

TP1: Método de Bayes

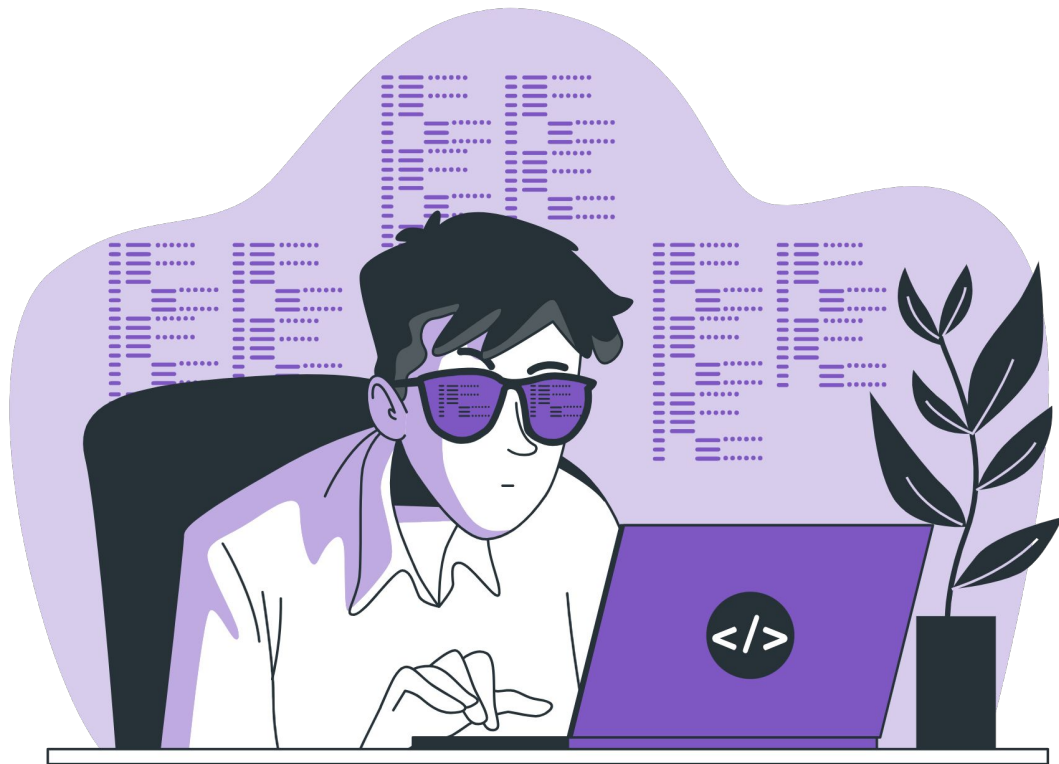
Grupo 3:

Mila Langone

Paula Oseroff

Luciana Diaz Kralj

Paula A. Domingues



Contenidos a presentar

01

Naive Bayes

Funcionamiento e implementación.

02

Aplicación de métricas

Matriz de confusión, medidas de evaluación y curvas ROC.

