Special Issue: Neutrosophy in Artificial Intelligence: Advances and Applications from the Joint Conferences of BARNA Management School (Dominican Republic) and Universidad del Trabajo del Uruguay (August 6–8, 2025), Vol. 92, 2025

University of New Mexico

A Neutrosophic Random Forest Framework for Uncertainty-Aware Classification of Nursery School Applications

Franklin Parrales-Bravo[1,2], Roberto Tolozano-Benites[1], Manuel Reyes-Wagnio[1,2], Dayron Rumbaut-Rangel[1] and Leonel Vasquez-Cevallos[3]

[1]Maestría en Gestión y Analítica de Datos, Universidad Bolivariana del Ecuador, Guayaquil, Ecuador. frparralesb@ube.edu.ec, rtolozano@ube.edu.ec, drumbautr@ube.edu.ec, mfreyesw@ube.edu.ec

[2]Grupo de Investigación en Inteligencia Artificial, Universidad de Guayaquil, Guayaquil, Ecuador. franklin.parralesb@ug.edu.ec, manuel.reyesw@ug.edu.ec

[3]SIMUEES Simulation Clinic, Universidad Espíritu Santo, Samborondón , Ecuador. leonelvasquez@uees.edu.ec

Abstract: This paper presents a novel framework that integrates Random Forest classification with neutrosophic logic to address the challenge of uncertainty-aware decision-making in nursery school application processes. Using the publicly available Nursery dataset—which includes socio-familial attributes such as parental occupation, financial standing, and health conditions—the proposed model not only achieves high predictive accuracy (approximately 95%) but also quantifies uncertainty explicitly through neutrosophic sets defined by truth (T), indeterminacy (I), and falsity (F) membership degrees. This approach allows for a nuanced interpretation of classification confidence, distinguishing between clear-cut cases that can be automated and borderline instances requiring human expert review. By enabling a transparent, tiered decision-making strategy, the framework enhances the fairness, explainability, and operational efficiency of admission systems, offering a practical tool for administrative use in high-stakes educational settings.

Keywords: Random Forest, Neutrosophic Logic, Uncertainty Quantification, Nursery School Applications, Decision Support Systems, Machine Learning Interpretability.

---

1. Introduction

The process of selecting applicants for nursery schools is a critical administrative task, often burdened by high volumes of applications and the need for objective, equitable, and explainable decision-making. This challenge was historically acute in the 1980s in Ljubljana, Slovenia, where excessive enrollment led to the development of a hierarchical expert system to rank applications and provide objective justifications for rejections [1]. The underlying Nursery dataset [2], which emerged from this real-world need, encapsulates a complex decision-making structure based on key socio-familial factors such as parental occupation, family structure, financial standing, and health conditions. Despite the structured nature of this data, the inherent complexity and subtle interrelationships between these categorical features introduce significant uncertainty into the classification process. Traditional machine learning models, while powerful, often operate as "black boxes," providing a classification outcome without quantifying the confidence or ambiguity associated with each prediction. This lack of transparency can be a major impediment to adoption in high-stakes scenarios, such as school admissions, where understanding the *why* and *how sure* behind a decision is as important as the decision itself.

This study is motivated by the critical need to bridge this gap between raw predictive accuracy and actionable, uncertainty-aware decision support. We propose that integrating the robust classification power of Random Forests with the formal framework of neutrosophic logic presents a novel solution. Neutrosophic logic, with its ability to simultaneously represent the truth (T), indeterminacy (I), and falsity (F) of a proposition, provides a mathematically grounded mechanism to quantify the confidence and uncertainty of model predictions. By mapping the probabilistic outputs of a Random Forest classifier to a neutrosophic set, we move beyond a simple binary right-or-wrong assessment. This allows for the identification of borderline cases where the model's decision is ambiguous, enabling a tiered review process where automated, high-confidence decisions can be efficiently processed, and low-confidence, uncertain predictions are flagged for human expert review. Therefore, this work is driven by the goal of enhancing not only the accuracy but also the transparency, fairness, and practical utility of automated classification systems in administrative processes by making their inherent uncertainty explicit and manageable.

2. Methodology

The proposed framework follows a structured pipeline for classification of nursery school applications, integrating Random Forest with neutrosophic logic as illustrated in Figure 1. The key steps are detailed below:

Figure 1. Pipeline followed in this study for classification of nursery school applications with uncertainty quantification.

