

Simulador docente para la comprensión de la construcción de árboles de decisión ID3

Ismael Ramos-Pérez, Gadea Lucas-Pérez, Daniel Drefs-Fernandes, Pedro Latorre-Carmona, Carlos López-Nozal
ismaelrp@ubu.es, glucas@ubu.es, ddx1015@alu.ubu.es, plcarmona@ubu.es, clopezno@ubu.es

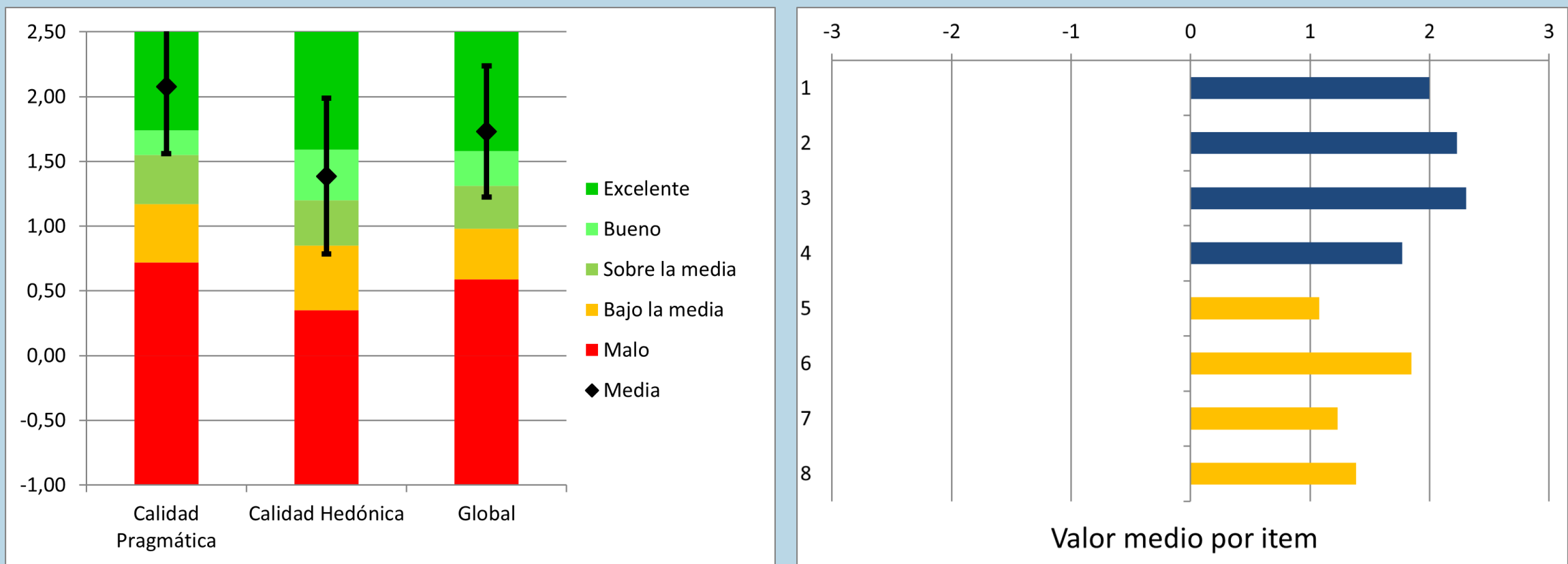


Motivación y objetivos

Los árboles de decisión son modelos populares en aprendizaje automático por su claridad e interpretabilidad [1]. No obstante, su construcción paso a paso puede resultar compleja para principiantes. Los simuladores didácticos han demostrado ser herramientas eficaces como recursos de apoyo en el proceso de enseñanza y aprendizaje. Presentamos una herramienta docente interactiva y de código abierto que facilita la comprensión del algoritmo ID3. Esta necesidad surge ante la escasez de simuladores centrados en IA explicativa. La herramienta permite visualizar, paso a paso y de forma interactiva, el cálculo de la entropía, entropía condicional y la ganancia de información [2], ayudando a entender el proceso de construcción del árbol.

Evaluación de la herramienta

La herramienta fue evaluada con el **User Experience Questionnaire Short (UEQ-S)** por 13 docentes con experiencia en Minería de Datos. Los resultados muestran una **calidad pragmática excelente** (media de 2.08), buena calidad hedónica (media de 1.38) y una valoración global excelente (media de 1.73), según el *benchmark* oficial del test.



