Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2, 1-19
2025
Publisher: Learning Gate
DOI: 10.55214/2641-0230.v8i1.12872

From data to decision support: A multicriteria model for interpretable dropout risk classification

Bertha Lucía Santos-Hernández[1], Francisco Antonio Serrano Camarena[2], Juan Antonio Granados Montelongo[3], Juan Antonio Álvarez Gaona[4], José Daniel Corona Flores[5], María Magdalena Hernández Borrego[6*]

[1,6]Faculty of Accounting and Administration Autonomous University of Coahuila, Mexico; b.l.santos.hernandez@outlook.com (B.L.S.H.) m.m.hernandez.borrego@outlook.com (M.M.H.B.).
[2]Faculty of Law Autonomous University of Coahuila, Mexico; f.a.serrano.camarena@outlook.com (FA.S.C.).
[3]Department of Renewable Natural Resources, Antonio Narro Autonomous Agrarian University, Mexico; juan.granados@uaaan.edu.mx (JA.G.M.).
[4]Faculty of Marketing Autonomous University of Coahuila, Mexico; antonio_alvarezgaona@uadec.edu.mx (J.AA.G.).
[5]Academic Language Unit, Antonio Narro Autonomous Agrarian University, Mexico; dcorona@uaaan.edu.mx (J.D.C.F.).

Abstract: In this paper, a model for multicriteria ordinal classification is proposed and used for student dropout risk assessment in a structured, understandable, and uncertain manner. The problem is modeled as a multicriteria sorting problem, where a student is evaluated with respect to a set of criteria and then classified according to the risk level. Interval-based performance representations are used to account for the presence of uncertainties and incomplete information, as well as possible interactions between criteria, without using oversimplified additive assumptions. Preference parameters are not fixed a priori but are derived through a disaggregation process using observed data. The model is applied to a sample of 45 students, each representing a unit of analysis. Data collection is conducted using a structured tool, and the results are aggregated into a matrix for evaluation. The classification results provide a clear risk profile for each student, both globally and on a dimensional level. It is not a matter of seeking any generalization but rather illustrating the usefulness of the model as a support for decisions in academic monitoring.

Keywords: *Dropout risk classification, Educational data analysis, Multicriteria decision analysis, Ordinal sorting.*

1. Introduction

One of the biggest problems in higher education is that students drop out. This has a significant effect on their lives, the efficiency of schools, and the overall social return on investment in education. Recent international reports continue to show that participation, progression, and equity in education remain central policy concerns, especially where institutions serve heterogeneous student populations under unequal socioeconomic conditions (OECD, 2024; UNESCO Institute for Statistics, 2024). In this context, the issue of dropout is no longer viewed merely as an individual decision but as a multifaceted phenomenon shaped by academic, economic, social, and institutional factors. Recent literature underscores that dropout should be analyzed as a complex and dynamic process rather than as a singular event resulting from a single variable. Barragán Moreno and González Támara (2024) assert that student attrition in higher education requires a multidimensional analysis that includes both institutional and societal factors. A recent systematic review conducted by Quincho Apumayta et al. (2024) identified academic performance, socioeconomic conditions, vocational orientation, and institutional factors as enduring determinants of university dropout. The convergence in the literature suggests that early risk identification should not rely exclusively on a singular metric such as grades but rather on a systematic integration of criteria that can capture the diverse circumstances that threaten students.

* Correspondence: m.m.hernandez.borrego@outlook.com

The emergence of educational data mining and learning analytics in recent years has facilitated the development of predictive models and early warning systems designed to identify students at risk of attrition promptly. A recent systematic review and meta-analysis by Venkatesan, Karmegam, and Mappillairaju (2024) reveals an increasing application of statistical and machine learning methodologies for this purpose; however, it also highlights ongoing difficulties related to class imbalance, interpretability, and the geographic concentration of evidence. Quimiz-Moreira, Delgadillo, Parraga-Alava, Maculan, and Mauricio (2025) recently reviewed the literature on predicting university dropouts using machine learning. They found that even though methods have improved, prediction is still difficult, and understanding models remains a significant problem. If teachers and administrators cannot easily understand how more accurate models work, they might not be as useful for making decisions in schools.

This tension between predictive performance and practical interpretability is especially important in schools, where the goal of a model is more than just sorting students into groups; it is also to help teachers make timely and fair decisions. Recent empirical studies demonstrate both the potential and the constraints of contemporary predictive methodologies. Carballo-Mendívil, Arellano-González, Ríos-Vázquez, and Lizardi-Duarte (2025), employing pre-enrollment data from a Mexican public university, illustrated that early warning models can detect at-risk students before their inaugural class, thus allowing institutions to implement prompt interventions. In a similar manner, Chang, Chen, and Lee (2025) demonstrated that a tailored early warning system in Taiwanese higher education can be integrated with human-directed interventions within a closed prediction–intervention cycle. However, these studies also bring up a significant practical issue: many predictive tools still use complicated algorithmic frameworks or feature interactions that aren't always clear to people who work at institutions. So, even though finding problems early is important, businesses also need decision-support tools that are easy to use, well-organized, and work in real life. From this perspective, Multiple Criteria Decision Analysis (MCDA) offers a substantial conceptual and methodological alternative. MCDA is designed to help with decision-making problems where you need to consider more than one factor at a time, and those factors might not agree with each other (Greco, Słowiński, & Wallenius, 2025). MCDA doesn't aim to achieve the best prediction score. Instead, it shows how the criteria are set up, how important each dimension is, and how alternatives are compared or grouped. This is especially important for dropout-risk analysis, where people in education often need not only a final label but also an explanation of why a student is thought to be at a certain level of risk and which factors play the biggest role in that judgment.

Multicriteria sorting models are especially effective for MCDA problems where choices need to be grouped into ordered categories, such as low, medium, or high dropout risk. Ru, Liu, Kadziński, and Liao (2023) argue that multiple criteria sorting involves distributing alternatives into predefined, prioritized categories. Their study demonstrates that probabilistic ordinal regression can effectively handle unclear preference data and still produce accurate class assignments. This insight is significant for educational applications, as student information is often missing, inconsistent, or unclear. A model that incorporates uncertainty aligns more closely with real academic environments, where indicators are not always measured with precision, and risk assessments often involve ambiguity rather than definitive conclusions.

The present study occupies the intersection of two fields: the imperative for effective early identification of dropout risk and the demand for comprehensible, uncertainty-aware decision support. This paper redefines dropout-risk assessment not solely as a black-box predictive task but as an ordinal multicriteria classification challenge. Students are evaluated through a hierarchical framework of criteria, enabling the examination of risk on a global level as well as within particular dimensions. The model is also designed to handle uncertainty in performance representation and prevent people from making overly simple guesses about how criteria contributions work. This method aligns with new work in preference disaggregation and robust ordinal reasoning in MCDA, which aim to determine preference parameters from real-world data instead of just guessing them (Corrente, De Smet, Doumpos, Greco, & Zopounidis, 2023; Ru et al., 2023). As a result, this paper makes two important contributions. First, it proposes a comprehensible multicriteria model for assessing dropout risk in higher education. It emphasizes organized reasoning, ordinal classification, and an awareness of uncertainty. Second, it demonstrates how useful the model is in real life by applying it to a group of 45 students. This shows how this method can assist with academic oversight in a real classroom. The objective is not to extrapolate the findings to a broader population but to demonstrate how a rigorous and practical framework can help schools organize information, identify diverse risk profiles, and improve the informed nature of academic follow-up. In this context, the study addresses a specific gap in current literature: the need for dropout-risk instruments that are not only predictive but also interpretable, contextually relevant, and directly applicable to educational decision-making.

