



## Predicción de resultados en educación inicial mediante un modelo de Random Forest neutrosófico: gestión de la incertidumbre en la toma de decisiones pedagógicas

### Predicting outcomes in early childhood education using a neutrosophic Random Forest model: Managing uncertainty in pedagogical decision-making

Franklin Parrales-Bravo <sup>1,2</sup>, Roberto Tolozano-Benitas <sup>1</sup>, Manuel Reyes-Wagnio <sup>1,2</sup>, Dayron Rumbaut-Rangel <sup>1</sup> y Leonel Vásquez-Cevallos <sup>3</sup>

<sup>1</sup> Maestría en Gestión y Analítica de Datos, Universidad Bolivariana del Ecuador, Guayaquil, Ecuador; [frparralesb@ube.edu.ec](mailto:frparralesb@ube.edu.ec), [rtolozano@ube.edu.ec](mailto:rtolozano@ube.edu.ec), [drumbautr@ube.edu.ec](mailto:drumbautr@ube.edu.ec), [mfreyesw@ube.edu.ec](mailto:mfreyesw@ube.edu.ec)

<sup>2</sup> Grupo de Investigación en Inteligencia Artificial, Universidad de Guayaquil, Guayaquil, Ecuador; [franklin.parralesb@ug.edu.ec](mailto:franklin.parralesb@ug.edu.ec), [manuel.reyesw@ug.edu.ec](mailto:manuel.reyesw@ug.edu.ec)

<sup>3</sup> Clínica de Simulación SIMUEES, Universidad Espíritu Santo, Samborondón, Ecuador; [leonelvasquez@uecs.edu.ec](mailto:leonelvasquez@uecs.edu.ec)

#### Resumen:

Este artículo presenta un nuevo marco que integra la clasificación de Bosque Aleatorio con la lógica neutrosófica para abordar el desafío de la toma de decisiones consciente de la incertidumbre en los procesos de solicitud de admisión a escuelas infantiles. Utilizando el conjunto de datos Nursery, disponible públicamente, que incluye atributos sociofamiliares como la ocupación de los padres, la situación financiera y las condiciones de salud, el modelo propuesto no solo logra una alta precisión predictiva (aproximadamente el 95%), sino que también cuantifica la incertidumbre explícitamente a través de conjuntos neutrosóficos definidos por los grados de pertenencia a la verdad (V), la indeterminación (I) y la falsedad (F). Este enfoque permite una interpretación matizada de la confianza de la clasificación, distinguiendo entre casos claros que pueden automatizarse e instancias límite que requieren la revisión de expertos humanos. Al permitir una estrategia de toma de decisiones transparente y escalonada, el marco mejora la imparcialidad, la explicabilidad y la eficiencia operativa de los sistemas de admisión, ofreciendo una herramienta práctica para uso administrativo en entornos educativos de alto riesgo.

**Palabras clave:** Bosque Aleatorio, Lógica Neutrosófica, Cuantificación De La Incertidumbre, Aplicaciones En Escuelas Infantiles, Sistemas De Apoyo A La Toma De Decisiones, Interpretabilidad Del Aprendizaje Automático.

#### Abstract:

This article 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 status, and health conditions, the proposed model not only achieves high predictive accuracy (approximately 95%) but also explicitly quantifies uncertainty through neutrosophic sets defined by degrees of truthfulness (V), indeterminacy (I), and falsity (F). This approach enables a nuanced interpretation of classification confidence, distinguishing between clear cases that can be automated and borderline instances that require human expert review. By enabling a transparent and tiered decision-making strategy, the framework improves the fairness, explainability, and operational efficiency of admissions systems, offering a practical tool for administrative use in high-risk educational settings.

**Keywords:** Random Forest, Neutrosophic Logic, Uncertainty Quantification, Applications in Preschools, Decision Support Systems, Machine Learning Interpretability.

## 1. Introducción

La selección de solicitantes en guarderías constituye una función administrativa crítica, frecuentemente afectada por un alto volumen de postulaciones y la necesidad de decisiones que sean objetivas, equitativas y justificables. Este problema se evidenció de manera significativa en Liubliana (Eslovenia) durante la década de 1980, cuando el exceso de matrículas llevó a la implementación de un sistema experto jerárquico para priorizar solicitudes y explicar objetivamente los rechazos [1]. Los datos generados en este contexto [2] reflejan un proceso complejo basado en múltiples factores sociofamiliares, como ocupación y nivel educativo de los padres, composición del hogar, situación económica y estado de salud de los niños. Aunque estructurados, estos datos presentan interacciones sutiles y dependencias categóricas que introducen un grado considerable de incertidumbre en la clasificación.