03

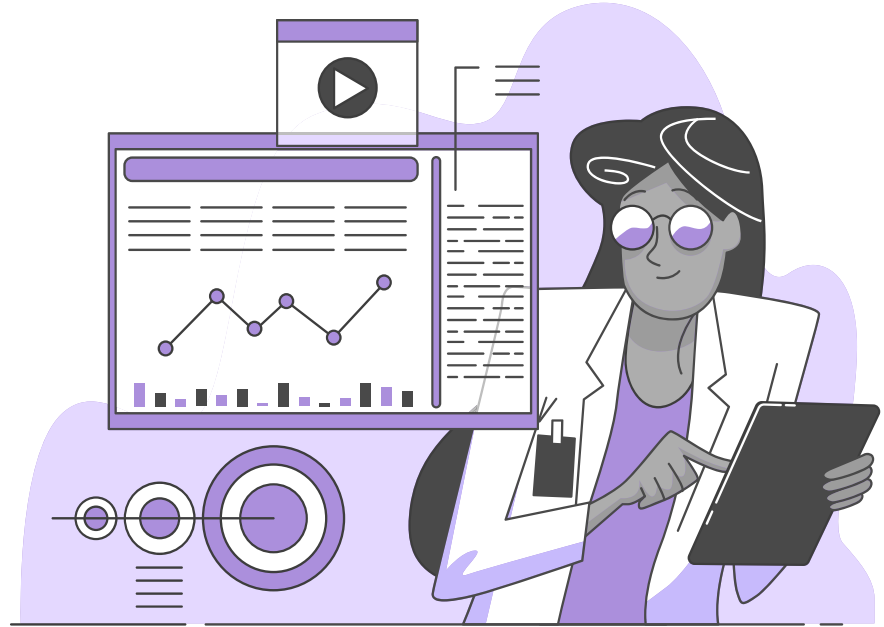
Redes Bayesianas

Funcionamiento e implementación.

04

Conclusiones

Análisis de consecuencias y conocimientos adquiridos.



01

Naive Bayes

Funcionamiento e
implementación



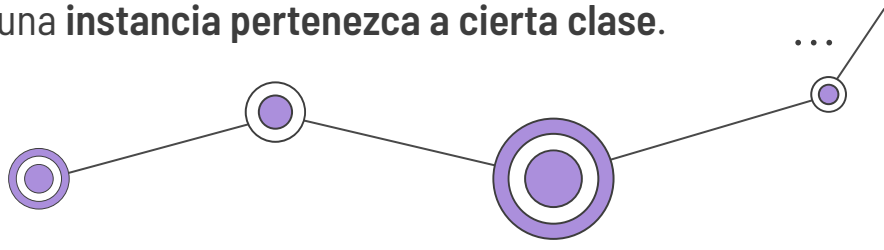


Teorema de Bayes



Repasamos

$$P(h|D) = \frac{P(D|h) \cdot P(h)}{P(D)}$$

- **h** es la **hipótesis** o la clase
 - **D** es el **conjunto de valores** para los atributos
 - **P(h)**: probabilidad **a priori** de cada una de las clases
 - **P(D)**: probabilidad a priori de que los atributos tengan ciertos valores
 - **P(D|h)**: probabilidad de que los atributos tengan ciertos valores dado que la instancia pertenece a una clase
 - **P(h|D)** es la probabilidad **a posteriori** de que una **instancia pertenezca a cierta clase**.
- 



Clasificador **Ingenuo** de Bayes



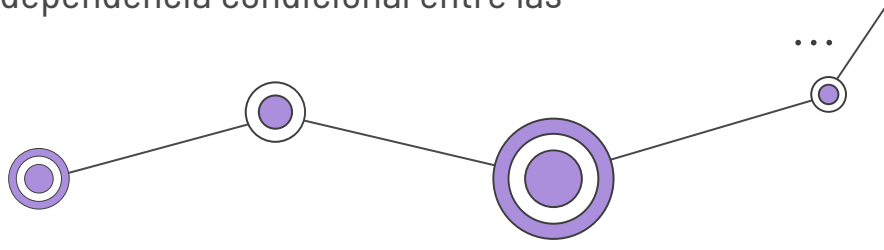
Objetivo Principal: Determinar la clase más probable para una nueva instancia en base a la información de las características del conjunto de entrenamiento.

Requisitos:

- Conjunto de datos etiquetado → aprendizaje supervisado
- Independencia condicional de los datos dada una clase
- Conocimiento a priori de las probabilidades asociadas
- Conjunto de Entrenamiento Representativo (diverso y equilibrado)

¿Por qué es Ingenuo?

Por la suposición simplificada que hace sobre la independencia condicional entre las características dada la clase a la que pertenecen.





Clasificador **Ingenuo** de Bayes



Funcionamiento

- Dado un ejemplo representado por un conjunto de atributos $\{a_1, a_2, \dots, a_n\}$
- Debemos asignar este ejemplo a una clase V dentro de un conjunto finito de clases $V = \{v_1, v_2, \dots, v_k\}$.
- Aplicamos el Teorema de Bayes para calcular la probabilidad de cada clase dado los atributos observados:

$$P(v_j | (a_1, \dots, a_n))$$

- Por último, de cada una de estas probabilidades a posteriori, seleccionamos la que implique la **máxima probabilidad**

$$V_{\text{opt}} = \max_{\forall j} (P(v_j | (a_1, \dots, a_n)))$$

Utilizaremos el teorema de Bayes para despejar esta última ecuación.