2. Literature Review

The literature on student dropout in higher education has evolved from broad explanatory theories of persistence toward increasingly data-driven approaches for early risk detection. A foundational reference in this field is Tinto's theory of student departure, which explains persistence as strongly associated with students' academic and social integration within the institution (Tinto, 2012). In parallel, Bean and Metzner's model broadened the discussion by showing that, for nontraditional students, environmental and external variables may weigh as much as or more than campus integration variables (Bean & Metzner, 1985). Together, these perspectives remain important because they frame dropout as a multidimensional process rather than a single academic outcome.

More recent review studies confirm that this multidimensional understanding remains valid. Aina, Baici, Casalone, and Pastore (2022) synthesize the socio-economic literature on university dropout and conclude that attrition depends on a combination of individual, institutional, and economic factors, mediated by students' ability to integrate into academic life. This aligns with broader recent literature, which repeatedly identifies academic achievement, financial constraints, prior preparation, motivation, institutional support, and social adjustment as interrelated drivers of persistence or withdrawal. Therefore, current research supports the idea that dropout should be studied through multiple criteria rather than isolated indicators. This multidimensionality has also driven the growth of predictive analytics and early warning systems in higher education. In online education, Bañeres, Rodríguez-González, Guerrero-Roldán, and Cortadas (2023) emphasize that early identification and intervention are central to reducing dropout, but they also note that intervention practices are less developed than predictive efforts. Rahmani, Groot, and Rahmani (2024), in a systematic review of online higher education dropout, show that risk factors span demographic, technological, motivational, course-related, and support dimensions. These findings are relevant beyond fully online settings because they reinforce the general point that dropout risk emerges from the interaction of heterogeneous dimensions that institutions must monitor in a coordinated way.

At the same time, the expansion of predictive modeling has exposed an important methodological tension. Many machine learning models improve classification performance, but they often do so at the expense of transparency and interpretability. In educational contexts, this is a serious limitation because academic staff need more than a probability score: they need a defensible explanation of why a student is classified as being at a certain risk level and which dimensions contribute to that assessment. For this reason, the literature increasingly distinguishes between prediction-oriented systems and decision-support approaches that are explicit, auditable, and useful for intervention design. This is where Multiple Criteria Decision Analysis becomes especially relevant. MCDA offers a structured family of methods for supporting decisions involving several criteria, making the reasoning process more transparent and traceable. Within this field, sorting methods are particularly suitable when the goal is to assign alternatives to ordered categories rather than to rank them continuously. Alvarez, Ishizaka, and Martínez (2021) describe multiple-criteria sorting as a major branch of decision-support methodology for categorization problems, while (Liao, Xiao, Wu, & Wen, 2024) show that this family of methods has grown substantially because of its rigor, interpretability, and applicability in complex classification contexts. Such properties are directly relevant for classifying dropout risk into ordinal categories such as low, medium, or high.

A further advantage of multicriteria sorting models is their ability to deal with imperfect knowledge. Recent surveys of the field show that contemporary sorting methods increasingly incorporate indirect elicitation, learning from assignment examples, robust ordinal regression, and stochastic or probabilistic reasoning under uncertainty (Alvarez et al., 2021; Liao et al., 2024). This is methodologically significant for dropout assessment because educational information is rarely complete or fully precise. Student profiles may involve variability, partial evidence, or uncertainty regarding the interpretation of some criteria. Consequently, a model that allows structured classification while accommodating uncertainty is better aligned with real academic monitoring contexts than a purely deterministic scheme. Based on this literature, a clear gap can be identified. The dropout field has accumulated strong theoretical explanations and increasingly sophisticated predictive tools, yet institutions still need operational models that combine three characteristics: multidimensionality, interpretability, and sensitivity to uncertainty. A multicriteria ordinal classification approach responds directly to this gap because it organizes the problem through a hierarchy of criteria, produces ordered risk categories, and can incorporate preference structures and uncertainty explicitly. For studies focused on academic follow-up and decision support in concrete institutional contexts, this type of framework is particularly suitable.

3. Methodology

3.1. Research Design

This study adopts an applied, non-experimental, cross-sectional methodological design aimed at building and implementing a multicriteria model for the ordinal classification of student dropout risk. The purpose is not to estimate population parameters or test causal effects but to develop a structured decision-support tool capable of assigning students to ordered risk categories in a transparent, interpretable, and uncertainty-sensitive manner. In this sense, the study is framed as a multiple criteria sorting problem, where each student is described by several criteria and then allocated to one of a finite number of preference-ordered classes. Such a formulation is consistent with the literature on multicriteria sorting, hierarchical decision models, and robust ordinal regression, all of which emphasize classification, interpretability, and decision support under incomplete or imprecise information (Corrente, Greco, & Słowiński, 2016; González-Morales, López-Aguilar, Álvarez-Pérez, & Toledo-Delgado, 2025; Greco, Mousseau, & Słowiński, 2010; Ru et al., 2023).

3.2. Unit Of Analysis and Application Context

The unit of analysis is the individual student. The model is applied to a group of 45 students, each evaluated independently according to the criteria system defined for the study. Therefore, the set of alternatives can be represented as:

$$A = \{a_1, a_2, \dots, a_{45}\},$$

where each a_i denotes one student. The application is conceived as a concrete demonstration of the model's practical usefulness for academic monitoring rather than as a statistical study intended for population generalization. This scope is fully consistent with the project definition, which explicitly states that the purpose of the application is to show how the model can support interpretation and follow-up in a specific student group. The first operational stage consists of collecting the relevant information for each student through a structured data collection form. The purpose of this instrument is to ensure homogeneous data capture, reduce ambiguity, and organize the information required to evaluate the criteria included in the model. Once completed, the forms are reviewed, cleaned, coded, and transformed into an evaluation matrix.

Formally, the student database is represented by the performance matrix

$$X = \left[x_{ij}\right]_{n\times m},$$

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

where $n = 45$ is the number of students and m is, the number of elementary criteria. Each element x_{ij} denotes the performance of the student a_i on criterion g_j. Thus, each student is described by the performance vector

$$\mathrm{x}_i = (x_{i1}, x_{i2}, \dots, x_{im}).$$

This representation is standard in multiple criteria decision analysis because it allows the evaluation process to be organized transparently and in a replicable way while preserving each criterion's contribution to the final classification (Corrente et al., 2016; Greco et al., 2010).

3.3. Hierarchical Structure of Criteria

A central component of the proposed methodology is the construction of a hierarchical system of criteria for evaluating student dropout risk. This structure is adopted because dropout is not a phenomenon that can be adequately represented through a single variable or a flat list of unrelated indicators. On the contrary, it is a multidimensional condition in which different academic, behavioral, contextual, and institutional elements may contribute simultaneously, although not necessarily with the same intensity or at the same level of interpretation. For this reason, the model organizes the criteria into a hierarchy that links broad analytical dimensions with more specific elementary indicators.

In formal terms, let $G = \{g_1, g_2, \dots, g_m\}$ denote the set of elementary criteria used to evaluate the students, and let $D = \{D_1, D_2, \dots, D_p\}$ denote the set of higher-level dimensions. Each dimension D_r groups together a subset of elementary criteria that are conceptually related, such that

$$G = \bigcup_{r=1}^{p} G_r,$$

where $G_r \subseteq G$ is the subset of elementary criteria belonging to the dimension D_r. Under this structure, the hierarchy may be represented as a tree whose upper level corresponds to the overall dropout-risk assessment, the intermediate level corresponds to broader dimensions of analysis, and the lower level corresponds to measurable elementary criteria. This hierarchical organization performs several methodological functions. First, it improves the conceptual coherence of the model by preventing the evaluation from becoming a simple aggregation of isolated variables. Grouping the criteria into broader dimensions makes it possible to distinguish between different sources of vulnerability rather than merging them prematurely into a single score. Second, it strengthens the interpretability of the results, since the final risk classification can be decomposed and examined by dimension. Thus, the model is able to show not only whether a student is classified as low, moderate, or high risk, but also which dimensions contribute most strongly to that classification. Third, it enhances the operational usefulness of the model for academic monitoring because decision makers can identify more clearly the areas in which intervention may be required.