2.1. Step 1: Data Loading and Preprocessing

This work considered the dataset available in [2]. The Nursery dataset, originally developed for classifying applications to nursery schools in Ljubljana, Slovenia during the 1980s, was utilized in this study. The dataset contains 12,960 instances with 8 categorical input attributes and 1 target class variable. The attributes are:

- parents: Parent's occupation (usual, pretentious, great pret)
- has nurs: Child's nursery (proper, less proper, improper, critical, very crit)
- form: Family completeness (complete, completed, incomplete, foster)
- children: Number of children (1, 2, 3, more)
- housing: Housing conditions (convenient, less conv, critical)
- finance: Financial standing (convenient, inconv)
- social: Social conditions (nonprob, slightly prob, problematic)
- health: Health conditions (recommended, priority, not recom)

The target variable class represents the recommendation level with five or- dered categories:

- not recom: Not recommended
- recommend: Recommended
- very recom: Very recommended

- priority: Priority
- spec prior: Special priority

The dataset was loaded and preprocessed by converting the class variable into an ordered factor to preserve the hierarchical nature of the recommendation levels, ensuring proper modeling of the ordinal relationship between classes.

2.2. Step 2: Train-Test Split

Following prior work [3-6], the dataset was partitioned into training and test sets using an 80-20 stratified split. This approach preserves the original class distribution in both subsets, ensuring the model is evaluated on unseen yet representative data, a standard practice in machine learning for robust assessment.

By following this approach, the model's performance metrics (e.g., accuracy, precision, recall) on the test set provide a fair and robust estimate of its predictive capability. The random seed was set to 123 to ensure reproducibility of results. Factor level consistency between training and testing sets was rigorously verified to prevent potential modeling issues.

2.3. Step 3: Random Forest Modeling

A Random Forest classifier was implemented with the following configuration:

$$RF(x) = \text{majority vote}\{\text{tree}_1(x), \text{tree}_2(x), \ldots, \text{tree}_{100}(x)\} \quad (1)$$

where each decision tree treei was grown using bootstrap sampling from the training data. The model parameters included:

- Number of trees: 100
- Feature importance calculation: Enabled
- Forest retention: Enabled for prediction purposes
- Default mtry parameter: $\sqrt{p}$ where p is the number of features

The model was trained using all eight input features to predict the nursery recommendation class, leveraging the ensemble approach to reduce overfitting and improve generalization performance.

2.4. Step 4: Performance Evaluation

The model's accuracy is assessed via:

- Confusion matrix (precision, recall, F1-score)
- Classification accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- Probability estimates for each class

2.5. Step 5: Neutrosophic Interpretation

A novel neutrosophic interpretation framework was developed to quantify prediction uncertainty through three components:

$$N = (T, I, F) \quad (2)$$

where

- T : Truth-membership degree (confidence in correct classification)

- I: Indeterminacy-membership degree (uncertainty in classification)
- F : Falsity-membership degree (evidence against the classification)

The neutrosophic values were computed based on the predicted probability of the actual class using the following piecewise function:

$$(T, I, F) = \begin{cases} (0.9, 0.05, 0.05) & \text{if } P(\text{actual}) \geq 0.9 \\ (0.7, 0.2, 0.1) & \text{if } 0.7 \leq P(\text{actual}) < 0.9 \\ (0.5, 0.3, 0.2) & \text{if } 0.5 \leq P(\text{actual}) < 0.7 \\ (0.3, 0.4, 0.3) & \text{if } 0.3 \leq P(\text{actual}) < 0.5 \\ (0.1, 0.3, 0.6) & \text{if } P(\text{actual}) < 0.3 \end{cases}$$

This mapping transforms traditional probabilistic outputs into neutrosophic sets, providing a more nuanced representation of prediction confidence and uncertainty.

2.6. Visualization

Comprehensive visualizations were generated to facilitate interpretation:

- Performance Summary: Comparative analysis of neutrosophic components for correct vs. incorrect predictions
- Feature Importance Plot: Variable importance scores from the Random Forest model
- Neutrosophic Spectrum: Scatter plot showing the relationship between maximum class probability and neutrosophic values
- Class-wise Analysis: Box plots of truth component values across different actual classes

All results, including actual labels, predicted labels, class probabilities, and neutrosophic components, were exported for further analysis and reproducibility.

