

Control 5 - IIC2433

Jueves 5 de Noviembre de 2020
Árboles de decisión, Random Forest, Naïve Bayes

Indicaciones

- El control es **individual**. La copia será castigada con nota 1.1 al curso, además de las sanciones disciplinarias correspondientes.
- El control tiene cuatro preguntas, para obtener el 7 deben contestarlas todas de forma correcta. Cada una tiene su propia ponderación.
- Para cada pregunta, si esta contiene más de una subpregunta, **debe contestar cada una de las subpreguntas para obtener el puntaje total en la pregunta.**
- La entrega de este control se realizará a través de un buzón en Canvas, que permanecerá abierto hasta las 23:59 del día Jueves 5 de Noviembre. **No se permitirán entregas atrasadas.**
- El control debe ser entregado en formato **PDF**. En caso de entregar con otro formato, este no se corregirá.

Preguntas Árboles de decisión (2.5 pts)

1. (2.0 pts.) Construya el árbol de decisión para los siguientes datos, donde tenemos 4 atributos (Profesional, Número de hijos, Con trabajo y Joven) y una clase (Aprobado). Escriba los cálculos paso a paso.

ID	Profesional	Número de Hijos	Con trabajo	Joven	Aprobado (Clase)
1	Sí	2	Sí	Sí	Sí
2	Sí	2	Sí	Sí	Sí
3	Sí	2	Sí	Sí	Sí
4	No	2	No	No	No
5	No	2	No	No	No
6	No	2	No	No	No
7	No	2	Sí	Sí	No
8	No	2	Sí	Sí	No

2. (0.5 pts.) Utilizando el árbol construido anteriormente, prediga el valor de los siguientes datos y diga el porcentaje de certeza que tiene sobre la decisión.

Profesional	Número de Hijos	Con trabajo	Joven	Aprobado (Clase)
Sí	4	No	Sí	? =
No	1	Sí	No	? =

Preguntas Random Forest (1.5 pts)

1. (1.5 pts.) ¿Cómo podría utilizar el algoritmo Random Forest como un regresor? Es decir, predecir el valor numérico y no una clase. Explique el proceso paso a paso. Por último, como podría reducir la cantidad de dimensiones de forma arbitraria de un *dataset* utilizando este algoritmo. Explique el procedimiento para partir de una dataset de N dimensiones a uno de M, con M menor que N.

Preguntas Naïve Bayes (2.0 pts)

- Dado el siguiente dataset, responda:

ID	Maní	Vino	Mariscos	¿Alergia?
1	No	No	Sí	Sí
2	No	Sí	No	Sí
3	No	No	Sí	Sí
4	Sí	Sí	Sí	Sí
5	Sí	No	Sí	No
6	No	Sí	No	No
7	Sí	No	No	No
8	No	No	No	No

1. (0.7 pts) ¿Cuál es la probabilidad de no haber consumido Maní, dado que hubo alergia, es decir $P(\text{Maní}=\text{No}|\text{Alergia}=\text{Si})$? Escriba el desarrollo que justifique la respuesta.
2. (0.7 pts) ¿Cuál es la probabilidad de no haber consumido Vino, dado que no hubo alergia, es decir $P(\text{Vino}=\text{No}|\text{Alergia}=\text{No})$? Escriba el desarrollo que justifique la respuesta.
3. (0.6 pts) Basado en los datos previos y usando un clasificador Naïve Bayes, clasifique los siguientes dos casos. Escriba el desarrollo que justifique la respuesta.

Maní	Vino	Mariscos	¿Alergia?
Sí	Sí	No	??
No	Sí	Sí	??