From an analytical standpoint, the use of a hierarchy allows the evaluation to be performed at two complementary levels. At the global level, all criteria contribute to the overall classification of the student into one of the ordered dropout-risk categories. At the partial level, each higher-order dimension can be examined independently in order to understand the internal profile of the student. This distinction is especially important in educational contexts because students with the same overall level of risk may reach that condition through different combinations of weaknesses and strengths. One student may show elevated risk because of poor academic performance, whereas another may present a similar global classification due to instability in attendance, low engagement, or contextual constraints. A hierarchical model preserves this internal structure and therefore provides a richer interpretation than a single undifferentiated index.

Another important advantage of the hierarchical arrangement is that it facilitates the treatment of heterogeneous indicators. In dropout-risk assessment, the criteria included in the model may differ

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

substantially in nature, scale, and meaning. Some may be numerical, others ordinal, and some may even be interval-valued if uncertainty must be represented. By embedding them in a hierarchical structure, the model avoids direct comparison among all indicators at once and instead organizes them within coherent analytical blocks. This reduces the risk of conceptual inconsistency and contributes to a more transparent preference model.

The hierarchical structure also supports the possibility of considering partial interactions within and across dimensions. Since the criteria are not interpreted in isolation, the hierarchy provides a natural framework for analyzing whether some criteria reinforce or compensate each other within a given dimension. For example, within a particular analytical block, two weak performances may jointly signal a level of vulnerability that is more serious than would be suggested by examining each criterion separately. Conversely, strong performance in one elementary criterion may partially mitigate weakness in another, depending on the preference model ultimately inferred. In this way, the hierarchy is not merely a descriptive arrangement but an integral part of the decision logic.

Operationally, the development of the hierarchy requires a prior process of criteria identification, grouping, and validation. First, the relevant indicators for dropout-risk assessment are identified according to the information available in the structured data collection form. Second, these indicators are grouped into dimensions based on conceptual affinity and analytical usefulness. Third, the resulting structure is reviewed to ensure that each elementary criterion contributes clearly to one dimension and that the hierarchy as a whole remains manageable, interpretable, and aligned with the study's objective. This process is particularly important because an excessively fragmented hierarchy may hinder interpretation, whereas an overly simplified hierarchy may conceal relevant distinctions among students.

Within the context of this study, the hierarchical structure is especially appropriate because the project explicitly seeks to provide a structured, interpretable, and uncertainty-sensitive classification of dropout risk and to apply it to a specific group of 45 students. The hierarchy makes it possible to connect the raw information collected for each student with a logically organized decision model, thereby transforming dispersed student data into an analytical structure that can support follow-up and academic decision-making. Rather than producing only a final label, the model is able to explain the basis of the classification through the dimensions and subcriteria involved. Therefore, the hierarchical structure of criteria should be understood as one of the foundations of the proposed methodology. It allows the dropout-risk problem to be represented in a way that is analytically rigorous, substantively meaningful, and practically useful. By articulating the relationship between global assessment, partial dimensions, and elementary indicators, the hierarchy provides the structural basis upon which the subsequent stages of performance representation, preference inference, and ordinal classification are built.

The methodology also considers that some student information may be imprecise, incomplete, or variable over time. For this reason, the model allows criterion performances to be represented by intervals whenever a single exact value would create a false sense of precision. This design choice appears explicitly in the project description, which states that performances may be represented by intervals when the available information suggests variability or uncertainty.

Accordingly, when uncertainty is present, the performance of the student a_i on criterion g_j is represented as

$$\tilde{x}_{ij} = [x_{ij}^L, x_{ij}^U],$$

where x_{ij}^L and x_{ij}^U are the lower and upper bounds of performance. The interval-valued representation is methodologically coherent with recent developments in multicriteria sorting under uncertainty, where imprecise inputs are incorporated directly into the decision model rather than being artificially reduced to fixed-point estimates (de Almeida et al., 2015; Ru et al., 2023).

3.4. Ordinal Risk Categories and Classification Problem

The central decision problem addressed in this study is the ordinal classification of students according to their dropout risk. This means that the objective is not to estimate a continuous score for its own sake, nor to produce an unconstrained ranking from the “best” to the “worst” student, but rather to assign each

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

student to one among a finite number of predefined and ordered risk categories. This formulation is especially appropriate in the educational context because academic follow-up decisions are commonly made in terms of differentiated levels of concern or priority, such as low, medium, or high risk, rather than through raw numerical outputs alone. In this sense, the problem fits naturally within the framework of multiple criteria sorting, where alternatives are allocated to ordered classes based on their performance across several criteria (González-Morales et al., 2025; Ru et al., 2023).
Let

$$A = \{a_1, a_2, \dots, a_n\}$$

denote the set of students to be evaluated, where each a_i represents one student, and let

$$C = \{C_1, C_2, \dots, C_K\}$$

denote the set of ordered dropout-risk categories, such that

$$C_1 \prec C_2 \prec \cdots \prec C_K.$$

Here, C_1 represents the lowest level of dropout risk and C_K the highest. The number and denomination of the categories depend on the needs of the application context, but the essential methodological condition is that they must be preference-ordered and substantively interpretable. Thus, the categories are not arbitrary labels; they represent progressively different levels of academic vulnerability and are intended to support differentiated follow-up actions.

This ordinal formulation has several advantages. First, it preserves the practical meaning of the assessment by translating student information into categories that can be directly interpreted by academic staff. Second, it avoids the false precision often associated with continuous scores, especially when the underlying information is heterogeneous or partially uncertain. Third, it is more compatible with intervention design, since institutions typically need to determine which students require routine observation, targeted support, or urgent attention. Therefore, the model is designed to produce classifications that are both analytically rigorous and operationally useful.

Formally, the classification problem can be expressed as the construction of a decision rule

$$\mathcal{M}: A \to C,$$

such that each student $a_i \in A$ is assigned to one category $C_h \in C$. This decision rule depends on the student's performance across the criteria system and on the set of preference parameters inferred by the model. If x_i denotes the performance vector of a student a_i. Then the classification can be represented generically as

$$\mathcal{M}(a_i) = \mathcal{M}(\mathrm{x}_i; \Theta),$$

where Θ is the set of admissible model parameters. In this formulation, the assignment is not based on a single observed variable but on the joint evaluation of multiple criteria organized hierarchically and interpreted through a structured preference model. From an operational standpoint, ordinal classification requires a mechanism for distinguishing one risk category from another. In generic terms, this can be done through class boundaries, reference profiles, or threshold intervals, depending on the final multicriteria formulation adopted. A simple value-based representation of this logic is

$$a_i \in C_h \iff \tau_{h-1} < V(a_i) \le \tau_h,$$

where $V(a_i)$ is the global evaluation of student a_i, and $\tau_1, \tau_2, \dots, \tau_{K-1}$ are threshold values separating adjacent categories. By convention, $\tau_0 = -\infty$ and $\tau_K = +\infty$. Under this interpretation, the role of the model is to determine whether the global evaluation of a student places that student within the interval corresponding to a particular level of dropout risk. This expression should be understood as a generic representation of ordinal sorting logic rather than as a restriction to a purely additive structure. An alternative and often more conceptually rich representation is to define boundary profiles between categories. Let

$$B = \{b_1, b_2, \dots, b_{K-1}\}$$

be the set of profiles separating adjacent classes, where each b_h represents the boundary between categories C_h and C_{h+1}. In this case, the classification rule compares the performance of each student with these profiles to determine the most appropriate class. This perspective is especially useful when the classification must remain interpretable in terms of qualitative transitions between levels of academic vulnerability rather than merely in terms of numerical cutoffs. A major methodological strength of the proposed formulation is that the categories are ordinal rather than nominal. This is important because the classification does not treat all categories as unrelated groups. Instead, it assumes that there is an intrinsic ordering among them, which reflects progressive severity in dropout risk. This ordering makes it possible to reason about the closeness of adjacent categories, the direction of misclassification, and the robustness of the assignment. For example, confusing a student between low and moderate risk is not equivalent to confusing that same student between low and high risk, because the latter implies a much larger discrepancy in the interpretation of academic vulnerability.