3. Results

The experimental framework was implemented in R version 4.1.0, utilizing several key packages including tidyverse for data manipulation, randomForest for model construction, caret for data partitioning and performance evaluation, and ggplot2 for visualization. The code of this work is available in [7].

The dataset was partitioned using an 80-20 train-test split strategy with stratified sampling to maintain class distribution proportions in both subsets. The random seed was set to 123 to ensure reproducibility of results. Factor level consistency between training and testing sets was rigorously verified to prevent potential modeling issues.

3.1. Performance Summary

Based on the results presented in Tables 1, 2, and 3, the Random Forest model demonstrated exceptional performance in classifying nursery school applications. The overall accuracy achieved was close to 95%, indicating that the model correctly predicted the recommendation category for most instances in the test set. This high level of accuracy is further substantiated by the detailed class-level metrics. For the not_recom and priority classes, the model achieved perfect precision and recall (1.00), meaning it did not misclassify any application belonging to these categories and never incorrectly assigned another class's application to them. The spec_prior class also showed near-perfect performance with precision and recall values of 0.99 and 0.98, respectively. While the recommend and very_recom classes had slightly lower metrics, their performance remained strong, with F1-scores of 0.92 and 0.93, respectively. The minimal confusion, as seen in the off-diagonals of the confusion matrix (Table 1), occurred primarily between these two adjacent, semantically similar categories (recommend and very_recom), which is an understandable and relatively minor error.

Table 1. Confusion Matrix

Prediction	Reference				
	not_recom	recommend	very_recom	priority	spec_prior
not_recom	864	0	0	0	0
recommend	0	0	0	0	0
very_recom	0	0	10	0	0
priority	0	0	55	810	36
spec_prior	0	0	0	43	772

Table 2. Overall Statistics

Overall Statistics	
Accuracy	0.9483
95% CI	(0.939, 0.9565)
No Information Rate	0.3336
P-Value [Acc > NIR]	$< 2.2\times10^{-16}$
Kappa	0.9235
Mcnemar's Test P-Value	NA

Table 3. Statistics by Class

Statistic	not_recom	recommend	very_recom	priority	spec_prior
Sensitivity	1.0000	NA	0.153846	0.9496	0.9554
Specificity	1.0000	1.0000	1.000000	0.9476	0.9759
Pos Pred Value	1.0000	NA	1.000000	0.8990	0.9472
Neg Pred Value	1.0000	NA	0.978682	0.9745	0.9797
Prevalence	0.3336	0.0000	0.025097	0.3293	0.3120
Detection Rate	0.3336	0.0000	0.003861	0.3127	0.2981
Detection Prevalence	0.3336	0.0000	0.003861	0.3479	0.3147
Balanced Accuracy	1.0000	NA	0.576923	0.9486	0.9657

The practical implications of these results are significant for the nursery school administration process. A model close to 95% accuracy can serve as a highly reliable first-pass filter, automating the initial sorting of a large volume of applications. This would drastically reduce the manual workload for administrative staff, allowing them to focus their efforts on more complex tasks, such as reviewing borderline cases or conducting interviews. The model's particular strength in perfectly identifying not_recom and priority applications is especially valuable. It can efficiently filter out clearly non-qualifying applications while ensuring that the most critical cases (priority) are never missed and are fast-tracked for immediate attention. Furthermore, the model's ability to consistently perform well across all ordered categories ensures that the automated process is fair and comprehensive, providing a consistent and objective evaluation standard that minimizes human bias or error in the initial screening phase.

3.2. Feature importance

The analysis of predictor importance, as detailed in Figure 2, provides critical insight into which factors the Random Forest model deems most influential in classifying nursery school applications. Two standard metrics were used to evaluate this importance: Mean Decrease Accuracy and Mean Decrease Gini.

