number of premises in the fuzzy system	2 (Low, high)
Input	Number of rules that use one input variable	Number of rules in the fuzzy system with one input variable	2 (Low, high)
Input	Number of rules that use two input variables	Number of rules in the fuzzy system with two input variables	2 (Low, high)
Input	Number of rules that use three or more input variables	Number of rules in the fuzzy system with three or more input variables	2 (Low, high)

(Continued)

Table 1: Continued

Type	Name of the variable	Description	Labels
Input	Average number of labels defined by input variables	Number of labels defined by the input variables in the fuzzy system	2 (Low, high)
Output	Interpretability index	Interpretability index of the fuzzy system	5 (Very low, low, medium, high, very high)

3.4 Context Adaptation Module

As mentioned earlier, FS can be adapted to new contexts without needing to be retrained. Context is extremely important in correctly applying an LA model to predict future performance of the students. For instance, a class of size 30 might be considered large in case of an elective course but would be considered medium for a core course. However, research on context-adaptation for LA applications is scanty so far.

It is generally agreed that, among the two main components of an FIS, the RB is universal and context-independent whereas the KB is context-aware. Accordingly, an FIS can be adapted by transforming the KB according to the context without affecting the logical structure of the RB. The method given in [24] (for the construction domain) is used for applying the transformation functions to the KB. As an example, the procedure involved in adapting a fuzzy set A from a base context B to its adapted version \hat{A} in a new context N is described by the following steps:

Initially, the KB is populated with the definition of A according to the base context B defined over the DD $[b_l, b_h]$. Since Gaussian MF's are being used, the parameters of A will be (x, m, σ) , where m and σ are the mean and standard deviation respectively. When context adaptation is required, first of all determine the range of the new context N as $[n_l, n_h]$, where $n_l = \min(x_1, x_2, \dots, x_n)$ and $n_h = \max(x_1, x_2, \dots, x_n)$ where x_i are the feature values over the new domain. A context adaptation factor is then calculated as:

$$\tau(x) = Qx + n_l \quad (4)$$

$$Q = \frac{n_h - n_l}{b_h - b_l} \quad (5)$$

The context adapted version of A is calculated by replacing the base parameters with the adapted values as follows:

$$\hat{A}(x) = A(\tau(x)) \quad (6)$$

where $x \in [n_l, n_h]$ and $\tau(x) \in [b_l, b_h]$. The reader is encouraged to read for further details of these equations.

4 Model Implementation and Evaluation

In this section the implementation and evaluation of the proposed model is discussed.

4.1 Dataset

The validation of the proposed framework has been carried out using the Open University Learning Analytics Dataset (OULAD) [25], which is offered by the Open University, UK, one of the largest distance learning universities in the world. OULAD comprises anonymized student demographic and performance data for various course modules offered by the Open University in the years 2013 and 2014. The data captures registrations, demographics (gender, age band, region, index of multiple deprivation), assessments, and virtual learning environment interactions of the students. All data is tabular in CSV (comma separated values) format and the tables can be interlinked using key columns. The intent was to predict the passing or failing of a student early in a course by using machine learning and fuzzy logic on the student's data.

4.2 Preprocessing

To extract the tree, student data for 2 offerings of the same course in 2013 and 2014 has been used, represented in OULAD as course BBB with offerings 2013B and 2014B respectively. There were 8590 relevant records of 2531 distinct students, considering that only those assessments were chosen that were marked by the tutor (and not by the computer), as they carried the majority of the marks, and were due at most by the 120th day of the course, since an early intervention based on these is desired. There were 4 such assessments for each of the chosen course modules. All the relevant data was combined into one file and was then further processed by taking the average of the scores obtained in the aforementioned assessments for each student. As it was required to have a binary classification of the final result of a student as Pass or Fail, the instances of a final result of Distinction were renamed as Pass, and Withdrawn as Fail, without a loss of accuracy. Context adaptation was later applied to another cohort comprising 2 courses represented as AAA in OULAD with offerings 2013 J and 2014 J respectively. The data was preprocessed as for the pre-classification step and resulted in 678 relevant records that were then input to the context adaptation module (see Section 4.7).