Clasificador **Ingenuo** de Bayes



Surge un problema cuando una probabilidad $P(v_j | a_i)$ se calcula como 0 a partir de los datos de entrenamiento.

- Si $P(v_j | a_i) = 0$, toda la probabilidad de la clase a_i se anula por el atributo observado v_j .

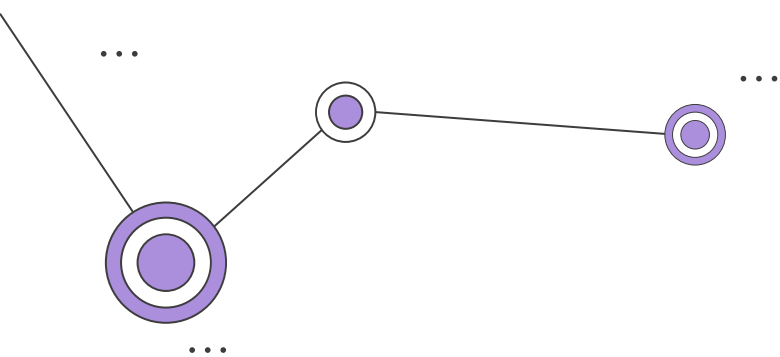
Suavización Laplaciana

- Consiste en agregar un valor (k) a las frecuencias para evitar probabilidades nulas.
- De esta forma una posible probabilidad nula en la productoria no anula la probabilidad final.

De esta forma, la frecuencia relativa se calcula en base a la fórmula:

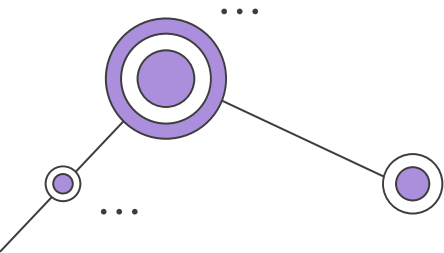
$$p = \frac{\text{cantidad de ocurrencias} + 1}{\text{total} + k}$$

donde k es el número de valores posibles que toma a_i .



Ejercicio 1:

Preferencias británicas



Ejercicio a Resolver

| scones | cerveza | whiskey | avena | futbol | Nacionalidad |
|--------|---------|---------|-------|--------|--------------|
| 0 | 0 | 1 | 1 | 1 | I |
| 1 | 0 | 1 | 1 | 0 | I |
| 1 | 1 | 0 | 0 | 1 | I |
| 1 | 1 | 0 | 0 | 0 | I |
| 0 | 1 | 0 | 0 | 1 | I |
| 0 | 0 | 0 | 1 | 0 | I |
| 1 | 0 | 0 | 1 | 1 | E |
| 1 | 1 | 0 | 0 | 1 | E |
| 1 | 1 | 1 | 1 | 0 | E |
| 1 | 1 | 0 | 1 | 0 | E |
| 1 | 1 | 0 | 1 | 1 | E |
| 1 | 0 | 1 | 1 | 0 | E |
| 1 | 0 | 1 | 0 | 0 | E |

En función de la información presentada, se deben clasificar dos ejemplos:

- $x_1 = (1, 0, 1, 1, 0)$
- $x_2 = (0, 1, 1, 0, 1)$

Utilizando el clasificador ingenuo de Bayes



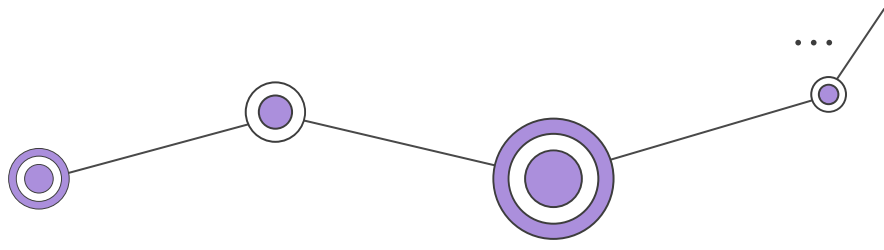
Ejercicio a Resolver



En este caso

- V es la nacionalidad, escocés (E) o irlandés (I)
- Los atributos: { scones, cerveza, whiskey, avena, futbol}