Los ítems mejor valorados fueron facilidad de uso, eficiencia y claridad. Las valoraciones hedónicas fueron más variables, lo que sugiere margen de mejora en el diseño visual y la originalidad.

Conclusiones y líneas futuras

La herramienta presentada facilita la comprensión del algoritmo ID3 al permitir visualizar paso a paso la construcción del árbol y los conceptos de entropía y ganancia de información. Su diseño interactivo y su disponibilidad como recurso abierto la convierten en un apoyo útil para la docencia. La evaluación con docentes mediante el test UEQ-S mostró resultados muy positivos, destacando especialmente la claridad y facilidad de uso.

Líneas futuras:

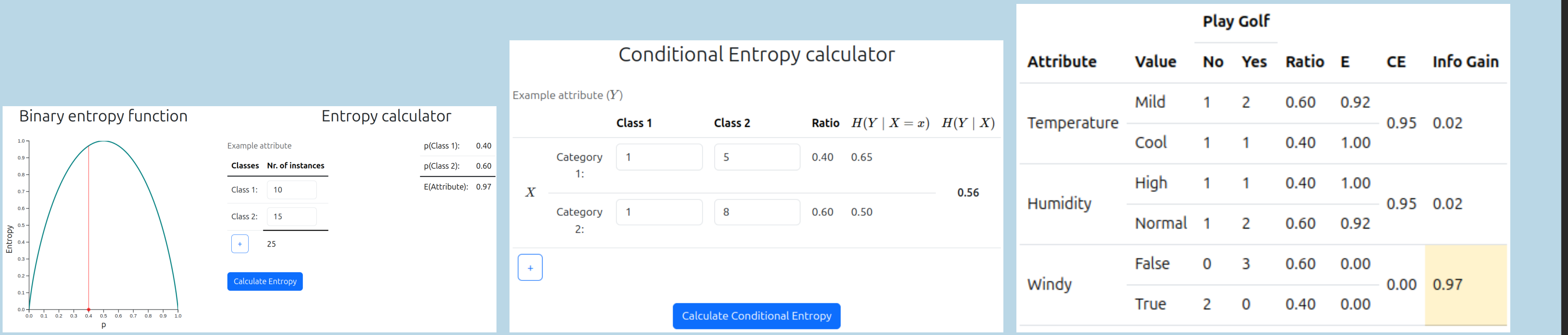
- Evaluar su impacto en el aprendizaje real de estudiantes.
- Añadir soporte para atributos numéricos y clases múltiples (C4.5).
- Incluir algoritmos alternativos como CART y técnicas de poda.
- Mejorar la interactividad: explorar nodos, crear datasets manualmente, mejorar el diseño visual.

Referencias

- [1] V. G. Costa, C. E. Pedreira, *Recent advances in decision trees: An updated survey*, Artificial Intelligence Review, vol. 56, no. 5, pp. 4765–4800 (2023).
- [2] T. M. Cover, J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

Fundamentos teóricos

El algoritmo ID3 construye árboles de decisión dividiendo datos según el atributo que maximiza la **ganancia de información**. Aplica el concepto de **entropía**, una medida de incertidumbre, buscando su minimización, en cada partición. **Entropía**: incertidumbre de una variable. **Entropía condicional**: incertidumbre restante al conocer otra variable. **Ganancia de información**: reducción de entropía al dividir por un atributo.



Funcionalidad del simulador

- La herramienta está dividida en tres pestañas que guían al usuario a través de los conceptos fundamentales del algoritmo ID3:
- 1. Entropía**: incluye una calculadora interactiva para experimentar con distribuciones de clases y visualizar la entropía calculada.
 - 2. Entropía condicional**: permite introducir valores categóricos y muestra paso a paso el cálculo completo, incluyendo la media ponderada.
 - 3. Árbol ID3**: simula la construcción del árbol mostrando:
 - El árbol generado paso a paso.
 - El conjunto de datos con filtros aplicados.
 - Los cálculos de ganancia de información por atributo.

Los usuarios pueden usar conjuntos de datos preseleccionados, o cargar los suyos en formato `.csv` (con datos categóricos y clase binaria).

La herramienta es de código abierto (<https://github.com/admirable-ubu/decision-trees-simulator>) y está disponible en línea (<https://admirable-ubu.github.io/decision-trees-simulator/>).