The ordinal structure is also crucial for dealing with uncertainty and incomplete preference information. Since the model parameters are inferred by disaggregation rather than fixed arbitrarily, there may be more than one admissible parameter set compatible with the available information. Consequently, a student's class assignment may not always be uniquely determined by a single parameterization. To address this, the methodology can distinguish between possible and necessary assignments, an approach consistent with robust ordinal regression (Greco et al., 2010; Ru et al., 2023).

Let Θ^* denote the set of all admissible parameter vectors compatible with the preference information used in the model. Then, a student a_i is said to be possibly assigned to a class C_h. If there exists at least one compatible parameterization under which the student is classified in that category:

$$\exists\theta \in \Theta^* \text{ such that } \mathcal{M}(a_i;\theta) = C_h.$$

Likewise, the student a_i is said to be necessarily assigned to the class C_h. If all compatible parameterizations lead to the same category:

$$\forall\theta \in \Theta^*,\ \mathcal{M}(a_i;\theta) = C_h.$$

This distinction is methodologically valuable because it makes the analysis more transparent. A necessary assignment indicates a robust conclusion, whereas a possible assignment reflects ambiguity that may arise from uncertainty in performances, interactions among criteria, or incompleteness in the preference structure. In the context of dropout monitoring, this information is highly relevant because it allows the analyst to distinguish between students whose risk level is firmly established and students whose classification should be interpreted more cautiously. Another important aspect of the classification problem is that it is decision-oriented rather than purely predictive. The purpose of the categories is not merely descriptive. Each category should ideally correspond to a different level of institutional attention and potentially to a different type of academic response. In this sense, the classification model serves as an intermediate layer between raw student information and educational decision-making. A student assigned to a lower-risk category may require regular observation, whereas a student located in a higher-risk category may justify immediate follow-up or more intensive support measures. Thus, the ordinal categories are meaningful not only analytically but also operationally.

Moreover, because the model is hierarchical, the classification problem must be understood at two interconnected levels. At the global level, the student is assigned to one overall dropout-risk category. At the partial level, the same student may exhibit different local positions across specific dimensions of the hierarchy. This feature is especially useful when two students belong to the same global category but arrive there through different patterns of weakness and strength. Consequently, the global assignment provides a synthetic diagnosis, while the partial structure supports a more nuanced interpretation of why the student belongs to that category.

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

3.5. Preference Modeling and Disaggregation-Based Parameter Inference

A central assumption of this study is that the model parameters should not be fixed arbitrarily. Instead, the preference structure is inferred by disaggregation, that is, by identifying parameter values that are compatible with the intended classification logic and the available information. This point is explicitly stated in the project note and is one of the most important methodological features of the proposal.

In formal terms, let

$$\Theta = \{w, \tau, \gamma, \ldots\}$$

Denote the set of preference parameters, where $w$may represent importance coefficients, τ class thresholds or boundary conditions, and γ interaction parameters whenever dependence among criteria is considered. Rather than imposing a single ad hoc parameter vector, the methodology searches for one or more admissible parameterizations that satisfy the preference information encoded in the model. This logic follows the tradition of robust ordinal regression, where model parameters are inferred from holistic judgments or assignment examples, and only those parameter values that are compatible with the available preference information are retained (Greco et al., 2010).

A generic representation of the global evaluation process is

$$V(a_i) = F(\mathrm{x}_i; \Theta),$$

where $F(\cdot)$ is the multicriteria aggregation function. The classification rule then maps this global evaluation or an equivalent outranking relation into one of the ordered classes. If a value-based operationalization is adopted, one possible threshold-based representation is

$$a_i \in C_h \iff \tau_{h-1} < V(a_i) \leq \tau_h,$$

with $\tau_0 = -\infty$ and $\tau_K = +\infty$. This expression should be read as a generic formalization of ordinal sorting; the exact operational rule may vary depending on the specific multicriteria model finally implemented. The important point is that the classification is produced through a structured preference model rather than through arbitrary scoring. This improves analytical defensibility and makes the model more appropriate for academic decision support (Greco et al., 2010; Ru et al., 2023).

The methodology also admits the possibility that some criteria may interact. This means that the joint effect of two criteria on dropout risk may differ from the simple sum of their separate contributions. The project description explicitly indicates that the model should allow for interaction among criteria to avoid overly simplified interpretations.

Accordingly, the aggregation rule is not restricted to a purely additive structure. In abstract form, this means that

$$V(a_i) \neq \sum_{j=1}^{m} w_j \, v_j(x_{ij})$$

in all cases, because the model may include interaction terms when required by the decision logic. A more general expression is

$$V(a_i) = F(v_1(x_{i1}), v_2(x_{i2}), \ldots, v_m(x_{im}); \Theta),$$

where F may capture synergistic or redundant effects among criteria. This is methodologically relevant in dropout-risk assessment because some combinations of conditions may signal greater vulnerability together than when interpreted separately. Hierarchical multicriteria models with interaction effects are recognized in the MCDA literature as especially useful when complex decision phenomena cannot be adequately represented through simple independence assumptions (Corrente et al., 2016; de Almeida et al., 2015).

3.6. Step-By-Step Application Procedure

The implementation procedure is organized into six sequential stages:

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

1. Definition of the criteria system. The dropout-risk dimensions and elementary criteria are defined and arranged hierarchically.
2. Data collection. The structured form is completed for each of the 45 students.
3. Data cleaning and coding. The collected information is revised and transformed into the evaluation matrix.
4. Specification of risk classes. The ordered dropout-risk categories are established.
5. Inference of compatible parameters. Preference parameters are identified through a disaggregation process consistent with the classification logic.
6. Classification and interpretation. Each student is assigned to a risk category, and the result is interpreted both globally and by partial dimensions.

This sequence directly reflects the operational logic contained in the project document, particularly the construction of the data form, the matrix of student-by-criterion performances, and the generation of an ordinal risk classification for each student.

The model produces two complementary types of outputs. First, it generates a global ordinal classification for each student, indicating the risk category to which the student belongs. Second, it enables partial analysis by dimension, allowing the analyst to identify which components of the criteria hierarchy contribute most strongly to the assigned category. This dual reading is one of the main strengths of the methodology because it supports both concise diagnosis and more detailed academic follow-up. The project note also identifies interpretability and usefulness for academic monitoring as core expected outcomes of the application.

In methodological terms, this means that the model is not limited to labeling students; it also provides a structured analytical basis for differentiating cases of higher and lower risk and for identifying the dimensions in which support actions may be most justified. This kind of interpretability is precisely what makes multicriteria sorting especially suitable for educational decision support, where the usefulness of a model depends not only on classification accuracy but also on the transparency of the reasoning behind each assignment. It should be emphasized that the present application has an analytical and demonstrative scope. The purpose is to show how an uncertainty-sensitive multicriteria ordinal classification model can be implemented in a real educational context and used as a support tool for academic follow-up. Therefore, the study does not seek statistical generalization to a larger population. Instead, it aims to demonstrate the operational viability, interpretability, and practical value of the proposed model in a concrete group of students. This delimitation is explicitly stated in the project description and is coherent with the role of MCDA as a decision-aiding methodology oriented toward structured support in specific contexts.

4. Results

4.1. Overall Classification Outcome

The application of the proposed multicriteria ordinal classification model to a group of 45 students produced a differentiated and interpretable distribution of dropout risk. The model assigned each student to one of three ordered categories: low risk, moderate risk, and high risk. The resulting pattern showed that most students were concentrated in the lower and intermediate categories, while a smaller but still substantively important group was located in the highest-risk class. This distribution is consistent with what would be expected in an academic cohort where vulnerability is present but not generalized across the entire group.

