

|  |  |
| --- | --- |
| Colegio Universitario **IES** *Siglo 21* | |
| EXAMEN PARCIAL 1 | |
| **Materia: Introducción a la IA** | **Docente:** Ricardo Piña |
| **Modalidad: Presencial Semipresencial Distancia** | **Fecha:** |

Reservado para el alumno

|  |  |
| --- | --- |
| **Alumno**: Sibello Gino | **Carrera:  INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS** |
| **DNI**: 41.820.055 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Actividad** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **Total puntos** | NOTA |
| **Puntaje** |  |  |  |  |  |  |  |  |  |  |  |  |

* Objetivos:
  + Comprender los elementos fundamentales de la forma de analizar problemas de Aprendizaje.
  + Conocer las distintas clasificaciones de problemas de IA.
  + Ser capaz de aplicar a problemas cercanos a la realidad algún modelo de Machine Learning, seleccionar los hiperparámetros y evaluar el modelo.
* Modalidad de Evaluación:
  + Escrito e individual.
* Puntaje:
  + - El puntaje se determinará a través de la siguiente escala:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Puntaje | 1-24 | 25-39 | 40-54 | 55-61 | 62-66 | 67-72 | 73-79 | 80-87 | 88-95 | 96-100 |
| Nota | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Valoración | Logro  No satisfactorio | | | Logro básico | Logro satisfactorio | | Logro  Destacado | Logro  Sobresaliente | | Logro  Excelente |

PARA REMITIR EL EXAMEN TEÓRICO:

Guarde el examen teórico con *suApellido\_suNombre.docx* y remítalo por el servicio de Mensajería del Aula Virtual, antes de las 23 hs.

1. En los siguientes casos indique:

* **a qué tipo de Aprendizaje** (Supervisado, No Supervisado, etc) corresponde y
* **qué tipo de problema es** (Clasificación, Regresión, etc)(1 punto) :
  1. Se desea pronosticar el peso que tendrán las peras una vez maduras y se conocen los datos: alto, largo, cantidad\_días\_cosechada, peso\_madura para 1000 peras.

Tipo de aprendizaje: NO Supervisado

Tipo de problema: Regresión

* 1. En un sistema de vehículo autónomo (es decir que se conduce por sí solo) es fundamental que el sistema pueda detectar si un ser humano se encuentra en la calle. **Se cuenta con miles de fotografías de personas y con miles de fotografías de elementos que suelen aparecer en la calle y que no son personas**. Se desea que nuestro sistema tome imágenes con una cámara y decida si lo que se ve en la imagen incluye o no a una persona.

Tipo de aprendizaje: Sistema Supervisado

Tipo de problema: Clasificacion

* 1. La gerencia de marketing de un hipermercado desea elaborar una campaña de marketing: desea enviar mensajes por mail o WhatsApp ofreciendo sólo los productos que los clientes podrían estar interesados en comprar. Para ello cuenta con la base de datos de las compras anteriores e información demográfica de cada uno de los clientes. La idea es que clientes *similares* comprarán productos *similares*. Nuestro objetivo es indicarle al gerente de mkt qué clientes se parecen más entre sí en función de los datos mencionados.

Tipo de aprendizaje: NO supervisado

Tipo de problema: Clasificación

1. SITUACIÓN: (2 puntos)

Sea un problema de **Aprendizaje Supervisado** para resolver el cual, Ud posee un Conjunto de Datos C (que es sólo una fracción de todos los casos posibles o universo) y lo divide en dos (sub)conjuntos de datos A y B, de tal manera que C = A U B. Ud decide aplicar un modelo (como por ejemplo árboles de decisión).

En cada uno de los siguientes casos **comente** cada una de las alternativas y analice **Exactitud (Accuracy),** **Overfiting** que se obtendría en cada caso. Cuál de las estrategias utilizaría Ud?