Sean $x_1 = (1, 0, 1, 1, 0)$ y $x_2 = (0, 1, 1, 0, 1)$ las instancias a clasificar:

- $P(I | x_1) = 0.236$
 - $P(E | x_1) = 0.764$
 - $P(I | x_2) = 0.832$
 - $P(E | x_2) = 0.168$
- } Es más probable que la persona con los gustos x_1 le corresponda la nacionalidad **escocesa**.
- } La persona con los gustos x_2 se clasificará como una persona **irlandesa**.
- 

¿Qué pasa si evaluamos la matriz original utilizando Naive Bayes?

| scones | cerveza | whiskey | avena | futbol | Nacionalidad | P-Nacionalidad |
|--------|---------|---------|-------|--------|--------------|----------------|
| 0 | 0 | 1 | 1 | 1 | I | I |
| 1 | 0 | 1 | 1 | 0 | I | E |
| 1 | 1 | 0 | 0 | 1 | I | E |
| 1 | 1 | 0 | 0 | 0 | I | E |
| 0 | 1 | 0 | 0 | 1 | I | I |
| 0 | 0 | 0 | 1 | 0 | I | I |
| 1 | 0 | 0 | 1 | 1 | E | E |
| 1 | 1 | 0 | 0 | 1 | E | E |
| 1 | 1 | 1 | 1 | 0 | E | E |
| 1 | 1 | 0 | 1 | 0 | E | E |
| 1 | 1 | 0 | 1 | 1 | E | E |
| 1 | 0 | 1 | 1 | 0 | E | E |
| 1 | 0 | 1 | 0 | 0 | E | E |


- No recomendable, usamos el mismo conjunto para **train** y **test**.
- Podemos ver que mismo usando instancias conocidas no predice con 100% de **accuracy** (77%).



02

Aplicación de Métricas

Matriz de confusión, medidas de
evaluación y curvas ROC





Métricas de evaluación



Buscamos **evaluar** y **comparar** distintos **métodos de clasificación**, estudiando las consecuencias ante el cambio de parámetros.

Rendimiento del modelo: evaluación de la clasificación

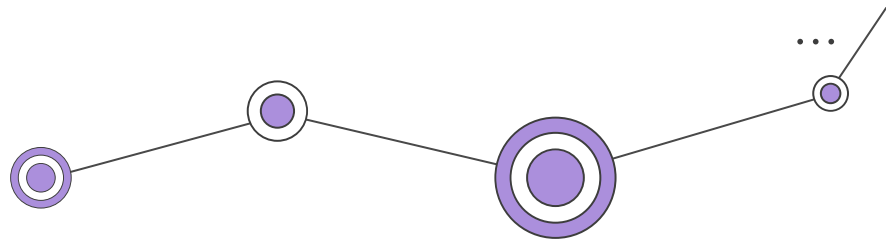
- $$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- $$\text{Tasa de FP} = \frac{FP}{FP + TN}$$

- $$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

- $$\text{Precision} = \frac{TP}{TP + FP}$$

- $$\text{Recall} = \frac{TP}{TP + FN}$$





Métricas de evaluación

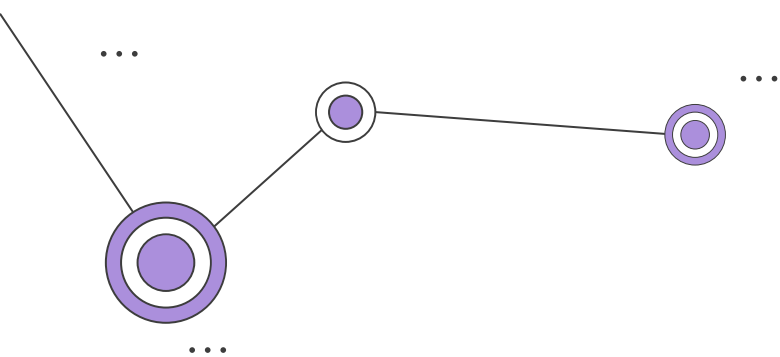


Rendimiento del modelo: evaluación de desempeño

- **Matriz de confusión:** Determina si el sistema confunde dos clases contrastando predicciones con los datos reales.
- **Curva Receiver Operating Characteristic (ROC):** Visualiza la tasa de verdaderos positivos contra la tasa de falsos positivos al variar el umbral de decisión.
- **Área debajo de la curva ROC (AUC):** Medida de la precisión del test de clasificación.