4.3 Classification Using Decision Tree

Decision tree learning was preferred as the predictive technique because it uses a white-box model and is thus easy to explain and interpret. Moreover, it is computationally less intensive and requires less data preprocessing. The Scikit-learn library for CART (Classification and Regression Trees) algorithm and REPTree (Reduced error pruning) algorithm] from WEKA (Waikato environment for knowledge analysis) library] were used.

Both algorithms gave similar results. CART's binary tree was built by splitting nodes on the basis of Gini impurity index. Pandas and Numpy were used for data manipulation. WEKA's REPTree algorithm builds a decision or regression tree using information gain/variance reduction and prunes it using reduced-error pruning. Optimized for speed, it only sorts values for numeric attributes once. It deals with missing values by splitting instances into pieces, as C4.5 does. You can set the minimum number of instances per leaf, maximum tree depth (useful when boosting trees), minimum proportion of training set variance for a split (numeric classes only), and number of folds for pruning.

The features found to be most important for classifying a student as likely to Pass or Fail are: i) the student's average score in the assessments, ii) highest education level attained by the student previously, and iii) the index of multiple deprivation, which is essentially a poverty measure. A training-test split of 80-20 was used to build the REPTree model. Fig. 2 shows the REPTree for the training data.

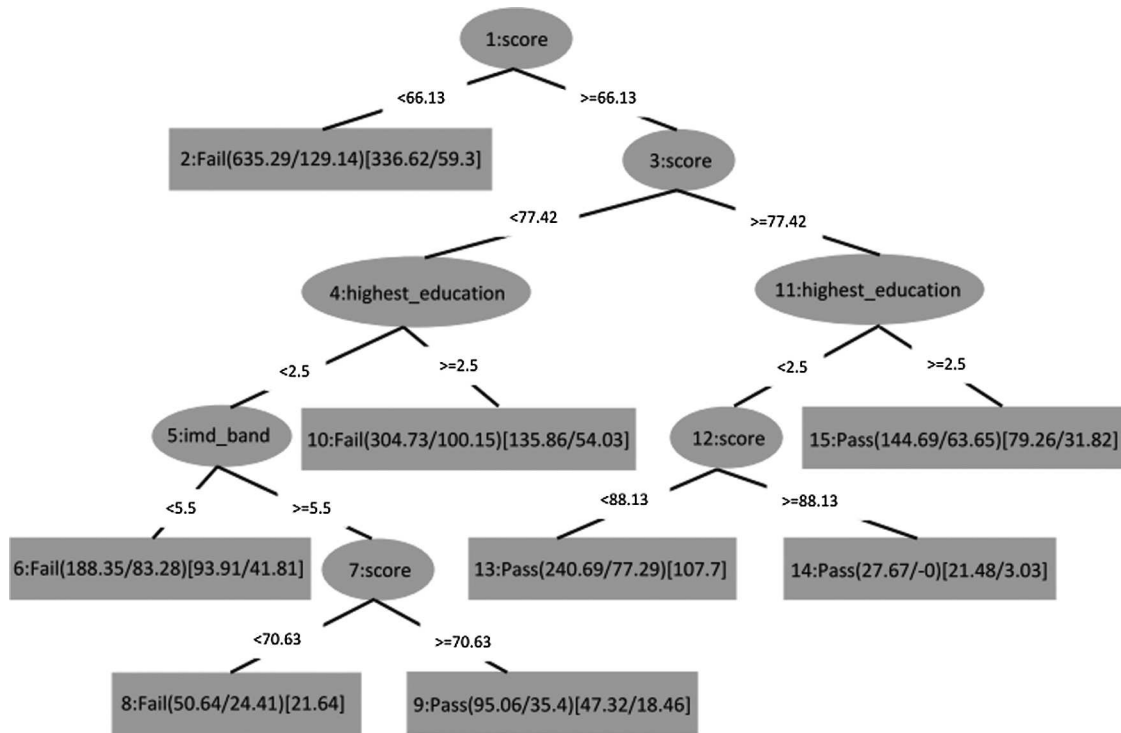


Figure 2: REPTree model of training data

4.4 Construction of Fuzzy Inference System

After obtaining the tree, each path from the root to a leaf is traversed and the feature sub-intervals expressed in each of the nodes are extracted. After examining the collected feature sub-intervals, fuzzy linguistic variables were defined to express these sub-intervals. Following input features were fuzzified with the help of linguistic variables: $F = \{Score, Education, IMDBand\}$. The output variable Outcome is also fuzzified. All linguistic variables employ normalized Gaussian membership functions defined by (1) and (2), as explained in the previous section. These MF's are incorporated into fuzzy IF-THEN rules of the form (3). An example MF is shown in Fig. 3. An FIS is constructed around these rules, with these properties: FIS Type: MAMDANI, Inference: MIN-MAX and defuzzification: Centroid. The FIS was further tuned to optimize the rule structure and composition. The FIS with three input features and one output feature is shown in Fig. 4. The system is capable of reflecting the complex and nonlinear relationships between various variables. Fig. 5 shows the rule surface underscoring the nonlinear relationship between the inputs and the output.