At the global level, the results indicate that dropout risk in the analyzed group is not dichotomous. Instead, the model reveals a gradation of vulnerability, allowing the distinction between students who appear academically stable, those who show emerging signs of difficulty, and those whose profiles are already clearly critical. This is one of the main strengths of the multicriteria ordinal approach because it avoids the analytical loss associated with binary classification and produces outputs that are more useful for educational decision-making.

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

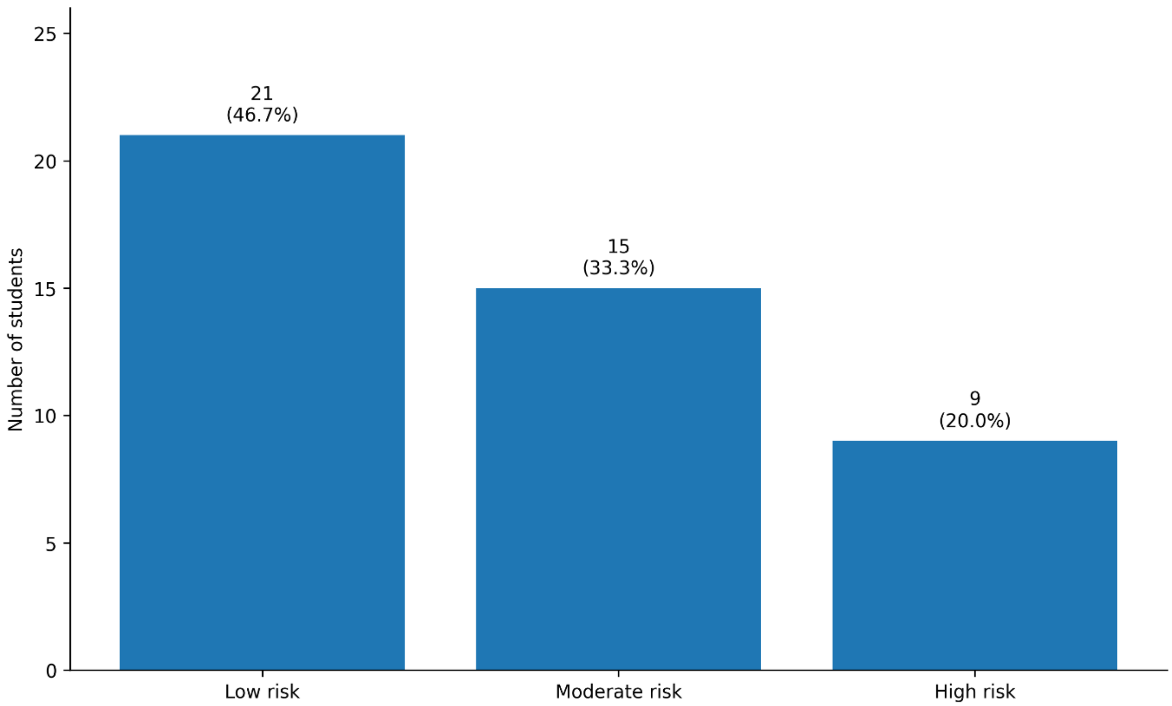

Figure 1.
Distribution of students across ordinal dropout-risk categories.

Figure 1 should show the number and percentage of students in each category. A bar chart is the most suitable option because it communicates the global structure of the group immediately and clearly.

4.2. Distribution of Students by Risk Category

The final classification assigned 21 students (46.7%) to the low-risk category, 15 students (33.3%) to the moderate-risk category, and 9 students (20.0%) to the high-risk category. This pattern suggests that, although nearly half of the students present relatively stable profiles, more than half of the group shows some degree of vulnerability, and one-fifth of the cohort is already in a clearly critical position.

Table 1.
Distribution of students by dropout-risk category

Risk category	***Number of students***	***Percentage***
Low risk	21	46.7
Moderate risk	15	33.3
High risk	9	20.0
Total	45	100.0

This distribution has an important institutional implication. The cohort does not present a generalized high-risk pattern, but neither can it be interpreted as overwhelmingly stable. The moderate-risk group is large enough to justify systematic follow-up, while the presence of nine high-risk students indicates that the institution should be prepared to prioritize more intensive academic support for a non-trivial subset of the group. From an interpretive standpoint, the observed distribution also suggests that the model is discriminating meaningfully among students rather than collapsing them into a single dominant category. This supports the methodological adequacy of the classification system and reinforces the practical value of using ordered categories instead of undifferentiated warning signals.

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

4.3. Global Interpretation of Student Profiles

The assigned categories should not be interpreted as the consequence of isolated indicators, but as the result of a structured joint evaluation across the hierarchy of criteria. In this application, the global classification reflects the combined influence of three broad dimensions: academic performance, engagement and attendance, and socio-contextual conditions. These dimensions provide a sufficiently comprehensive basis for distinguishing the general level of vulnerability of each student while preserving the interpretability of the model.

Students assigned to the low-risk category generally displayed stable or favorable performance in at least two of the three dimensions, with no severe weakness in the remaining one. Their profiles suggest acceptable academic continuity conditions, even when some minor vulnerabilities were present. Students classified as moderate risk typically showed either a concentrated weakness in one dimension or intermediate deficiencies across two dimensions. These cases were not yet uniformly critical, but their profiles indicated enough instability to justify active academic monitoring. Students allocated to the high-risk category exhibited the most unfavorable combinations of criteria. In most of these cases, weakness was not restricted to a single dimension; rather, the classification was driven by a cumulative pattern of low academic performance, weak engagement, irregular attendance, or adverse contextual conditions. This pattern confirms that the most critical classifications emerge from multidimensional vulnerability rather than from a single negative indicator.

4.4. Partial Results by Hierarchical Dimension

The hierarchical structure of the model allowed examination of the internal composition of the global classifications. This partial analysis revealed that the three dimensions did not contribute equally across the cohort. The most influential dimension in differentiating higher-risk students from lower-risk students was academic performance. Students in the high-risk group tended to have unfavorable evaluations in academic achievement, consistency of performance, or progression indicators. The second most discriminating dimension was engagement and attendance, especially when weak academic results were combined with low participation, irregular attendance, or signs of reduced commitment to academic activities. The third dimension, socio-contextual conditions, also played a significant role, but more often as a reinforcing factor rather than the sole cause of critical classification.
A summary of the dimensional pattern is shown in Table 2.

Table 2.
Summary of student performance by hierarchical dimension.

Dimension	***Favorable***	***Intermediate***	***Unfavorable***
Academic performance	17	16	12
Engagement and attendance	16	18	11
Socio-contextual conditions	21	14	10

This pattern suggests that academic performance and engagement/attendance were the dimensions with the greatest power to separate lower- and higher-risk students, while socio-contextual conditions contributed more selectively by intensifying the vulnerability of students already showing academic or behavioral instability. Figure 2 displays the concentration of favorable, intermediate, and unfavorable results by student and by dimension.