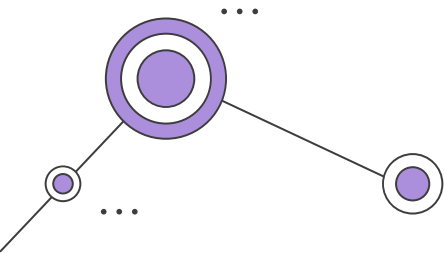
Diseño del experimento: validación del modelo

- **Train test split:** Simulación del comportamiento del modelo frente a datos inéditos.
- **Validación cruzada en k-partes:** División aleatoria de los datos en k partes de igual tamaño, usando 1 parte como test y las k-1 restantes como entrenamiento.

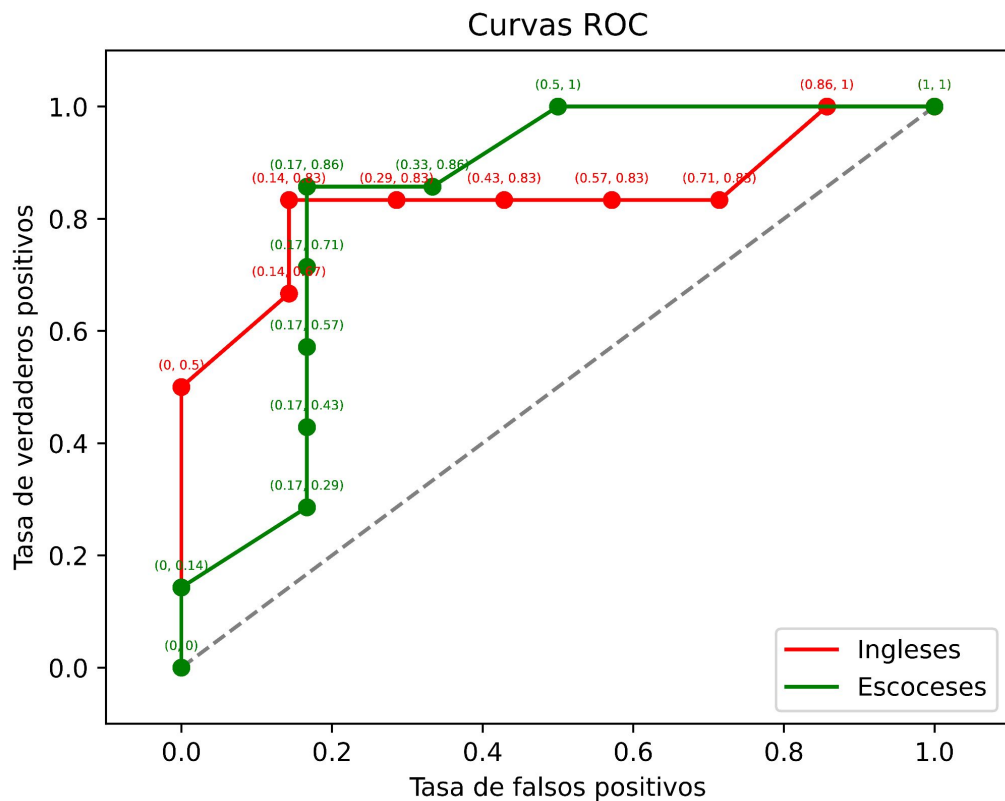


Ejercicio 1:

Preferencias británicas



Curva ROC



Valor AUC
0.833

A partir de un rango de
umbrales:

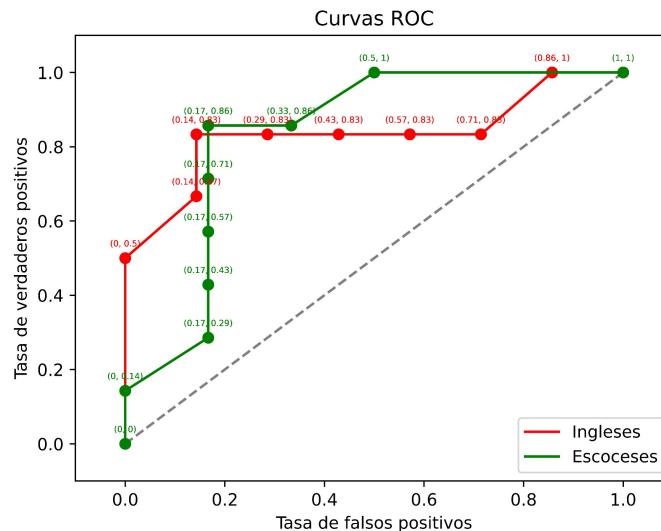
$U \in [0, 1]$
con paso 0.0001

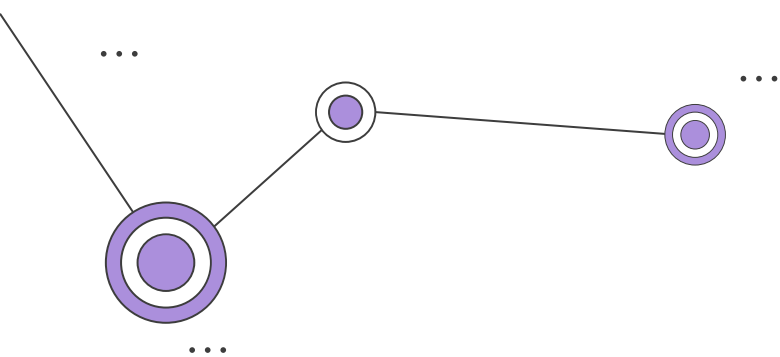
Curva ROC

{'threshold': {'min': 0.0, 'max': 0.1515},
'TPR': 1.0,
'FPR': 1.0,
'metrics': {'TP': 6.0,
'FP': 7.0,
'FN': 0.0,
'TN': 0.0,
'Precision': 0.46,
'Recall': 1.0,
'F1 Score': 0.63,
'Accuracy': 0.46}}

{'threshold': {'min': 0.1516, 'max': 0.1923},
'TPR': 1.0,
'FPR': 0.8571428571428571,
'metrics': {'TP': 6.0,
'FP': 6.0,
'FN': 0.0,
'TN': 1.0,
'Precision': 0.5,
'Recall': 1.0,
'F1 Score': 0.67,
'Accuracy': 0.54}}

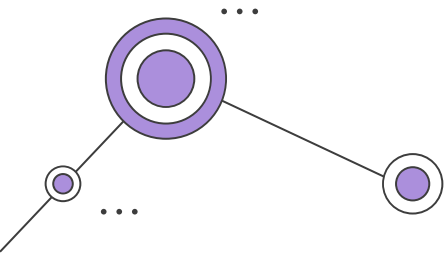
{'threshold': {'min': 0.19,
'max': 0.21}, 'TPR': 0.83, 'FPR': 0.71,
'metrics': {'TP': 5.0,
'FP': 5.0,
'FN': 1.0,
'TN': 2.0,
'Precision': 0.5,
'Recall': 0.83,
'F1 Score': 0.625,
'Accuracy': 0.54}}





Ejercicio 2:

Noticias argentinas



Ejercicio a Resolver

| fecha | titular | fuelle | categoria |
|------------------|--|------------------------|------------|
| 14/11/2018 9:08 | Trabajadores del Buenos Aires Design cortan... | Infobae.com | Nacional |
| 13/11/2018 14:14 | La boda del gobernador Gerardo Morales: tap... | Clarín.com | Nacional |
| 14/11/2018 10:08 | Cumbre del G20: qué calles estarán cortadas y... | iprofesional.com | Nacional |
| 14/11/2018 2:02 | Una fractura que confirma la candidatura de... | LA NACION (Argentina.) | Nacional |
| 14/11/2018 9:03 | Infierno grande: ola de divorcios en un pueblo... | Diario El Día | Nacional |
| 14/11/2018 10:18 | Presupuesto 2019: vallaron el Congreso, en... | LA NACION (Argentina.) | Nacional |
| 14/11/2018 7:47 | Qué calles estarán cortadas y cómo funciona... | Infobae.com | Destacadas |
| 14/11/2018 10:26 | Murió de vacaciones en Italia, la aerolínea per... | Clarín.com | Destacadas |
| 13/11/2018 23:23 | Mauricio Macri se probó productos para la cal... | LA NACION (Argentina.) | Destacadas |
| 14/11/2018 7:11 | La palabra clave que salvó a una niña en Esta... | Infobae.com | Destacadas |
| ... | ... | ... | ... |