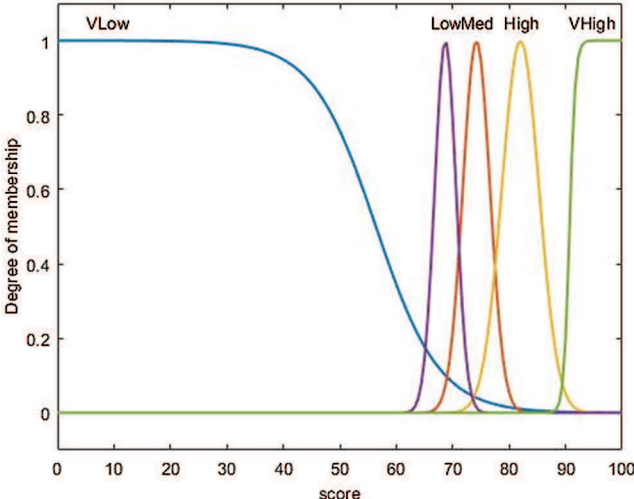


Figure 3: Fuzzy sets for score

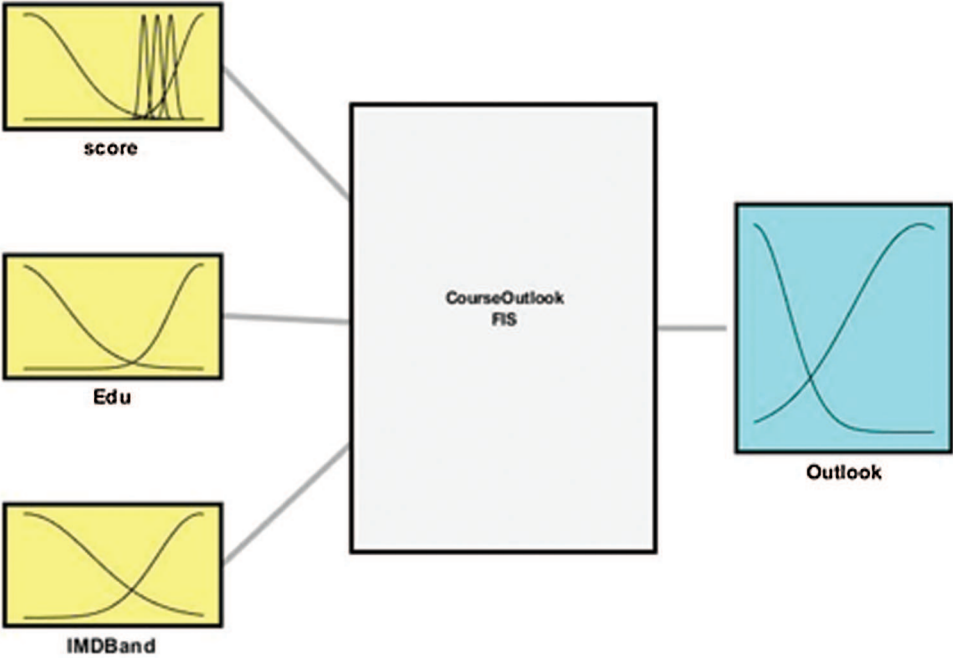


Figure 4: FIS structure for grade prediction

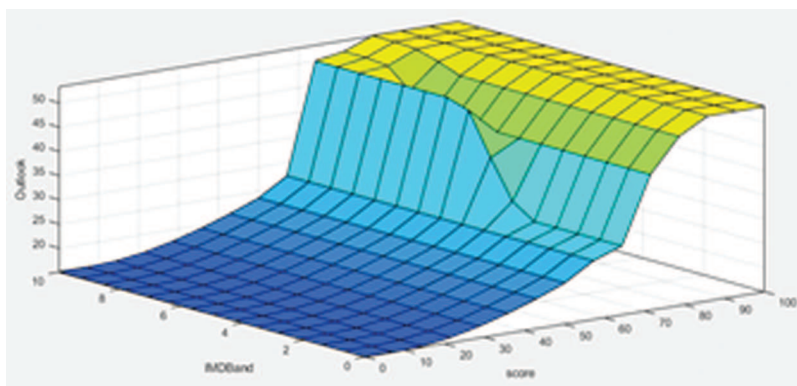


Figure 5: Non-linear relationships between the FIS variables

4.5 Interpretability Analysis

The parameters used for calculation of the fuzzy index of the FIS are shown in [Tab. 2](#):

Table 2: Input and output variables for the FIS to calculate the fuzzy index

Input variable	Value
Total number of rules	4
Total number of premises	6
Number of rules that use one input variable	2
Number of rules that use two input variables	2
Number of rules that use three or more input variables	0
Average number of labels defined by input variables	2.5

Using the values given in [Tab. 2](#), the outputs of the four knowledge bases designed to find the fuzzy index [10] for this system are given below:

- Rule Base Dimension of the system: Low.
- Rule Base Complexity of the system: Low.
- Rule Base Interpretability of the system: Very high.
- Fuzzy index of the system (final output): Very high.

The fuzzy index of the designed FIS is evaluated as very high which makes the system highly interpretable and hence, easy to adapt by professional and academic staff in educational institutions.

4.6 Context Adaptation Module

Context adaptation was also experimented with by using the existing data of previous cohorts to predict the performance of a new cohort (see Section 4.3) during the early phase of a semester. The parameters m and σ were adapted for the new cohort using the previous cohort as the base domain. It was observed that only the parameter m was enough to give sufficiently acceptable performance