Los enfoques tradicionales de aprendizaje automático, aunque precisos, suelen operar como “cajas negras”, ofreciendo resultados de clasificación sin información sobre la confiabilidad o ambigüedad de las predicciones. Esta opacidad limita su aplicabilidad en contextos de alto riesgo, como admisiones escolares, donde conocer el nivel de certeza de una decisión es tan relevante como la decisión misma.

Este estudio propone abordar esta limitación combinando el poder predictivo de los algoritmos Random Forest con la lógica neutrosófica, que representa simultáneamente la verdad (T), la indeterminación (I) y la falsedad (F) de cada proposición. Al transformar las probabilidades del clasificador en valores neutrosóficos, se logra no solo una clasificación binaria, sino también la cuantificación de la incertidumbre asociada. Esto permite identificar casos ambiguos que requieren revisión humana, mientras que las decisiones de alta confianza pueden ejecutarse automáticamente. Así, el enfoque mejora la precisión, transparencia, equidad y aplicabilidad práctica de los sistemas de selección automatizados, haciendo explícita la incertidumbre inherente al proceso.

## 2. Metodología

El marco propuesto sigue un proceso estructurado para la clasificación de aplicaciones en guarderías, integrando Random Forest con lógica neutrosófica, como se ilustra en la Figura 1. Los pasos clave se detallan a continuación:



Figura 1. Proceso seguido en este estudio para la clasificación de solicitudes de escuelas infantiles con cuantificación de incertidumbre.

### 2.1. Paso 1: Carga y preprocesamiento de datos

Este trabajo consideró el conjunto de datos disponible en [2]. En este estudio se utilizó el conjunto de datos Nursery, desarrollado originalmente para clasificar aplicaciones en guarderías de Liubliana, Eslovenia, durante la década de 1980. El conjunto de datos contiene 12 960 instancias con 8 atributos de entrada categóricos y una variable de clase objetivo. Los atributos son:

- padres: Ocupación de los padres (habitual, pretencioso, gran pret)
- tiene enfermeras: Guardería infantil (adecuada, poco adecuada, inadecuada, crítica, muy crítica)
- formulario: Completitud familiar (completo, completado, incompleto, acogida)
- niños: Número de hijos (1, 2, 3, más)
- Vivienda: Condiciones de vivienda (conveniente, menos conveniente, crítica)

- Finanzas: Situación financiera (conveniente, inconveniente)
- social: Condiciones sociales (no prob., ligeramente prob., problemáticas)
- salud: Condiciones de salud (recomendadas, prioritarias, no recomendadas)

La clase de variable objetivo representa el nivel de recomendación con cinco categorías ordenadas:

- no recom: No recomendado
- recomendar: Recomendado
- muy recom: Muy recomendable
- prioridad: Prioridad
- especificación previa: Prioridad especial

El conjunto de datos se cargó y preprocesó convirtiendo la variable de clase en un factor ordenado para preservar la naturaleza jerárquica de los niveles de recomendación, asegurando un modelado adecuado de la relación ordinal entre clases.

## 2.2. Paso 2: División de entrenamiento y prueba

Siguiendo trabajos previos [3-6], el conjunto de datos se dividió en conjuntos de entrenamiento y de prueba mediante una división estratificada de 80-20. Este enfoque conserva la distribución de clases original en ambos subconjuntos, lo que garantiza que el modelo se evalúe con datos no vistos, pero representativos, una práctica estándar en aprendizaje automático para una evaluación robusta.

Con este enfoque, las métricas de rendimiento del modelo (p. ej., exactitud, precisión y recuperación) en el conjunto de prueba proporcionan una estimación justa y robusta de su capacidad predictiva. El valor de semilla aleatorio se estableció en 123 para garantizar la reproducibilidad de los resultados. La consistencia a nivel de factor entre los conjuntos de entrenamiento y de prueba se verificó rigurosamente para evitar posibles problemas de modelado.