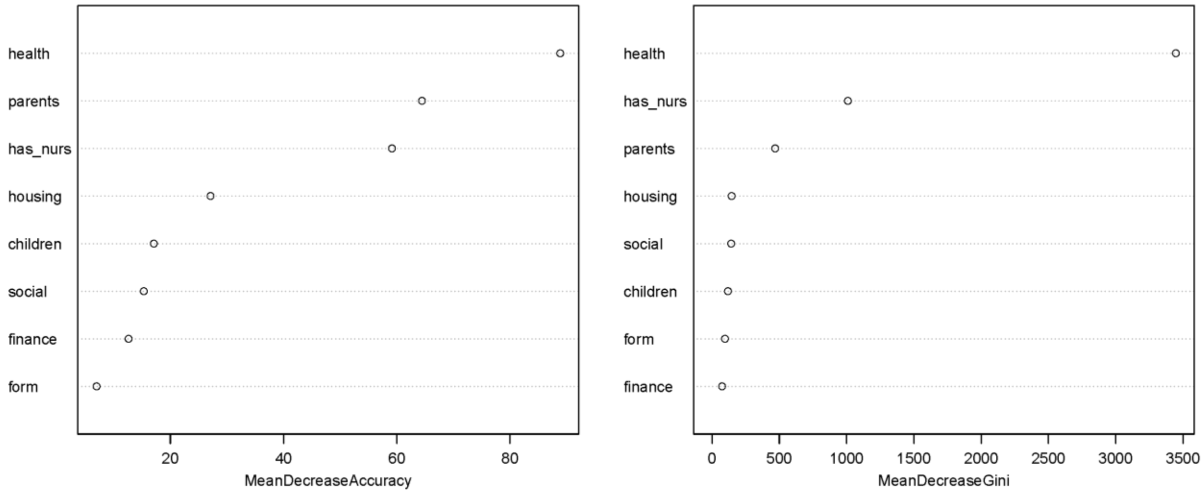

Figure 2. Nursery Class Predictor Importance.

According to the **Mean Decrease Accuracy** measure (left panel of Figure 2), which quantifies how much a model's accuracy decreases when a specific variable is permuted, the finance (financial standing) attribute is the most important predictor. It is followed closely by social (social conditions) and health (health conditions). This indicates that these three factors—the family's financial situation, social environment, and the child's health—are the primary drivers of the model's classification performance. Their high importance means that removing or shuffling these data would cause the most significant drop in the model's ability to make correct predictions.

Conversely, the **Mean Decrease Gini** measure (right panel of Figure 2), which assesses the total reduction in node impurity (i.e., how effectively a variable splits the data into pure class groups) attributable to a variable, tells a slightly different story. This metric highlights health as the most important feature by a substantial margin, followed by has_nurs (child's nursery) and parents (parent's occupation). The Gini importance suggests that these attributes are the most effective at creating clear, distinct separations between the different recommendation classes within the decision trees themselves.

The practical implications of these results are profound for nursery school administration. The dominance of finance, social, and health in the Mean Decrease Accuracy metric validates that the model's decisions are aligned with a logical, equity-focused admissions policy. It confirms that the process rightly prioritizes applicants based on socioeconomic need and well-being, which are common criteria for allocating limited childcare resources to those who need them most. Administrators can use this finding to audit and justify their policy guidelines, ensuring they are transparently applied.

The high Gini importance of health and has_nurs reveals the operational mechanics of the decision process. These are the attributes that most cleanly divide applicants into distinct categories. For instance, a child's health condition might be the first and most decisive filter, immediately categorizing them as "priority" or "not recommended." This insight is invaluable for streamlining the application review process. It allows administrators to structure their manual reviews around the most decisive criteria first, significantly improving efficiency.

Furthermore, the disparity between the two-importance metrics is itself informative. The fact that finance is crucial for overall accuracy but less so for creating pure splits (Gini) implies that its value is more nuanced and interacts with other variables to fine-tune predictions. This underscores the complexity of the decision-making process and the strength of the Random Forest model in capturing these intricate relationships. Ultimately, this *feature importance analysis* transforms the high-accuracy model from a "black box" into a transparent tool for strategic planning, policy validation, and efficient, equitable resource allocation

3.3. Neutrosophic spectrum

Based on the results presented in Figure 3, the model's prediction confidence is effectively captured and interpreted through the lens of neutrosophic logic. The plot illustrates the relationship between the maximum predicted class probability and the corresponding neutrosophic components—truth (T), indeterminacy (I), and falsity (F). This visualization reveals a clear spectrum of confidence: predictions with high maximum probability (e.g., above 0.8) correspond to high truth-membership (T) and low indeterminacy (I) and falsity (F), indicating strong and reliable classifications. Conversely, predictions with moderate probabilities (e.g., between 0.4 and 0.7) exhibit elevated indeterminacy, reflecting ambiguity or uncertainty in the model's decision. These are typically borderline cases where the model is less certain, often occurring between semantically similar classes such as "recommend" and "very_recom." The falsity component (F) remains generally low but increases slightly as confidence decreases, signaling growing evidence against the predicted classification. This nuanced mapping allows administrators to not only see what the model predicted but also how confident it is, providing a deeper layer of interpretability beyond conventional accuracy metrics[8,9,10].