in our domain, so only μ has been used for the reported results. For each of the input features, the corresponding linguistic variables in the KB were adapted using the method stated in Section 3.4 above. FIS performance was tested on the adapted KB on the new domain, without making any changes to the RB.

4.7 Results and Discussion

[Tabs. 3](#) and [4](#) depict the performance of the modules. For the particular scenario of enabling early intervention to help students likely to fail a course, recall is the most important performance metric since it is imperative not to miss any potentially failing student in the results. Hence, having less false negatives (high recall) at the expense of having more false positives (low precision) is acceptable. The proposed system achieves more than 82% recall in both the REPTree classifier and the context adaptation module ([Tabs. 3](#) and [4](#) respectively). An important aspect of the model is its ability to predict students' performance by considering only a subset of the assessments undertaken in the first few weeks of the semester. This provides an opportunity for early intervention by providing academic support to these students. This inherent requirement of using a subset of data to predict students' performance lowers the prediction accuracy but provides crucial data required for successful early intervention to support academically weak students. Moreover, the fuzzy index of the proposed system is Very High (calculated in Section 4.6), which makes the predictions easy to understand and use for early intervention purposes for all the stakeholders. Due to a tradeoff between system accuracy and interpretability, more interpretable systems tend to be less accurate. Despite this limitation, the proposed system shows acceptable accuracy and high interpretability.