## 2.3. Paso 3: Modelado de bosque aleatorio

Se implementó un clasificador Random Forest con la siguiente configuración:

$$RF(x) = \text{mayoría votar} \{ \text{árbol}_1(x), \text{árbol}_2(x), \dots, \text{árbol}_{100}(x) \} \quad (1)$$

Donde cada árbol de decisión treei se desarrolló mediante muestreo bootstrap a partir de los datos de entrenamiento. Los parámetros del modelo incluyeron:

- Número de árboles: 100
- Característica importancia cálculo: habilitado
- Bosque retención : Activado para predicción propósitos
- Por defecto lo intento parámetro:  $\sqrt{p}$  donde  $p$  es el número de características

El modelo se entrenó utilizando las ocho características de entrada para predecir la clase de recomendación de guardería, aprovechando el enfoque de conjunto para reducir el sobreajuste y mejorar el rendimiento de generalización.

## 2.4. Paso 4: Evaluación del desempeño



La precisión del modelo se evalúa mediante:

- Matriz de confusión (precisión, recuperación, puntuación F1)
- Precisión de la clasificación:  

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
- Estimaciones de probabilidad para cada clase

## 2.5. Paso 5: Interpretación neutrosófica

Se desarrolló un nuevo marco de interpretación neutrosófica para cuantificar la incertidumbre de la predicción a través de tres componentes:

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

dónde

- $T$ : Grado de pertenencia a la verdad (confianza en la clasificación correcta)
- $I$ : Grado de indeterminación-pertenencia (incertidumbre en la clasificación)
- $F$ : Grado de falsedad-pertenencia (evidencia en contra de la clasificación)

Los valores neutrosóficos se calcularon en función de la probabilidad prevista de la clase real utilizando la siguiente función por partes:

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

Este mapeo transforma los resultados probabilísticos tradicionales en conjuntos neutrosóficos, proporcionando una representación más matizada de la confianza y la incertidumbre de la predicción.

## 2.6. Visualización

Se generaron visualizaciones completas para facilitar la interpretación:

- Resumen del rendimiento: Análisis comparativo de los componentes neutrosóficos para predicciones correctas e incorrectas
- Gráfico de importancia de características: puntuaciones de importancia variable del modelo de bosque aleatorio
- Espectro neutrosófico: diagrama de dispersión que muestra la relación entre la probabilidad máxima de clase y los valores neutrosóficos
- Análisis por clase: diagramas de caja de los valores de los componentes de verdad en diferentes clases reales

Todos los resultados, incluidas las etiquetas reales, las etiquetas previstas, las probabilidades de clase y los componentes neutrosóficos, se exportaron para su posterior análisis y reproducibilidad.

## 3. Resultados



El marco experimental se implementó en la versión 4.1.0 de R, utilizando varios paquetes clave, como tidyverse para la manipulación de datos, randomForest para la construcción de modelos, caret para la partición de datos y la evaluación del rendimiento, y ggplot2 para la visualización. El código de este trabajo está disponible en [7].

El conjunto de datos se dividió mediante una estrategia de división 80-20 de entrenamiento y prueba con muestreo estratificado para mantener las proporciones de distribución de clases en ambos subconjuntos. El valor de semilla aleatoria se estableció en 123 para garantizar la reproducibilidad de los resultados. Se verificó rigurosamente la consistencia a nivel de factor entre los conjuntos de entrenamiento y prueba para evitar posibles problemas de modelado.

### 3.1. Resumen de rendimiento

Con base en los resultados presentados en las Tablas 1, 2 y 3, el modelo Random Forest demostró un rendimiento excepcional en la clasificación de aplicaciones para guarderías. La precisión general alcanzada fue cercana al 95%, lo que indica que el modelo predijo correctamente la categoría de recomendación para la mayoría de los casos en el conjunto de prueba. Este alto nivel de precisión se sustenta aún más con las métricas detalladas a nivel de clase. Para las clases not\_recom y priority, el modelo logró precisión y recuperación perfectas (1.00), lo que significa que no clasificó erróneamente ninguna aplicación perteneciente a estas categorías y nunca les asignó incorrectamente la aplicación de otra clase. La clase spec\_prior también mostró un rendimiento casi perfecto con valores de precisión y recuperación de 0.99 y 0.98, respectivamente. Si bien las clases recommend y very\_recom tuvieron métricas ligeramente inferiores, su rendimiento se mantuvo sólido, con puntuaciones F1 de 0.92 y 0.93, respectivamente. La confusión mínima, como se ve en las diagonales no perpendiculares de la matriz de confusión (Tabla 1), ocurrió principalmente entre estas dos categorías adyacentes, semánticamente similares (recomendar y muy\_recomendado), lo que es un error comprensible y relativamente menor.