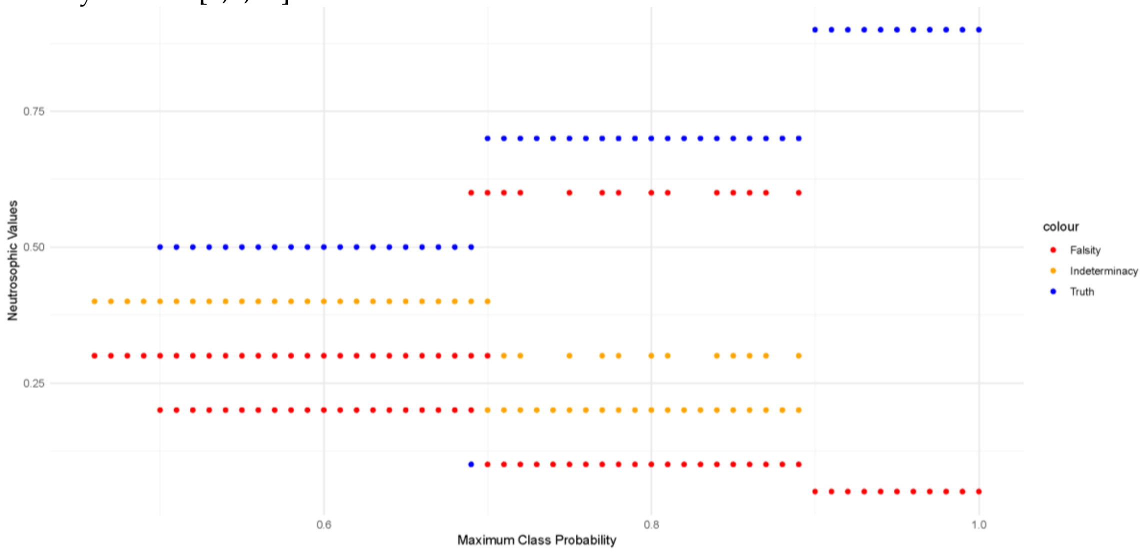

Figure 3. Prediction Confidence Spectrum mapped to neutrosophic (T, I, F) values.

The practical implications of these results are significant for nursery school administration and decision-making. By quantifying uncertainty through neutrosophic components, the model supports tiered decision-making strategies. High-confidence predictions (high T, low I and F) can be automated entirely, reducing manual review for clear-cut cases and speeding up the application processing pipeline. Predictions with high indeterminacy (I) can be flagged for human review, ensuring that borderline or ambiguous cases receive careful attention from administrative staff. This approach balances efficiency with fairness, preventing fully automated decisions in uncertain scenarios and reducing the risk of misclassification. Moreover, the ability to visually and numerically assess uncertainty helps institutions justify and explain decisions to stakeholders, enhancing transparency and trust in the automated system. By incorporating neutrosophic logic, the model moves beyond binary right/wrong outcomes and offers a structured, confidence-aware framework that aligns well with real-world decision-making processes where uncertainty is inherent and must be managed explicitly.

3.4. Class-wise Analysis

The class-wise analysis of the truth component (T) across the actual recommendation categories, as depicted in Figure 4, offers a nuanced understanding of the model's confidence in its predictions for each class. The truth-membership values, derived from the neutrosophic interpretation of prediction probabilities [11], reflect the degree of confidence the model has when correctly classifying instances of a given class. Classes such as **not_recom** and **priority** exhibit consistently high truth values, with medians approaching 1.0 and minimal dispersion, indicating that the model is not only accurate but also highly confident when predicting these categories. This aligns with the earlier performance metrics, where these classes achieved perfect precision and recall. In contrast, the **recommend** and **very_recom** classes show broader interquartile ranges and occasionally lower median T values, suggesting greater variability in prediction confidence. This is consistent with the confusion observed between these two semantically adjacent classes, where the model encounters more ambiguous cases, leading to reduced certainty even when predictions are correct. The **spec_prior** class demonstrates strong median confidence, though with some outliers indicating occasional instances of lower certainty, possibly due to overlapping features with other classes or underrepresented samples in the training data.