Table 3: REPTree detailed accuracy by class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.828	0.515	0.531	0.828	0.647	0.322	0.729	0.622	Fail
	0.485	0.172	0.800	0.485	0.604	0.322	0.729	0.765	Pass
Weighted average	0.626	0.314	0.689	0.626	0.622	0.322	0.729	0.705	

Table 4: Accuracy details for adapted FIS

Precision	Recall	TP Rate	FP Rate	F Score	Class
0.293478	0.823171	0.823171	0.056641	0.432692	Fail
0.865741	0.365234	0.365234	0.176829	0.513736	Pass

Considering the class imbalance present in the dataset, weighted averages of the performance measures are also reported, using the class size as a weight. It is clear from [Tab. 3](#) that the weighted average is well above 50% for both precision and recall.

5 Conclusion and Future Work

This paper presented a comprehensive approach to provide an early answer to the million-dollar question in learning analytics: which students are likely to fail this course? Early intervention can put

such students back on the path of success. This hybrid statistical fuzzy system identifies such students using their performance in the initial assessments of the course and a few other features through a learning decision tree and then generates a set of fuzzy rules. These rules are easy to understand, and this is quantified by measuring their interpretability using the fuzzy index. Finally, the FIS is context adapted and used to predict the likelihood of student success and failure in other courses, without any retraining being required. The performance of the system is high in terms of recall, the main parameter of success in this scenario.

Some of the limitations of the proposed system are due to the fact that the system is designed to identify students that are at risk of failing the course considering only the subset of assessments undertaken in the early weeks of the semester. Hence, the accuracy of the system is not very high when compared to other models based on data from all assessments of the course. The other limitation of the system arises from the requirement of high interpretability which necessitates keeping the decision tree simple.

The proposed system is an end-to-end solution to a key learning analytics problem and this work shall be extended by using the framework on several datasets to discover more features that contribute to student success and failure. Other learning algorithms will also be used, both supervised and unsupervised, in the classification stage of the framework and their performance and interpretability will be measured. Another research thread to be pursued in future is the human in the loop paradigm to make this framework even more accurate, interpretable, and sentient.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] G. Siemens and R. Baker, "Learning analytics and educational data mining: Towards communication and collaboration," in *Proc. 2nd Int. Conf. on Learning Analytics and Knowledge*, China, vol. 12, no. 1, pp. 252–254, 2012.
- [2] D. Gaevic, S. Dawson and G. Siemens, "Let's not forget: Learning analytics are about learning," *TechTrends*, vol. 59, pp. 64–71, 2015.
- [3] F. Pereira, E. T. D., Oliveira, D. D. Oliveira, A. Cristea, L. C. Valho *et al.*, "Using learning analytics in the amazonas: Understanding students' behavior in introductory programming," *British Journal of Educational Technology*, vol. 51, pp. 955–972, 2020.
- [4] D. Ifenthaler and J. Yau, "Utilising learning analytics to support study success in higher education: A systematic review," *Educational Technology Research and Development*, vol. 20, no. 1, pp. 1–13, 2020.
- [5] R. Marcinkevics and J. E. Vogt, "Interpretability and explainability: A machine learning zoo mini tour," *ArXiv*, vol. 20, no. 3, pp. 1–24, 2020.
- [6] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *ArXiv*, vol. 1, no. 3, pp. 1–13, 2017.
- [7] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 3, pp. 206–215, 2019.
- [8] O. González, "Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view," *IEEE Access*, vol. 7, no. 1, pp. 154096–154113, 2019.
- [9] S. Sohail, A. Khanum and A. Alvi, "Hybrid fuzzy-statistical system for learning analytics," in *Proc. IEEE Int. Conf. on Teaching, Assessment, and Learning for Engineering (TALE)*, Australia, pp. 989–994, 2018.