**Tabla 1.** Matriz de confusión

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

**Tabla 2.** Estadísticas generales

Overall Statistics	
Accuracy	0.9483
95% CI	(0.939, 0.9565)
No Information Rate	0.3336
P-Value [Acc > NIR]	$< 2.2 \times 10^{-16}$
Kappa	0.9235
McNemar's Test P-Value	NA

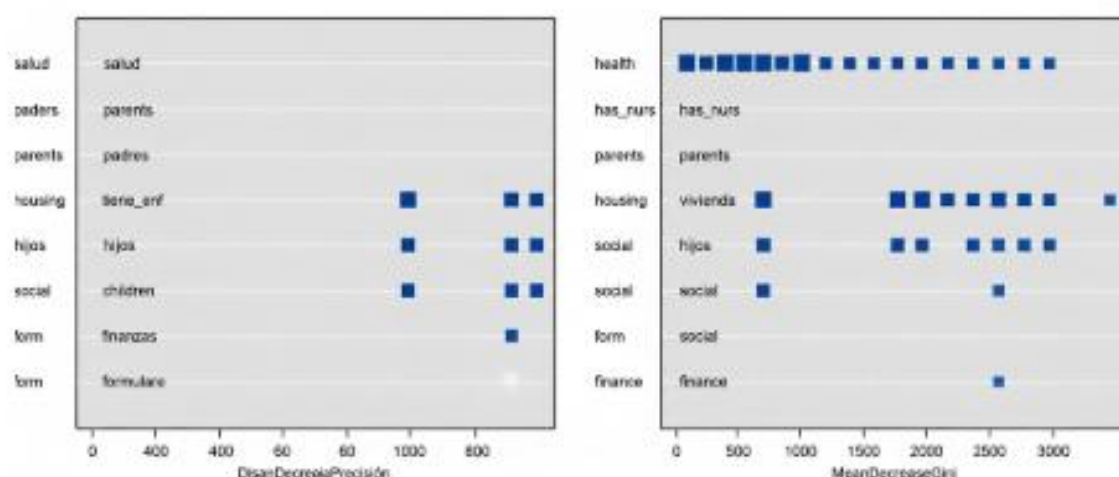
**Tabla 3.** Estadísticas por clase

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

Las implicaciones prácticas de estos resultados son significativas para el proceso de administración de escuelas infantiles. Un modelo cercano al 95% de precisión puede servir como un filtro de primera pasada altamente confiable, automatizando la clasificación inicial de un gran volumen de solicitudes. Esto reduciría drásticamente la carga de trabajo manual para el personal administrativo, permitiéndoles enfocar sus esfuerzos en tareas más complejas, como revisar casos límite o realizar entrevistas. La fortaleza particular del modelo en la identificación perfecta de solicitudes no recomendadas y prioritarias es especialmente valiosa. Puede filtrar eficientemente las solicitudes claramente no calificadas, al tiempo que garantiza que los casos más críticos (prioritarios) nunca se pasen por alto y se tramiten rápidamente para su atención inmediata. Además, la capacidad del modelo para funcionar consistentemente bien en todas las categorías ordenadas garantiza que el proceso automatizado sea justo y completo, proporcionando un estándar de evaluación consistente y objetivo que minimiza el sesgo o error humano en la fase de selección inicial.

### 3.2. Importancia de las características

El análisis de la importancia de los predictores, como se detalla en la Figura 2, proporciona una visión crucial de los factores que el modelo de Bosque Aleatorio considera más influyentes en la clasificación de las solicitudes para escuelas infantiles. Se utilizaron dos métricas estándar para evaluar esta importancia: la Precisión de la Disminución Media y el Índice de Gini de la Disminución Media.



**Figura 2.** Importancia de los predictores de la clase de guardería .

Según la medida de la **Precisión de Disminución Media** (panel izquierdo de la Figura 2), que cuantifica cuánto disminuye la precisión de un modelo al permutar una variable específica, el atributo financiero (situación financiera) es el predictor más importante. Le siguen de cerca el factor social (condiciones sociales) y el factor de salud (condiciones de salud). Esto indica que estos tres



factores (la situación financiera familiar, el entorno social y la salud del niño) son los principales impulsores del rendimiento de clasificación del modelo. Su alta importancia significa que la eliminación o reorganización de estos datos provocaría la mayor disminución en la capacidad del modelo para realizar predicciones correctas.

