

Examen-IA2-2022-Ordinaria.pdf



wJaKu_



Inteligencia Artificial II



3º Grado en Ingeniería Informática



Facultad de Informática
Universidad Complutense de Madrid

WUOLAH + BBVA

Hazte **cliente de BBVA y...**
ahórrate 6 meses
de suscripción

BOOM

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

NETFLIX

Spotify

HBOmax

Disney+

PlayStation Plus

DAZN

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

Abre tu cuenta



Hazte cliente de BBVA y ... ahórrate 6 meses de suscripción

WUOLAH
+ BBVA

NETFLIX

Spotify

HBOmax

Disney+

PlayStation.Plus

DAZN

Ahora, si te abres una Cuenta Online en BBVA, te reembolsamos una de estas suscripciones durante 6 meses (hasta 9,99€/mes) al pagarla con tu tarjeta Aqua Débito

Promoción solo para nuevos clientes de BBVA. Válida hasta el 30/06/2023. Estas empresas no colaboran en la promoción.

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

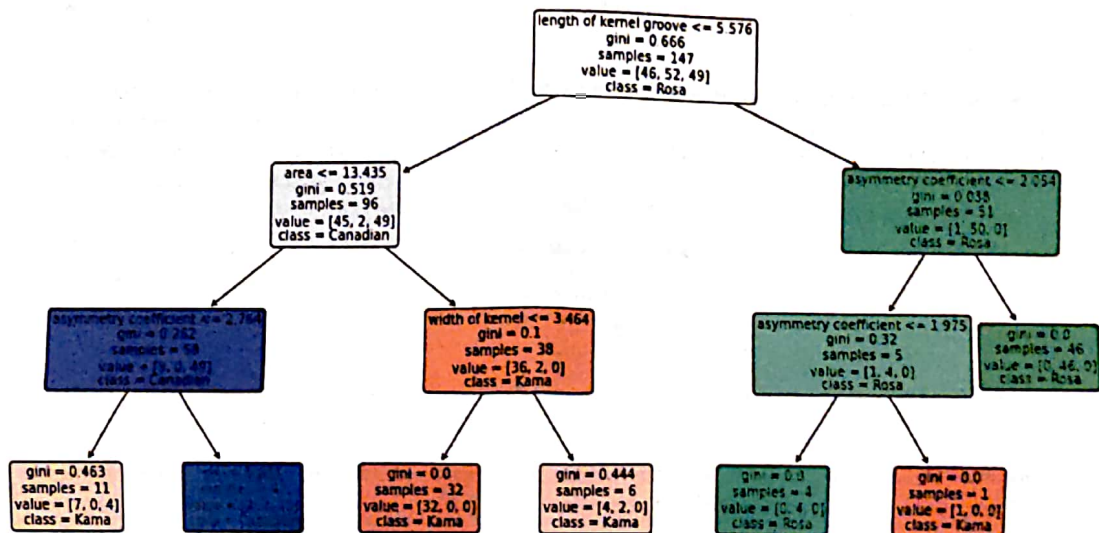
BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Examen de Inteligencia Artificial II

Convocatoria ordinaria - 20 de mayo de 2022

Duración: 2 horas

Pregunta 1 [2 puntos]. Se ha construido un árbol de decisión utilizando el criterio de Gini para un problema en el que se quiere clasificar unas semillas de trigo en tres variedades (*Rosa*, *Kama* y *Canadian*). Como variables de entrada se usa la longitud y el ancho de la hendidura de la semilla (*length and width of the kernel groove*), largo y ancho de la semilla (*length and width of kernel groove*), el área de la semilla (*area*) y una medida de asimetría de la semilla (*assymetry coefficient*).



Contesta de forma razonada y concisa a las siguientes cuestiones:

- Determina la precisión del árbol (en %) para las tres clases de semillas usando como referencia el conjunto de datos usado para crearlo.
- Considerando el árbol resultante, indica razonadamente el punto en el que verías más necesario realizar una poda.
- ¿Consideras que tendría sentido añadir un nivel adicional al árbol? Para razonar tu respuesta indica qué beneficio concreto esperarías obtener y/o qué problemas podrían aparecer.
- Si usamos este árbol tal cual como clasificador, ¿qué dos especies se confunden más entre sí (en los dos sentidos)? ¿Qué dos se confunden menos?

Pregunta 2 [1 punto]. Explica brevemente en qué consiste el diagrama del codo y para qué se usa en aprendizaje automático. Dibuja un diagrama del codo prototípico, explica qué suele representarse en sus ejes y qué conclusiones puedes extraer de él.