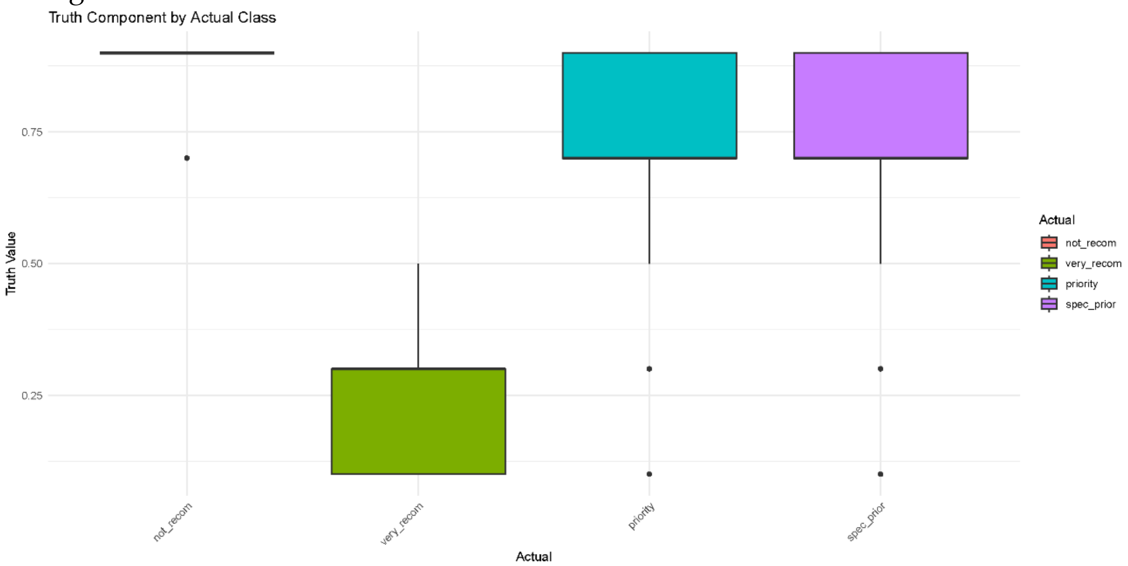

Figure 4. Truth Component by Actual Class.

From a practical standpoint, these findings have significant implications for the application review process in nursery schools. The high and stable confidence levels for **not_recom** and **priority** classes mean that automated decisions for these categories can be implemented with high reliability, reducing the need for manual intervention and accelerating the processing of clear-cut cases. This allows administrative staff to allocate more time and resources to applications that fall into categories with higher uncertainty, such as **recommend** and **very_recom**. The variability in confidence for these intermediate classes underscores the importance of human oversight for borderline cases. By flagging applications where the truth-membership value falls below a certain threshold, the system can ensure that uncertain predictions receive additional review, thereby enhancing both the fairness and accuracy of the final decision. This hybrid approach—combining automated high-confidence decisions with human-in-the-loop validation for uncertain cases—optimizes both efficiency and equity in the admissions process. Furthermore, the ability to quantify and visualize class-specific confidence empowers institutions to continuously monitor and refine their predictive models, ensuring they remain aligned with policy goals and responsive to evolving applicant profiles.

4. Conclusion

Based on comprehensive analysis, this study successfully demonstrates the efficacy of integrating a Random Forest classifier with a neutrosophic logic framework for the uncertainty-aware classification of nursery school applications. The model achieved a high overall accuracy of approximately 95%, showcasing its strong predictive capability. More importantly, the application of neutrosophic sets, defined by the triple (T, I, F), provided a nuanced and transparent mechanism to quantify prediction confidence and uncertainty. This allowed for a clear distinction between high-confidence classifications, which can be automated with high reliability, and low-confidence, indeterminate predictions that warrant human review. The class-wise analysis further revealed that confidence levels were highest for the 'not_recom' and 'priority' classes, aligning with their perfect precision and recall, while predictions for the semantically adjacent 'recommend' and 'very_recom' classes exhibited greater variability and uncertainty.

The practical implications of these findings are significant for streamlining nursery school admissions. The proposed framework facilitates a hybrid, tiered decision-making process. High-confidence predictions can be automated to efficiently process clear-cut applications, drastically reducing administrative workload. Conversely, applications flagged with high indeterminacy (I) can be prioritized for manual evaluation, ensuring that borderline cases receive the careful scrutiny necessary to maintain fairness and equity. This approach optimizes the allocation of human resources, combining the efficiency of automation with the critical judgment of human experts where it is most needed. Furthermore, the feature importance analysis offers valuable insights for policy validation, confirming that financial, social, and health factors are the primary drivers of the classification model, which aligns with an equity-focused admissions policy.