Por el contrario, la medida **de Gini de Disminución Media** (panel derecho de la Figura 2), que evalúa la reducción total de la impureza de los nodos (es decir, la eficacia con la que una variable divide los datos en grupos de clases puras) atribuible a una variable, presenta una situación ligeramente distinta. Esta métrica destaca la salud como la característica más importante por un margen considerable, seguida de *has\_nurs* (guardería infantil) y *parents* (ocupación de los padres). La importancia de Gini sugiere que estos atributos son los más eficaces para crear separaciones claras y diferenciadas entre las diferentes clases de recomendación dentro de los propios árboles de decisión.

Las implicaciones prácticas de estos resultados son profundas para la administración de guarderías. El predominio de las finanzas, lo social y la salud en la métrica de precisión de la disminución media valida que las decisiones del modelo se ajustan a una política de admisión lógica y centrada en la equidad. Confirma que el proceso prioriza correctamente a los solicitantes en función de sus necesidades socioeconómicas y su bienestar, criterios comunes para asignar recursos limitados de cuidado infantil a quienes más los necesitan. Los administradores pueden utilizar este hallazgo para auditar y justificar sus directrices políticas, garantizando su aplicación transparente.

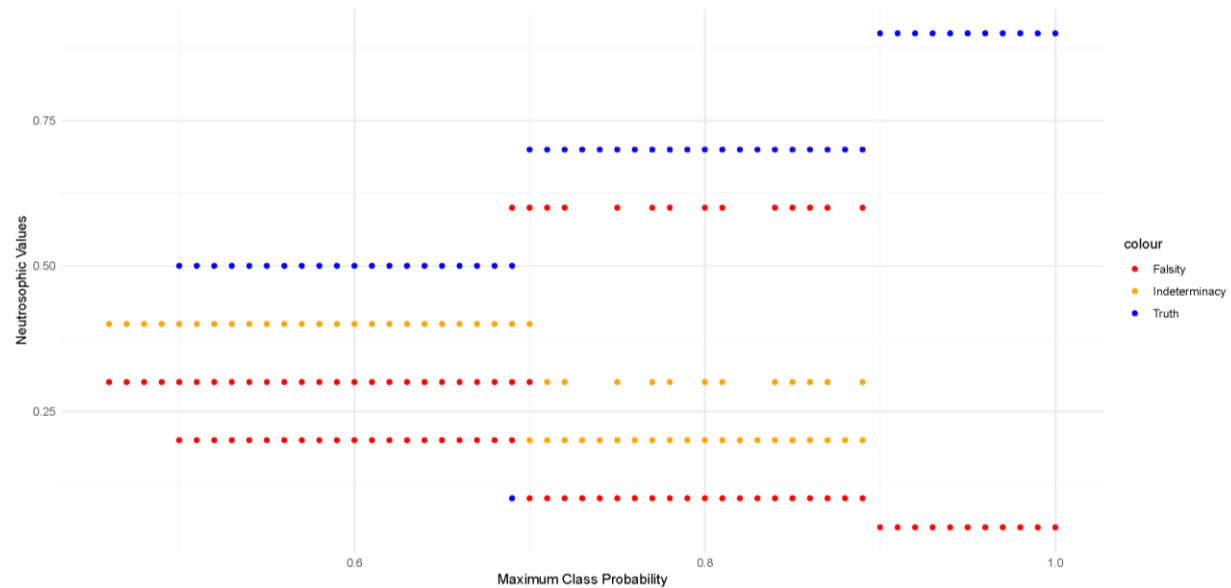
La alta importancia de Gini de la salud y de *has\_nurs* revela la mecánica operativa del proceso de decisión. Estos son los atributos que dividen con mayor precisión a los solicitantes en categorías distintas. Por ejemplo, el estado de salud de un niño podría ser el primer filtro, y el más decisivo, categorizándolo inmediatamente como "prioritario" o "no recomendado". Esta información es invaluable para agilizar el proceso de revisión de solicitudes. Permite a los administradores estructurar sus revisiones manuales priorizando los criterios más decisivos, lo que mejora significativamente la eficiencia.

