

2021examordinaria.pdf



bobesponja



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.

Examen de Inteligencia Artificial II Convocatoria ordinaria - 14 de junio de 2021

Duración: 2 horas

1/6

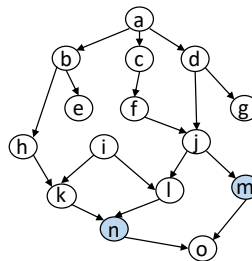
Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

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.

Pregunta 1 [1 punto]. Escribe una consulta SPARQL que recupere los títulos y las fechas de estreno de todas las películas dirigidas por Spielberg de menos de 90 minutos de duración. Las películas deben aparecer ordenadas alfabéticamente por título. La fecha de estreno puede no estar disponible en todas las películas. La información se recupera a partir de una base de conocimiento RDF con las siguientes características (los identificadores de las entidades y relaciones aparecen en cursiva):

- Entidades: *Spielberg*, *Film*.
- Relaciones:
 - instanceOf*: Instance → Class
 - subclassOf*: Class → Class
 - title*: Film → string
 - duration*: Film → int
 - directedBy*: Film → Person
 - releaseDate*: Film → date
- Las películas se representan mediante instancias directas o indirectas de la clase *Film*.

Pregunta 2 [1 punto]. En una de las prácticas del curso usamos una medida de similitud semántica basada en el cálculo del *Least Common Subsumer* (LCS). Define qué es el LCS y qué representa, escribe la fórmula de la medida de similitud y úsala para calcular la similitud entre los conceptos *n* y *m* en la siguiente taxonomía donde las flechas conectan cada concepto con sus subconceptos.



Pregunta 3 [2 puntos]. Representa en Prolog, siguiendo los convenios explicados en clase, una base de conocimiento a partir de la siguiente información. Escribe un comentario por cada predicado que utilices explicando su significado.

- El Metro está compuesto por distintas líneas de conectan estaciones. Por ejemplo, la línea amarilla en dirección sur conecta las estaciones de Moncloa y Argüelles, Argüelles y Ventura Rodríguez, etc. Las estaciones de Moncloa y Argüelles también están conectadas por la línea gris.
- Siempre que se va a un sitio en Metro es posible regresar haciendo el camino inverso.
- Las líneas de este Metro no tienen ciclos (por ejemplo, no existen líneas circulares). Tampoco tienen sentido los trayectos que contengan ciclos.
- Dos estaciones de la misma línea están a distancia N si para llegar de una a otra debo pasar por N-1 estaciones intermedias.
- Dos estaciones están próximas si es posible viajar de una a otra pasando por 3 estaciones intermedias o menos, y cambiando como máximo una vez de línea.



WUOLAH
+ BBVA

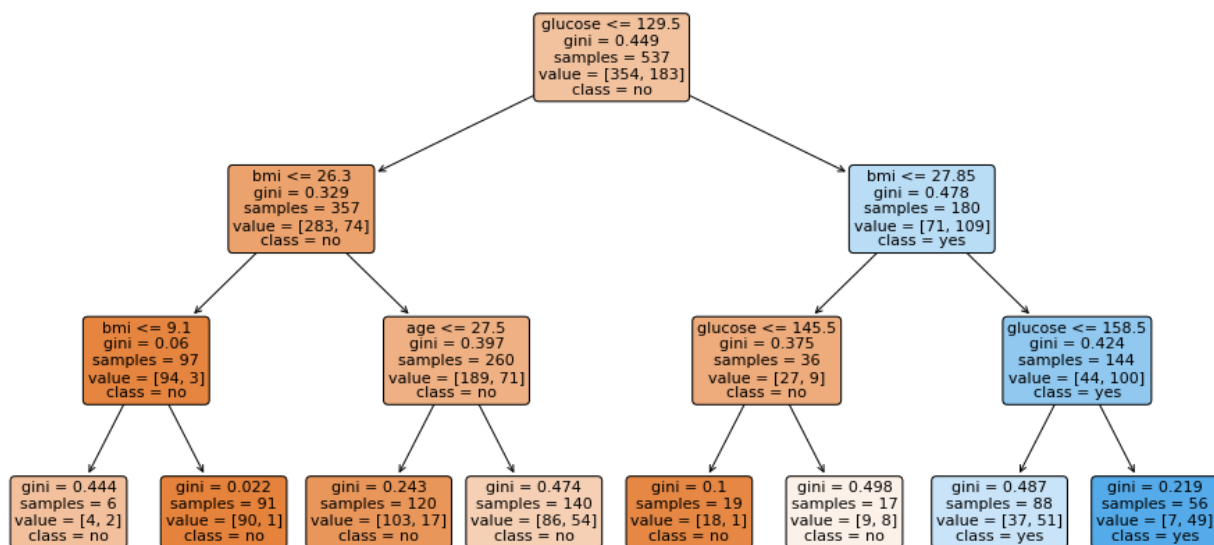
WUOLAH

- Las estaciones pueden o no tener distintos servicios. Por ejemplo, Moncloa tiene tiendas, Bibliometro, cajeros automáticos, etc. Pero Ventura Rodríguez no tiene ninguno de esos servicios.
- Las estaciones exclusivamente comerciales son las que tienen tiendas y cajeros automáticos, y no tienen Bibliometro.