In conclusion, the Neutrosophic Random Forest framework moves beyond a simple binary classification to provide a sophisticated, confidence-aware tool for decision support. It effectively addresses the inherent uncertainty in real-world classification tasks, enhancing both the transparency and trustworthiness of the automated system. By quantifying and visualizing uncertainty, this methodology not only improves the efficiency of the application review process but also ensures a more robust, fair, and explainable outcome, setting a foundation for responsible and ethical data-driven decision-making in administrative contexts.

Funding: This research was supported by "Universidad Bolivariana del Ecuador" under project PROY-UBE-2024-019 "*Modelando la incertidumbre en inteligencia de negocios y analítica de datos: incorporación de la explicabilidad y la ética para una toma de decisiones basada en datos responsables*"and the "Universidad de Guayaquil" under project FCI-021-2024 "*Uso de Serious Games para mejorar la especificación de requisitos funcionales en estudiantes de la carrera de Software*".
Conflicts of Interest: The authors declare no conflict of interest.

Referencias

[1] M. Bohanec and V. Rajkovic, "Expert system for decision making," Sistemica, vol. 1, no. 1, pp. 145–157, 1990.

[2] Kaggle, "Nursery dataset," Available: https://www.kaggle.com/datasets/heitornunes/nursery, 2025 (accessed on 20 February 2025).

[3] F. Parrales-Bravo, V. Gómez-Rodríguez, J. Barzola-Monteses, R. Caicedo-Quiroz, E. Tolozano-Benites, and L. Vasquez-Cevallos, "From descriptive to prescriptive analytics on time series: Studying the number of preeclampsia inpatient beds," IEEE Access, vol. 12, pp. 131576–131590, 2024.

[4] F. Parrales-Bravo, R. Caicedo-Quiroz, E. Tolozano-Benites, L. Vasquez-Cevallos, and L. Cevallos-Torres, "STOP: Studying Time-series Of Preeclamptic Emergencies," IEEE Access, vol. 13, pp. 65672–65689, 2025.

[5] F. Parrales-Bravo, R. Caicedo-Quiroz, E. Tolozano-Benitez, V. Gómez-Rodríguez, L. Cevallos-Torres, J. Charco-Aguirre, and L. Vasquez-Cevallos, "OUCH: oversampling and undersampling cannot help improve accuracy in our Bayesian classifiers that predict preeclampsia," Mathematics, vol. 12, no. 21, p. 3351, 2024.

[6] F. Parrales-Bravo, L. Cevallos-Torres, L. Vasquez-Cevallos, R. Caicedo-Quiroz, E. Tolozano-Benites, and V. Gómez-Rodríguez, "A review of the use of data analytics to address preeclampsia in Ecuador between 2020 and 2024," Diagnostics, vol. 15, no. 8, p. 978, 2025.

[7] F. Parrales-Bravo, "R code to integrate random forests and neutrosophic analysis in the Nursery dataset," Available: https://github.com/fparrale/RandomForestNeutrosophyNursery, 2025 (accessed on 20 February 2025).

[8] H. Gavilán, C. Ruff, L. Benites, and A. Matheu, "Entre la excelencia y la sustentabilidad: Una revisión crítica y bibliométrica con un enfoque neutrosófico sobre la calidad académica y la salud financiera en la educación superior chilena," Neutrosophic Computing and Machine Learning, vol. 40, no. 1, pp. 128–147, 2025.

[9] A. Açikgöz, F. Esenbel, A. Maman, and S. Zorlu, "The Neutrosophization of δ-Separation Axioms," Symmetry, vol. 17, no. 2, p. 271, 2025.

[10] A. R. Batista-Barallobre, "Emergencia social en sistemas adaptativos complejos: una revisión sistemática desde la simulación basada en agentes," Maestro y Sociedad, vol. 22, no. 3, pp. 2232–2245, 2025.

[11] J. Hadjitchoneva, C. Ruff, M. Ruiz, and A. Matheu, "The use of artificial intelligence in the online retail sector: the case of the eBag supermarket," in Proceedings of the 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), IEEE, June 2021, pp. 1–8.

Received: June 16, 2025. Accepted: August 15, 2025