Además, la disparidad entre las dos métricas de importancia es en sí misma reveladora. El hecho de que las finanzas sean cruciales para la precisión general, pero menos para la creación de divisiones puras (índice de Gini), implica que su valor es más matizado e interactúa con otras variables para afinar las predicciones. Esto subraya la complejidad del proceso de toma de decisiones y la solidez del modelo de Bosque Aleatorio para capturar estas intrincadas relaciones. En definitiva, este *análisis de importancia de las características* transforma el modelo de alta precisión de una "caja negra" en una herramienta transparente para la planificación estratégica, la validación de políticas y la asignación eficiente y equitativa de recursos.

### 3.3. Espectro neutrosófico

Con base en los resultados presentados en la Figura 3, la confianza de predicción del modelo se captura e interpreta eficazmente a través de la lente de la lógica neutrosófica. El gráfico ilustra la relación entre la probabilidad máxima de clase predicha y los componentes neutrosóficos correspondientes: verdad (T), indeterminación (I) y falsedad (F). Esta visualización revela un espectro claro de confianza: las predicciones con una probabilidad máxima alta (p. ej., superior a 0,8) corresponden a una alta pertenencia a la verdad (T) y una baja indeterminación (I) y falsedad (F), lo que indica clasificaciones sólidas y fiables. Por el contrario, las predicciones con probabilidades moderadas (p. ej., entre 0,4 y 0,7) presentan una indeterminación elevada, lo que refleja ambigüedad o incertidumbre en la decisión del modelo. Estos suelen ser casos límite donde el modelo es menos seguro, y a menudo ocurren entre clases semánticamente similares, como "recommend" y "very\_recom". El componente de falsedad (F) se mantiene generalmente bajo, pero aumenta ligeramente a medida que disminuye la confianza, lo que indica una creciente evidencia en contra de la clasificación prevista. Este mapeo matizado permite a los administradores no solo ver lo que predijo el modelo, sino también su nivel de confianza, lo que proporciona una mayor

interpretabilidad que va más allá de las métricas de precisión convencionales.



**Figura 3.** Espectro de confianza de predicción asignado a valores neutrosóficos (  $T$ ,  $I$ ,  $F$  ).

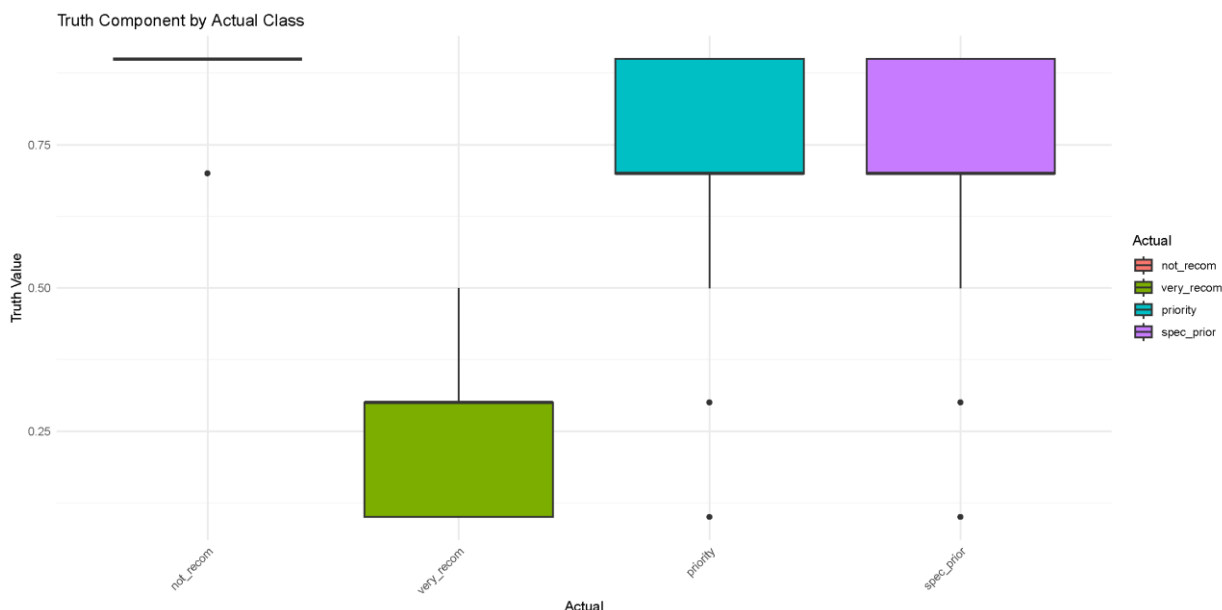
Las implicaciones prácticas de estos resultados son significativas para la administración y la toma de decisiones en escuelas infantiles. Al cuantificar la incertidumbre mediante componentes neutrosóficos, el modelo facilita la adopción de estrategias de toma de decisiones escalonadas. Las predicciones de alta confianza ( $T$  alta,  $I$  baja y  $F$ ) pueden automatizarse por completo, lo que reduce la revisión manual en casos claros y agiliza el proceso de tramitación de solicitudes. Las predicciones con alta indeterminación ( $I$ ) pueden marcarse para su revisión humana, lo que garantiza que el personal administrativo preste una atención minuciosa a los casos límite o ambiguos. Este enfoque equilibra la eficiencia con la imparcialidad, evitando la toma de decisiones totalmente automatizadas en escenarios inciertos y reduciendo el riesgo de clasificación errónea. Además, la capacidad de evaluar la incertidumbre visual y numéricamente ayuda a las instituciones a justificar y explicar las decisiones a las partes interesadas, lo que mejora la transparencia y la confianza en el sistema automatizado. Al incorporar la lógica neutrosófica, el modelo va más allá de los resultados binarios correctos/incorrectos y ofrece un marco estructurado y basado en la confianza que se adapta bien a los procesos de toma de decisiones del mundo real, donde la incertidumbre es inherente y debe gestionarse explícitamente.