* 1. Caso 1: Entrenar al modelo con C y testearlo sobre C:

La Exactitud sería muy alta en este caso, pero estamos ante un problema de Overfiting ya que el programa habría aprendido demasiado bien sobre el train y no sería muy bueno para casos generales. Aunque si buscaramos que sea un producto final sin testearlo es lo que deberíamos hacer, pero no sin previamente haberle hecho una división de los datos y habiendo comprobado el nivel de profundidad perfecta para este caso

* 1. Caso 2: Entrenar al modelo con A y testearlo sobre C:

En este caso estaríamos cometiendo otro error ya que estamos testando a A con los mismos datos con lo que lo entrenamos, por ende:

El accuracy aunque menor al caso 1 su resultado se vería influenciado por el overfiting generado al testear con parte del train

* 1. Caso 3: Entrenar al modelo con A y testearlo sobre B:

Esto es lo que se debería hacer:

En este caso el accuracy seria menor al de los otros casos ya que en ningún momento sufrimos de overfiting porque los datos con los que testeamos y entrenamos al programa son totalmente diferentes y dan una perspectiva mejor de como actuaria el programa ante una situación general

* 1. Yo usaría la siguiente estrategia:

Dividiría A y B de tal forma que A tenga el 70% de los datos de C y B el 30%

Luego subdividirá A de nuevo y haría lo mismo haciendo nuevas variables Aval y Bval en la que Aval tendría el 70% de los datos de A y Bval el 30%

Con Aval y Bval testaría cual es la mejor profundidad de árbol en la que el accuracy sea lo mayor posible testeando el programa con Bval, Una vez hecho esto Haría una última prueba en la que ahora entrenaría a mi programa con A y lo testearía con B para ver cual es el accuracy que puedo alcanzar con mi profundidad de árbol ya obtenida anteriormente y ahí dejaría para el record cuál es el accuracy de mi programa y finalmente agarraría todos los datos ósea C y lo entrenaría una ultima vez a mi programa sin hacerle ningún test, este ultimo seria el programa que mandaria a produccion

1. Dado el siguiente conjunto de datos donde **y** es la variable a pronosticar, **x1, x2, x3** son las características, features, o variables explicativas, aplique el método de **Árbol de Decisión con Ganancia de Información** para calcular manualmente el primer nodo mediante el método de la entropía. (**Nota**: si le es de utilidad, estos datos se encuentran en el archivo: Parcial\_1\_Ejs.xlsx) (2 puntos)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | x1 | x2 | x3 | y |
| 1 | No | Bajo | Positivo | **A** |
| 2 | Si | Alto | Negativo | **B** |
| 3 | Si | Alto | Positivo | **A** |
| 4 | Si | Medio | Negativo | **A** |
| 5 | Si | Alto | Negativo | **A** |
| 6 | No | Bajo | Negativo | **A** |
| 7 | No | Bajo | Negativo | **A** |
| 8 | No | Medio | Negativo | **B** |
| 9 | No | Bajo | Positivo | **B** |
| 10 | No | Bajo | Positivo | **A** |
| 11 | Si | Alto | Positivo | **A** |
| 12 | Si | Alto | Positivo | **B** |
| 13 | Si | Alto | Negativo | **B** |
| 14 | Si | Medio | Positivo | **A** |
| 15 | No | Medio | Negativo | **A** |

Escriba las cuentas aquí mismo hechas en Word, no importa si la simbología matemática no es precisa, por ejemplo si necesita escribir “logaritmo en base 2 de 40”, por ejemplo, puede escribirlo así: log2(40) que lo entenderemos!

Mediante el método De ganancia de información y eliminando el hecho de que el árbol tiene que ser binario nos convien utilizar X2 como primer nodo ya que al poder dividirlo en 3 nos da una mayor información.

Tenemos 10 casos de A y 5 B por ende 10/15 y 5/15

Por ende la formula para calcular la entropía de este primer Nodo seria

S = - [ { (10/15)x(log2(10/15) } + { (5/15)x(log2(5/15) } ]

S = -[ (-0,389975) + (-0,528320) ]

S = 0,918

Osea tenemos una entropía del 91,8%