Abre tu cuenta



WUOLAH
+ BBVA

Pregunta 3 [1 punto]. Disponemos del siguiente corpus de frases:

- *la caja roja está cerrada*
- *juan abrió la puerta roja*
- *juan abrió una puerta verde*
- *arturo abrió la caja roja*
- *juan cerró una puerta verde*
- *la caja azul se rompió*
- *una puerta está rota*
- *abrió juan*

Considerando el modelo de bigramas, responde a las siguientes cuestiones:

- Escribe la frase que se obtendría eligiendo en cada momento la palabra más probable.
- Calcula la probabilidad de encontrar la frase "una puerta verde" indicando los cálculos necesarios para obtenerla.

Pregunta 4 [1.5 puntos]. Supongamos un problema de clasificación de sentimiento en textos cortos (por ejemplo, tuits) en el que vamos a usar k-NN. En nuestro conjunto de datos tenemos un número pequeño de tuits de vocabulario muy variado. Nos planteamos caracterizar los textos mediante el vector medio de los vectores de una *word-embedding* GloVe de 50 dimensiones que se correspondan con las palabras del texto, ponderando los vectores por la frecuencia de aparición de las palabras en el texto.

Contesta de manera razonada y concisa a las siguientes preguntas:

- ¿Consideras que la aproximación basada en *word-embeddings* podría funcionar bien?
- ¿En qué aspectos consideras superior esa aproximación a usar k-NN con la aproximación de bolsa de palabras y su frecuencia de aparición en los textos?
- Indica un aspecto de los textos que pueda representar un problema para las dos aproximaciones propuestas.

Pregunta 5 [1.5 puntos]. Considera una base de datos sobre móviles en forma de hechos Prolog de la forma `dato(Modelo, Propiedad, Valor)`, por ejemplo:

```
dato(iphone13, cpu, a15).  
dato(iphone13, ram, 4).  
dato(pixel5, cpu, g765).  
dato(pixel5, memoria, 128).  
dato(pixel5, peso, 151).  
...
```

Crea una DCG que use como símbolo inicial pregunta (ver más abajo) que sea capaz de contestar preguntas en lenguaje natural del tipo:

- cuál es la cpu del iphone13
- dime el peso del pixel5
- cuánta memoria tiene el pixel5
- qué ram tiene el iphone13

No es necesario que la DCG compruebe los fenómenos de concordancia de género o número entre las palabras. Te facilitamos la siguiente regla Prolog que debe imprimir el valor sobre el que se pregunta en la sentencia X.

```
procesa(X) :- pregunta(M, P, X, []), dato(M, P, V), write(V).  
% Comienzo de la primera regla de la DGC  
pregunta(M, P) --> % Continúa la regla
```


Pregunta 6 [1.5 puntos]. Indica, de forma razonada y concisa, si cada una de las siguientes afirmaciones es verdadera o falsa.

- a. Siempre es recomendable usar el formalismo más expresivo para representar el conocimiento de un dominio.
- b. Las redes semánticas, los sistemas de marcos y las ontologías OWL-DL permiten representar conocimiento prototípico de forma sencilla (conocimiento general en los conceptos abstractos y excepciones en conceptos más específicos o en las instancias).
- c. La diferencia entre razonamiento en mundo abierto y mundo cerrado está en cómo interpretamos el conocimiento que no aparece en la base de conocimiento.

Pregunta 7 [1.5 puntos]. Considera una base de conocimiento RDF con las siguientes relaciones (para cada relación te indicamos su dominio y rango usando la notación Relación: Dominio -> Rango):

```
instanceOf: Instance -> Class
subclassOf: Class -> Class
hasNationality: Human -> Country
hasYearOfBirth: Human -> int
hasTitle: LiteraryWork -> str
hasGenre: LiteraryWork -> Genre
hasAuthor: LiteraryWork -> Human
hasPublicationYear: LiteraryWork -> int
```

Además, en esta base de conocimiento Novel es un tipo de obra literaria, ScienceFiction es un género de literatura y USA es un país concreto, es decir, una instancia de Country.

Escribe una consulta SPARQL que recupere las novelas de ciencia ficción (o cualquiera de sus subgéneros) escritas por autores estadounidenses nacidos entre 1950 y 1980. Para cada novela se debe mostrar su título y, si está disponible, el año de publicación. Los resultados se mostrarán ordenados alfabéticamente por título.