En función de la información presentada, implementar un **clasificador de texto** utilizando naive Bayes



Dataset original



Posibles categorías

- Ciencia y Tecnología: 3.856
- Deportes: 3.855
- Destacadas: 3.859
- Economía: 3.850
- Entretenimiento: 3.850
- Internacional: 3.850
- Nacional: 3.860
- Noticias destacadas: 133.819
- Salud: 3.840

← contiene otras categorías

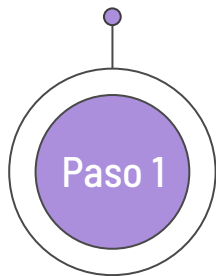
← contiene otras categorías

← excede y contiene otras categorías



Preprocesamiento de titulares

Eliminar números
y símbolos



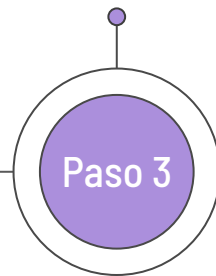
Paso 2

Tokenizar



Quitar
stopwords

Paso 3



Lematizar

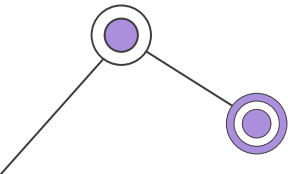
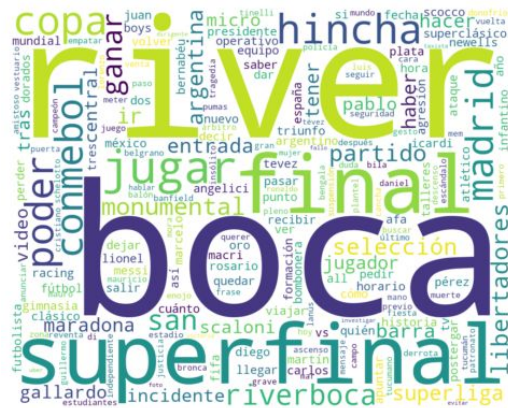
Paso 4



A diagram illustrating a network structure. A central node, represented by a large purple circle with a white border, is connected to several peripheral nodes. Each peripheral node is a smaller purple circle with a white border. Lines connect the central node to the peripheral nodes. Ellipses (...) indicate that there are more nodes and connections in the network.

Deportes

Destacadas



A diagram illustrating a network structure. A central node, represented by a large purple circle with concentric white and purple rings, is connected by lines to several peripheral nodes. These peripheral nodes are smaller purple circles with concentric white and purple rings. The connections form a star-like pattern, with the central node acting as a hub. Ellipses (...) are used to indicate that there are more nodes and connections than shown.

[illegible]

Entretenimiento





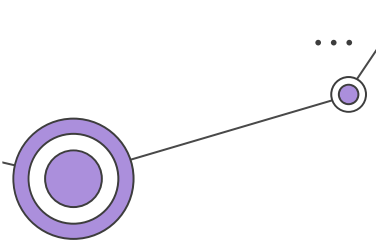
Evaluación de la clasificación



Nacional:

- TP: 1319
- FP: 741
- FN: 589
- TN: 12762
- Precisión: 0.64
- Recall: 0.69
- F1 Score: 0.66
- Accuracy: 0.91

Destacadas:

- TP: 567
 - FP: 668
 - FN: 1372
 - TN: 12804
 - Precisión: 0.46
 - Recall: 0.29
 - F1 Score: 0.36
 - Accuracy: 0.87
- 



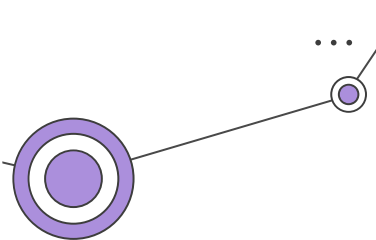
Evaluación de la clasificación



Deportes:

- TP: 1786
- FP: 199
- FN: 120
- TN: 13306
- Precisión: 0.90
- Recall: 0.94
- F1 Score: 0.92
- Accuracy: 0.98