Escribe a continuación las siguientes consultas:

- Estaciones próximas donde sólo una de ellas tenga Bibliometro.
- Estaciones que no sean exclusivamente comerciales.

Pregunta 4 [1.5 puntos]. Se ha construido un árbol de decisión para tratar de determinar si los pacientes sufren diabetes a partir de distintas variables descriptoras: el índice de masa corporal (*bmi*), la edad (*age*), los niveles de insulina (*insulin*), los niveles de glucosa (*glucose*), la presión sanguínea (*bp*), etc.



Contesta razonadamente a las siguientes cuestiones:

- Indica si las clases a predecir están equilibradas en el conjunto de datos de entrenamiento y cómo debemos tenerlo en cuenta al interpretar el error del modelo.
- Indica cuáles son las variables más relevantes en este problema y cómo podemos saberlo.
- Calcula la entropía total del segundo nivel del árbol (siendo la raíz el primer nivel).
- Indica los valores del parámetro r para los que se podría el nodo "bmi <= 9.1" usando poda pesimista.
- Calcula la tasa de aciertos, y la precisión y exhaustividad por clase de un árbol que sólo hiciera una pregunta sobre el mismo conjunto de datos con el que se ha construido.

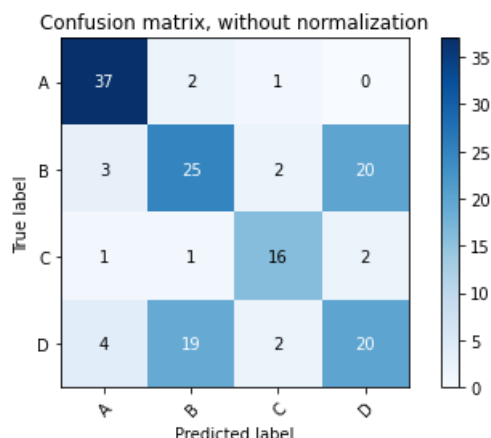
Pregunta 5 [1.5 puntos]. Explica brevemente la veracidad o falsedad de las siguientes afirmaciones.

- Los árboles de decisión no se ven penalizados si el conjunto de datos contiene variables de entrada (x_i) redundantes o irrelevantes para el problema a resolver.
- En aprendizaje supervisado, una baja correlación entre las variables en entrada (x_i) indica que el problema será difícil de modelar de forma precisa.
- El tiempo de cómputo de un clasificador k -NN es proporcional al valor de k .
- Un modelo con una exactitud (*accuracy*) de 99% en un problema de clasificación biclase tendrá una precisión (*precision*) y exhaustividad (*recall*) elevadas en ambas clases.

- e) El método de validación leave one out es recomendable en problemas donde se disponga de pocos datos.
- f) El algoritmo de descenso de gradiente, con una tasa de aprendizaje suficientemente pequeña, encontrará el mínimo global cuando la función a minimizar es convexa.

Pregunta 6 [1 punto]. Explica brevemente 3 estrategias distintas que se pueden utilizar para evitar el sobreaprendizaje al entrenar modelos basados en redes neuronales (las estrategias no tienen por qué ser exclusivas para redes neuronales, pero deben poder usarse con ellas).

Pregunta 7 [1 punto]. Dada la siguiente matriz de confusión, responde razonadamente a las siguientes cuestiones:



- a) ¿Cuántos ejemplos de cada clase contiene el conjunto de entrenamiento? ¿Cuál es la exactitud (*accuracy*) del modelo?
- b) ¿Cuál es la precisión (*precision*) y exhaustividad (*recall*) de la clase C?
- c) ¿Qué clases se distinguen mejor? ¿Qué clases se confunden más?
- d) Indica si las matrices de confusión se usan tanto en problemas de clasificación como de regresión.

Pregunta 8 [1 punto]. Se han analizado 200 críticas positivas y 100 negativas de restaurantes y se ha obtenido la siguiente tabla de frecuencias de aparición de términos.

Términos	Positivas	Negativas
reserva	75	30
servicio	100	80
limpio	50	35
precio	50	65
ruido	5	60
repetir	150	15

- a) Usa un clasificador Naive Bayes para determinar la probabilidad de que un mensaje con los términos *reserva*, *servicio*, *precio* y *repetir* corresponda a una crítica positiva o negativa.
- b) ¿Es Naive Bayes un modelo interpretable o de caja negra? Razona tu respuesta.