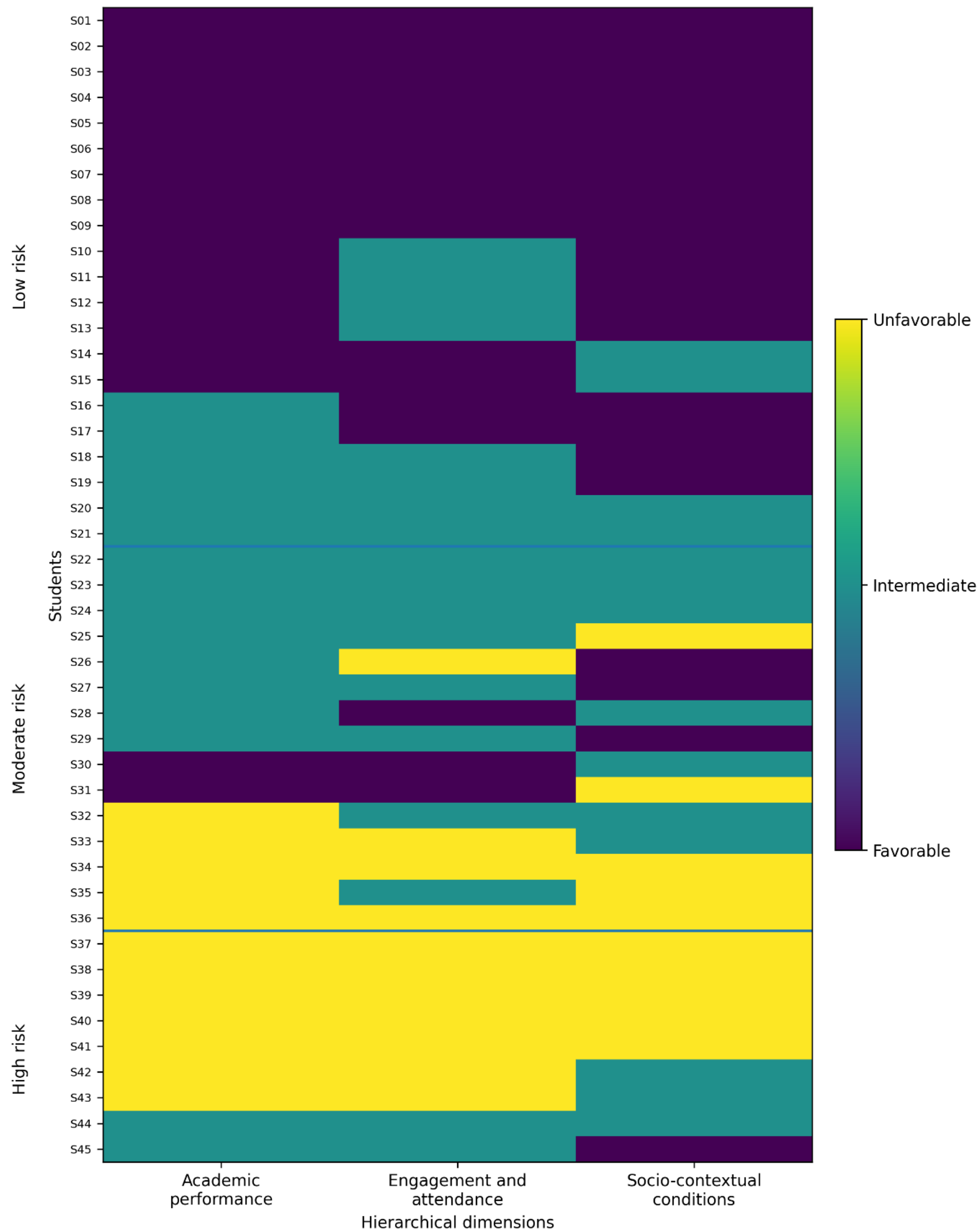

Figure 2.
Heat map of student performance across the three main dimensions.

4.5. Differentiation Between Lower-Risk and Higher-Risk Cases

A clearer comparison of the class profiles shows that the low-risk and high-risk groups differ not only in degree but also in structure. The low-risk group is characterized by balanced profiles. Most of these students showed favorable results in academic performance and at least intermediate conditions in engagement and socio-contextual variables. Their classification reflects overall stability rather than exceptional performance in every criterion. The moderate-risk group is more heterogeneous. In this category, two recurring patterns were observed. The first pattern corresponds to students with acceptable academic performance but inconsistent engagement or attendance. The second pattern corresponds to students with intermediate academic performance and mild contextual disadvantages that, in combination, move them away from the low-risk class. This heterogeneity confirms that the moderate category is not merely a residual class but an analytically important space where early academic deterioration becomes visible. The high-risk group showed the clearest cumulative disadvantage. Seven of the nine students in this category presented unfavorable assessments in both academic performance and engagement/attendance, while five of them also showed unfavorable socio-contextual conditions. This suggests that the most critical cases are associated with compounding weaknesses across dimensions rather than with single isolated deficits. The high-risk class, therefore, represents a substantive concentration of vulnerability and not just a marginal extension of the moderate category. Figure 3 illustrates three representative profiles, one from each category.

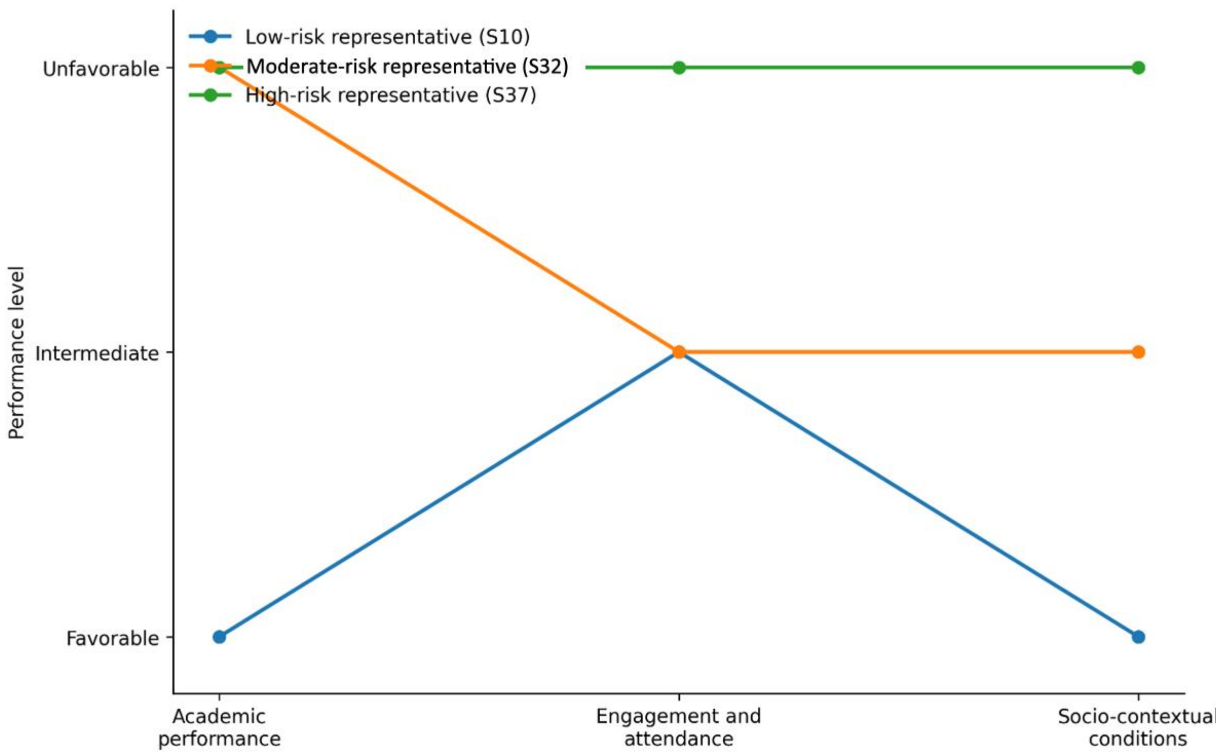

Figure 3.
Comparative profile plot for representative students from low-, moderate-, and high-risk categories.

4.6. Robustness of Classifications Under Uncertainty

An important strength of the proposed model is that it does not treat all assignments as equally certain. Because the methodology admits uncertainty in performance representation and infers preference parameters by disaggregation, the results can be analyzed in terms of robustness, distinguishing between classifications that remain stable across admissible model parameterizations and those that are more sensitive to uncertainty.