### 3.4. Análisis por clases

El análisis clase por clase del componente de verdad ( $T$ ) en las categorías de recomendación reales, como se muestra en la Figura 4, ofrece una comprensión matizada de la confianza del modelo en sus predicciones para cada clase. Los valores de pertenencia a la verdad, derivados de la interpretación neutrosófica de las probabilidades de predicción, reflejan el grado de confianza que tiene el modelo al clasificar correctamente las instancias de una clase dada. Clases como **not\_recom** y **priority** presentan valores de verdad consistentemente altos, con medianas cercanas a 1,0 y una dispersión mínima, lo que indica que el modelo no solo es preciso, sino también altamente confiable al predecir estas categorías. Esto coincide con las métricas de rendimiento anteriores, donde estas clases lograron una precisión y una recuperación perfectas. Por el contrario, las clases "**recom**" y "**very\_recom**" muestran rangos intercuartiles más amplios y, ocasionalmente, valores medianos de  $T$  más bajos, lo que sugiere una mayor variabilidad en la confianza de la predicción. Esto es coherente con la confusión observada entre estas dos clases semánticamente adyacentes, donde el modelo encuentra casos más ambiguos, lo que reduce la certeza incluso cuando las predicciones son correctas. La clase **spec\_prior** demuestra una fuerte confianza media, aunque con algunos valores atípicos que indican instancias ocasionales de menor certeza, posiblemente debido a características



superpuestas con otras clases o muestras subrepresentadas en los datos de entrenamiento.



**Figura 4.** Componente de verdad por clase real.

Desde un punto de vista práctico, estos hallazgos tienen implicaciones significativas para el proceso de revisión de solicitudes en las escuelas infantiles. Los niveles de confianza altos y estables para las clases **not\_recom** y **priority** significan que las decisiones automatizadas para estas categorías se pueden implementar con alta confiabilidad, reduciendo la necesidad de intervención manual y acelerando el procesamiento de casos claros. Esto permite al personal administrativo asignar más tiempo y recursos a las solicitudes que caen en categorías con mayor incertidumbre, como **recommend** y **very\_recom**. La variabilidad en la confianza para estas clases intermedias subraya la importancia de la supervisión humana para los casos límite. Al marcar las solicitudes donde el valor de pertenencia a la verdad cae por debajo de cierto umbral, el sistema puede garantizar que las predicciones inciertas reciban una revisión adicional, mejorando así tanto la imparcialidad como la precisión de la decisión final. Este enfoque híbrido, que combina decisiones automatizadas de alta confianza con validación humana en el circuito para casos inciertos, optimiza tanto la eficiencia como la equidad en el proceso de admisión. Además, la capacidad de cuantificar y visualizar la confianza específica de cada clase permite a las instituciones monitorear y refinar continuamente sus modelos predictivos, garantizando que permanezcan alineados con los objetivos de las políticas y respondan a los perfiles cambiantes de los solicitantes.