- [10] J. Alonso, S. Guillaume and L. Magdalena, “A hierarchical fuzzy system for assessing interpretability of linguistic knowledge bases in classification problems,” in *Proc IPMU*, 2006.
- [11] F. P. Sangdeh, D. Goldstein, J. Hofman, J. W. Vaughan and H. Wallach, “Manipulating and measuring model interpretability,” in *Proc. Conf. on Human Factors in Computing Systems*, Australia, vol. 20, no. 1, 2021.
- [12] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson *et al.*, “Predicting academic performance by considering student heterogeneity,” *Knowledge-Based Systems*, vol. 161, no. 3, pp. 134–146, 2018.
- [13] A. Cano and J. Leonard, “Interpretable multiview early warning system adapted to underrepresented student populations,” *IEEE Transactions on Learning Technologies*, vol. 12, pp. 198–211, 2019.
- [14] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. Sarker *et al.*, “Predicting student performance in higher educational institutions using video learning analytics and data mining techniques,” *Applied Sciences*, vol. 10, no. 4, pp. 1–20, 2020.
- [15] G. Badr, A. Algobail, H. Almutairi and M. Almutery, “Predicting students’ performance in university courses: A case study and tool in KSU mathematics department,” *Procedia Computer Science*, vol. 82, no. 5, pp. 80–89, 2016.
- [16] C. Villagfa, F. J. Duán, F. Llorens, P. Rosique, R. Satorre-Cuerda *et al.*, “Improving the expressiveness of black-box models for predicting student performance,” *Computers in Human Behavior*, vol. 72, no. 4, pp. 1–13, 2016.
- [17] A. Pardo, N. Mirriahi, R. Martinez-Maldonado, J. Jovanovic, S. Dawson *et al.*, “Generating actionable predictive models of academic performance,” in *Proc. of the 6th Int. Conf. on Learning Analytics & Knowledge*, China, vol. 16, no. 5, pp. 474–478, 2016.
- [18] Q. Hu and H. Rangwala, “Reliable deep grade prediction with uncertainty estimation,” in *Proc. 9th Int. Conf. on Learning Analytics & Knowledge*, China, vol. 19, no. 1, pp. 76–85, 2019.
- [19] S. Sorour, S. Rahman, S. Kahouf and T. Mine, “Understandable prediction models of student performance using an attribute dictionary,” in *Proc. Int. Conf. on Web-Based Learning*, Russia, vol. 21, no. 4, pp. 161–171, 2016.
- [20] R. Alamri and B. Alharbi, “Explainable student performance prediction models: A systematic review,” *IEEE Access*, vol. 9, no. 2, pp. 33132–33143, 2021.
- [21] A. Nguyen and M. R. Martinez, “On quantitative aspects of model interpretability,” *ArXiv*, vol. abs/2007.07584, 2020.
- [22] T. R. Razak, J. Garibaldi, C. Wagner, A. Pourabdollah and D. Soria, “Interpretability indices for hierarchical fuzzy systems,” in *Proc. IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, Italy, pp. 1–6, 2017.
- [23] M. Pota, M. Esposito and G. Pietro, “Interpretability indexes for fuzzy classification in cognitive systems,” in *Proc. IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)*, Canada, pp. 24–31, 2016.
- [24] A. Tsehayae and A. Fayek, “Context adaptation of fuzzy inference system-based construction labor productivity models,” *Advances in Fuzzy Systems*, vol. 18, no. 6, pp. 1–16, 2018.
- [25] J. Kuzilek, M. Hlosta and Z. Zdrahal, “Open university learning analytics dataset,” *Scientific Data*, vol. 4, no. 1, pp. 1–8, 2017.