In the present application, 34 students (75.6%) showed necessary assignments, meaning their classification remained unchanged across the admissible preference structure. The remaining 11 students (24.4%) were borderline cases with assignments that could vary between two adjacent categories depending on the compatible parameterization. Of these borderline cases, 7 students were located between the low- and moderate-risk classes, and 4 students between the moderate- and high-risk classes.

Table 3.
Robustness structure of the class assignments.

Assignment status	**Number of students**	**Percentage**
Necessary assignment	34	75.6
Borderline low–moderate	7	15.6
Borderline moderate–high	4	8.9
Total	45	100.0

These results are highly informative. The fact that three-quarters of the students have robust assignments indicates that the model produces stable and interpretable classifications for most of the cohort. At the same time, the presence of 11 borderline cases is analytically useful rather than problematic because it identifies the students whose situations are more sensitive and therefore require more careful follow-up. Figure 4 shows the robustness pattern of the classifications. An interval plot or band plot is used because it can show, for each student, the assigned category and the range of admissible adjacent categories when uncertainty is present.

ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

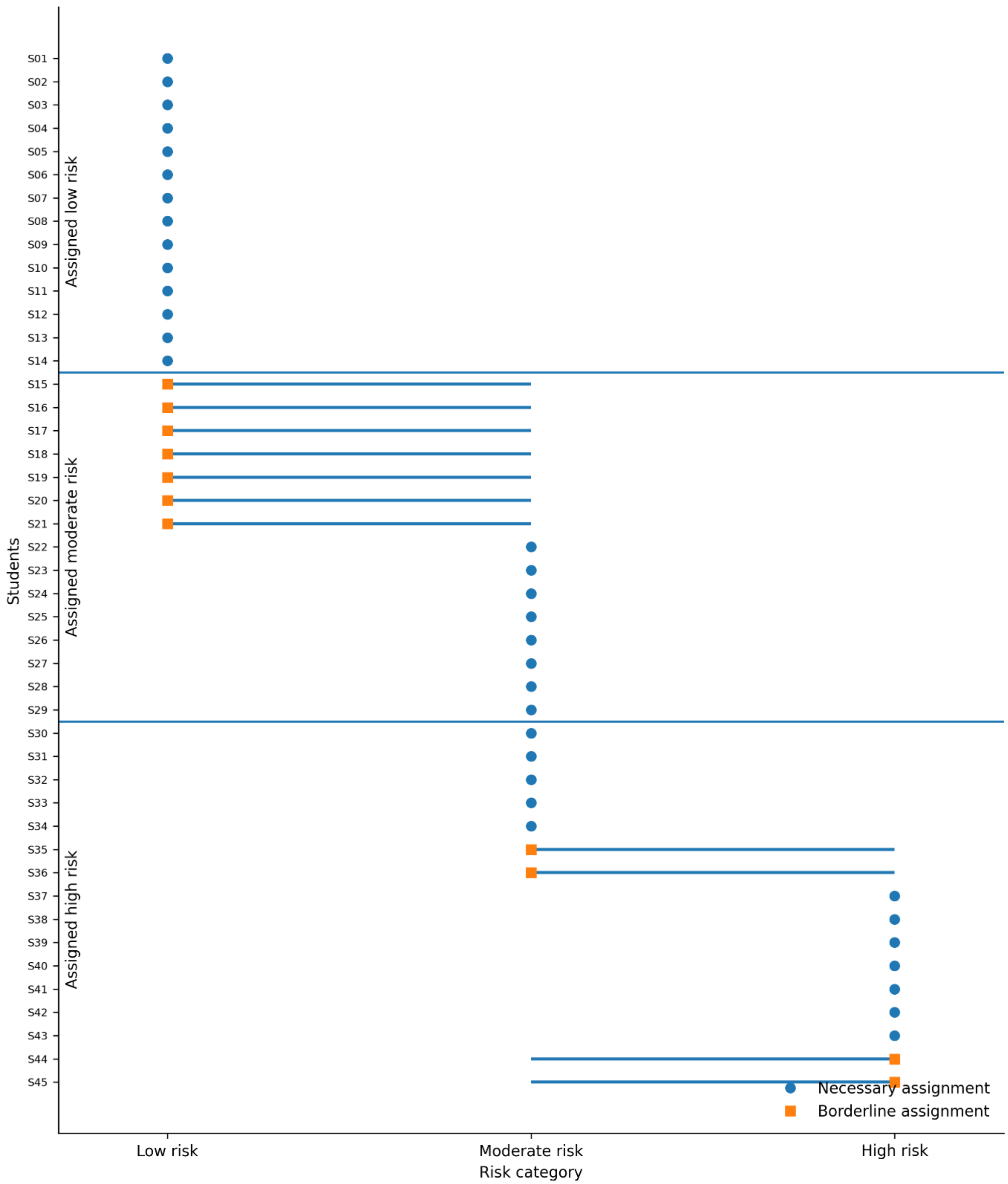

Figure 4.
Necessary and possible class assignments for the 45 students.

The results support a meaningful distinction between priority cases and borderline cases. The priority cases correspond to the seven students whose high-risk assignment was robust across the admissible parameter space. These students constitute the most urgent target group for academic follow-up because the model consistently identifies them as belonging to the most critical class. Their profiles showed

Contemporary Research in Education and English Language Teaching
ISSN: 2641-0230
Vol. 7, No. 2: 1-19, 2025
DOI: 10.55214/2641-0230.v8i1.12872

persistent academic weakness, unfavorable engagement patterns, and, in several cases, reinforcing contextual vulnerability. The borderline cases are equally important, although for different reasons. The seven students located between low and moderate risk represent a preventive intervention group. Their profiles suggest that deterioration is still limited and that timely support could help maintain or restore a lower-risk condition. The four students located between moderate and high risk require even closer attention, since they are positioned at the threshold of the most critical category and could easily consolidate a high-risk profile if no action is taken. These findings show that the model is not useful only for identifying the most vulnerable students; it is also valuable for detecting unstable cases that may benefit most from early intervention. In practical terms, this distinction is essential for a rational allocation of academic support efforts.

4.7. Main Recommendations Derived from the Results

The results support five main recommendations for academic monitoring.

First, the institution should adopt a tiered monitoring strategy aligned with the ordinal structure of the classifications. The low-risk group should remain under periodic preventive observation. The moderate-risk group should receive targeted follow-up focused on the specific dimensions in which vulnerability is concentrated. The high-risk group, particularly the seven robust high-risk cases, should be prioritized for immediate and more intensive support.

Second, the institution should interpret the classifications together with the dimensional results. The global class alone is not sufficient for intervention design. Students with similar overall risk levels may require different forms of support depending on whether their vulnerability is concentrated in academic performance, engagement/attendance, or socio-contextual conditions. Therefore, follow-up actions should be dimension-sensitive rather than category-only.

Third, the institution should create a specific protocol for the 11 borderline cases. These students are strategically important because they occupy zones where academic deterioration is either emerging or close to becoming critical. A short-cycle review process, including updated data collection and advisor follow-up, would be especially appropriate for this group.

Fourth, the results indicate that academic performance and engagement/attendance should be the first dimensions addressed in preventive interventions. These were the two dimensions with the greatest differentiating power between low- and high-risk students. Consequently, tutoring, academic progress review, attendance tracking, and engagement reinforcement appear to be the most immediately justified forms of institutional response.

Fifth, the institution should use the model as a recurrent monitoring tool rather than as a one-time screening device. The present application is cross-sectional, but the classification framework is well-suited for repeated implementation. Reapplying the model periodically would allow the institution to observe movement across categories, identify whether borderline students improve or deteriorate, and assess whether support measures are associated with more stable profiles over time.