#### 4. Conclusión

Basado en un análisis exhaustivo, este estudio demuestra con éxito la eficacia de integrar un clasificador Random Forest con un marco lógico neutrosófico para la clasificación consciente de la incertidumbre de las aplicaciones de la guardería. El modelo logró una alta precisión general de aproximadamente el 95%, mostrando su fuerte capacidad predictiva. Más importante aún, la aplicación de conjuntos neutrosóficos, definidos por el triple  $(T, I, F)$ , proporcionó un mecanismo matizado y transparente para cuantificar la confianza y la incertidumbre de la predicción. Esto permitió una distinción clara entre las clasificaciones de alta confianza, que se pueden automatizar con alta fiabilidad, y las predicciones indeterminadas de baja confianza que justifican la revisión humana. El análisis por clase reveló además que los niveles de confianza fueron más altos para las clases 'not\_recom' y 'priority', alineándose con su perfecta precisión y recuperación, mientras que las predicciones para las clases semánticamente adyacentes 'recommend' y 'very\_recom' exhibieron mayor variabilidad e incertidumbre.

Las implicaciones prácticas de estos hallazgos son significativas para optimizar el proceso de admisión a escuelas infantiles. El marco propuesto facilita un proceso de toma de decisiones híbrido y escalonado. Las predicciones de alta confianza pueden automatizarse para procesar eficientemente solicitudes claras, reduciendo drásticamente la carga administrativa. Por el contrario, las solicitudes marcadas con alta indeterminación (I) pueden priorizarse para la evaluación manual, garantizando que los casos límite reciban el escrutinio minucioso necesario para mantener la imparcialidad y la equidad. Este enfoque optimiza la asignación de recursos humanos, combinando la eficiencia de la automatización con el juicio crítico de expertos humanos donde más se necesita. Además, el análisis de importancia de las características ofrece información valiosa para la validación de políticas, confirmando que los factores financieros, sociales y de salud son los principales impulsores del modelo de clasificación, lo cual se alinea con una política de admisión centrada en la equidad.

En conclusión, el marco de Bosque Aleatorio Neutrosófico va más allá de una simple clasificación binaria para proporcionar una herramienta sofisticada y confiable para el apoyo a la toma de decisiones. Aborda eficazmente la incertidumbre inherente a las tareas de clasificación del mundo real, mejorando tanto la transparencia como la fiabilidad del sistema automatizado. Al cuantificar y visualizar la incertidumbre, esta metodología no solo mejora la eficiencia del proceso de revisión de solicitudes, sino que también garantiza un resultado más sólido, justo y explicable, sentando las bases para una toma de decisiones responsable y ética basada en datos en contextos administrativos.

## 5. Referencias

- [1] Bohanec, M., & Rajkovic, V. (1990). Expert system for decision making. *Sistemica*, 1(1), 145-157.
- [2] Kaggle, "Nursery dataset." <https://www.kaggle.com/datasets/heitornunes/nursery> , 2025. Accessed: 2025-02-20.
- [3] Parrales-Bravo, F., Gómez-Rodríguez, V., Barzola-Monteses, J., Caicedo-Quiroz, R., Tolozano-Benites, E., & Vasquez-Cevallos, L. (2024). From descriptive to prescriptive analytics on time series: Studying the number of preeclampsia inpatient beds. *IEEE Access*.
- [4] Parrales-Bravo, F., Caicedo-Quiroz, R., Tolozano-Benites, E., Vasquez-Cevallos, L., & Cevallos-Torres, L. (2025). STOP: Studying Time-series Of Preeclamptic Emergencies. *IEEE Access*.
- [5] Parrales-Bravo, F., Caicedo-Quiroz, R., Tolozano-Benitez, E., Gómez-Rodríguez, V., Cevallos-Torres, L., Charco-Aguirre, J., & Vasquez-Cevallos, L. (2024). OUCH: oversampling and undersampling cannot help improve accuracy in our bayesian classifiers that predict preeclampsia. *Mathematics*, 12(21), 3351.
- [6] Parrales-Bravo, F., Cevallos-Torres, L., Vasquez-Cevallos, L., Caicedo-Quiroz, R., Tolozano-Benites, R., & Gómez-Rodríguez, V. (2025). A Review of the Use of Data Analytics to Address Preeclampsia in Ecuador Between 2020 and 2024. *Diagnostics*, 15(8), 978.
- [7] Parrales-Bravo, F. "R code to integrate random forests and neutrosophic analysis in the Nursery dataset." <https://github.com/fparrale/RandomForestNeutrosophyNursery> , 2025. Accessed: 2025-02-20.