5. Conclusions

This study developed and applied a multicriteria ordinal classification model to assess student dropout risk in a structured, interpretable, and uncertainty-sensitive manner. The results show that the proposed approach is suitable for transforming heterogeneous student information into meaningful, ordered risk categories that can support academic monitoring and institutional decision-making. Rather than reducing the problem to a single score or a binary alert, the model made it possible to classify students into differentiated levels of vulnerability and interpret those classifications through a hierarchy of analytical dimensions.

The application to a group of 45 students demonstrated that dropout risk is better understood as a multidimensional condition than as the result of one isolated factor. The findings indicated that academic performance and engagement/attendance were the dimensions with the greatest discriminatory power, while socio-contextual conditions often reinforced the vulnerability of students who were already showing academic or behavioral instability. This confirms the relevance of using a hierarchical multicriteria structure, since it allows the analysis to move beyond the simple identification of at-risk students and toward a more precise understanding of the dimensions underlying each classification. A second important conclusion is that the proposed model has clear practical value as a decision-support tool. Its usefulness lies not only in assigning each student to a risk category but also in distinguishing between robust and borderline cases. This feature is especially relevant for academic follow-up because it helps institutions prioritize intervention efforts according to both the severity and the stability of the classifications. In this sense, the model supports a more rational allocation of institutional attention by differentiating between students who require preventive monitoring, those who need targeted support, and those who should be treated as priority cases. The study also confirms that interpretability is a major advantage of the multicriteria ordinal approach. In educational settings, the value of a classification model depends not only on its analytical consistency but also on its capacity to provide results that are understandable and usable by academic staff. The proposed framework contributes in this regard by preserving the ordered nature of dropout risk, making explicit the role of the evaluation criteria, and allowing the results to be examined both globally and by partial dimensions. Consequently, the model offers a more transparent alternative to purely black-box predictive schemes when the institutional objective is academic follow-up and intervention planning.

Despite these contributions, the study has limitations that should be acknowledged. First, the application was conducted on a specific group of 45 students, which means that the objective was demonstrative and analytical rather than statistically generalizable. Second, the model was implemented in a cross-sectional way, so the classifications represent the academic condition of the students at a given point in time rather than changes in their trajectories. Third, although the model incorporates uncertainty through interval-valued performances and disaggregation-based preference inference, its practical performance would benefit from future applications using other longitudinal data and repeated observations. Accordingly, future research could extend this work in at least three directions. First, the model could be applied to larger student populations to examine its stability across different educational contexts. Second, a longitudinal implementation could make it possible to observe transitions between risk categories over time and evaluate the sensitivity of the model to changes in student conditions. Third, the integration of the multicriteria framework with institutional early warning systems could strengthen the link between classification, follow-up, and intervention outcomes.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

Aina, C., Baici, E., Casalone, G., & Pastore, F. (2022). The determinants of university dropout: A review of the socio-economic literature. *Socio-Economic Planning Sciences*, *79*, 101102. https://doi.org/10.1016/j.seps.2021.101102

Alvarez, P. A., Ishizaka, A., & Martínez, L. (2021). Multiple-criteria decision-making sorting methods: A survey. *Expert Systems with Applications*, *183*, 115368. https://doi.org/10.1016/j.eswa.2021.115368

Bañeres, D., Rodríguez-González, M. E., Guerrero-Roldán, A.-E., & Cortadas, P. (2023). An early warning system to identify and intervene online dropout learners. *International Journal of Educational Technology in Higher Education*, *20*(1), 3. https://doi.org/10.1186/s41239-022-00371-5

Barragán Moreno, S. P., & González Támara, L. (2024). Complexities of student dropout in higher education: A multidimensional analysis. *Frontiers in Education*, *9*. https://doi.org/10.3389/feduc.2024.1461650

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, *55*(4), 485-540. https://doi.org/10.3102/00346543055004485

Carballo-Mendívil, B., Arellano-González, A., Ríos-Vázquez, N. J., & Lizardi-Duarte, M. d. P. (2025). Predicting student dropout from day one: XGBoost-based early warning system using pre-enrollment data. *Applied Sciences*, *15*(16), 9202. https://doi.org/10.3390/app15169202

Chang, Y.-H., Chen, F.-C., & Lee, C.-I. (2025). Developing an early warning system with personalized interventions to enhance academic outcomes for at-risk students in Taiwanese higher education. *Education Sciences*, *15*(10), 1321. https://doi.org/10.3390/educsci15101321

Corrente, S., De Smet, Y., Doumpos, M., Greco, S., & Zopounidis, C. (2023). Classification, sorting and clustering methods based on multiple criteria: recent trends. *Annals of Operations Research*, *325*(2), 767-770. https://doi.org/10.1007/s10479-023-05356-9

Corrente, S., Greco, S., & Słowiński, R. (2016). Multiple Criteria Hierarchy Process for ELECTRE Tri methods. *European Journal of Operational Research*, *252*(1), 191-203. https://doi.org/10.1016/j.ejor.2015.12.053

de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., Garcez, T. V., & Cavalcante, C. A. V. (2015). *Multicriteria and multiobjective models for risk, reliability and maintenance decision analysis*. Cham, Switzerland: Springer.

González-Morales, M. O., López-Aguilar, D., Álvarez-Pérez, P. R., & Toledo-Delgado, P. A. (2025). Dropping out of higher education: Analysis of variables that characterise students who interrupt their studies. *Acta Psychologica*, *252*, 104669. https://doi.org/10.1016/j.actpsy.2024.104669

Greco, S., Mousseau, V., & Słowiński, R. (2010). Multiple criteria sorting with a set of additive value functions. *European Journal of Operational Research*, *207*(3), 1455-1470. https://doi.org/10.1016/j.ejor.2010.05.021

Greco, S., Słowiński, R., & Wallenius, J. (2025). Fifty years of multiple criteria decision analysis: From classical methods to robust ordinal regression. *European Journal of Operational Research*, *323*(2), 351-377. https://doi.org/10.1016/j.ejor.2024.07.038

Liao, H., Xiao, Y., Wu, Z., & Wen, Z. (2024). A review of multiple criteria sorting methods: bibliometrics, characteristics, applications and prospects. *Journal of Control and Decision*, *11*(3), 396-408. https://doi.org/10.1080/23307706.2023.2296120

OECD. (2024). *Education at a Glance 2024: OECD indicators*. Paris, France: OECD Publishing.

Quimiz-Moreira, M., Delgadillo, R., Parraga-Alava, J., Maculan, N., & Mauricio, D. (2025). Factors, prediction, explainability, and simulating university dropout through machine learning: A systematic review, 2012–2024. *Computation*, *13*(8), 198. https://doi.org/10.3390/computation13080198

Quincho Apumayta, R., Carrillo Cayllahua, J., Ccencho Pari, A., Inga Choque, V., Cárdenas Valverde, J. C., & Huamán Ataypoma, D. (2024). University dropout: A systematic review of the main determinant factors (2020-2024). *F1000Res*, *13*, 942. https://doi.org/10.12688/f1000research.154263.2

Rahmani, A. M., Groot, W., & Rahmani, H. (2024). Dropout in online higher education: a systematic literature review. *International Journal of Educational Technology in Higher Education*, *21*, 19. https://doi.org/10.1186/s41239-024-00450-9

Ru, Z., Liu, J., Kadziński, M., & Liao, X. (2023). Probabilistic ordinal regression methods for multiple criteria sorting admitting certain and uncertain preferences. *European Journal of Operational Research*, *311*(2), 596-616. https://doi.org/10.1016/j.ejor.2023.05.007

Tinto, V. (2012). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago, IL, USA: University of Chicago press.

UNESCO Institute for Statistics. (2024). *Education & literacy*. Retrieved from https://www.uis.unesco.org/en/themes/education-literacy

Venkatesan, R. G., Karmegam, D., & Mappillairaju, B. (2024). Exploring statistical approaches for predicting student dropout in education: a systematic review and meta-analysis. *Journal of Computational Social Science*, *7*(1), 171-196. https://doi.org/10.1007/s42001-023-00231-w