Salud:

- TP: 1858
 - FP: 69
 - FN: 56
 - TN: 13428
 - Precisión: 0.96
 - Recall: 0.97
 - F1 Score: 0.97
 - Accuracy: 0.99
- 



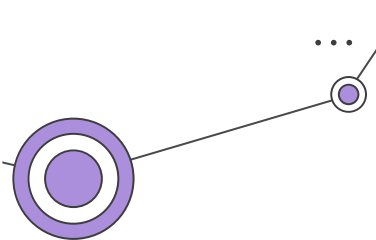
Evaluación de la clasificación



Entretenimiento:

- TP: 1865
- FP: 177
- FN: 103
- TN: 13266
- Precisión: 0.91
- Recall: 0.95
- F1 Score: 0.93
- Accuracy: 0.98

Ciencia y Tecnología:

- TP: 1934
 - FP: 88
 - FN: 41
 - TN: 13348
 - Precisión: 0.96
 - Recall: 0.98
 - F1 Score: 0.97
 - Accuracy: 0.99
- 



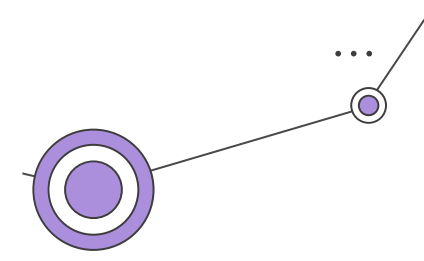
Evaluación de la clasificación



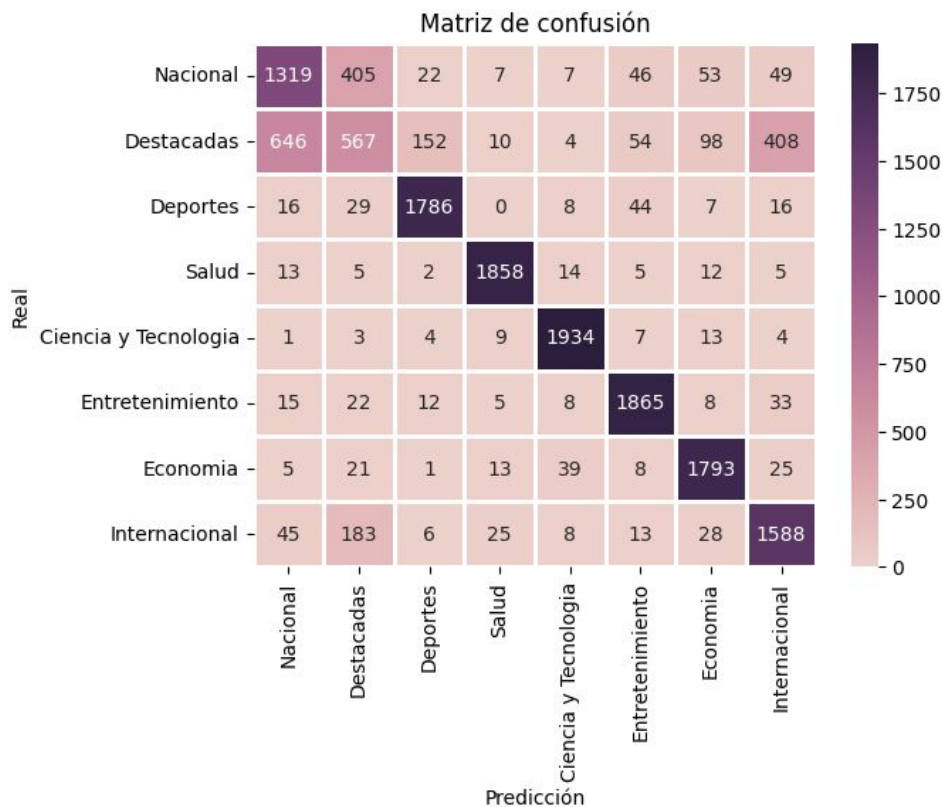
Economía:

- TP: 1793
- FP: 219
- FN: 112
- TN: 13287
- Precisión: 0.89
- Recall: 0.94
- F1 Score: 0.91
- Accuracy: 0.98

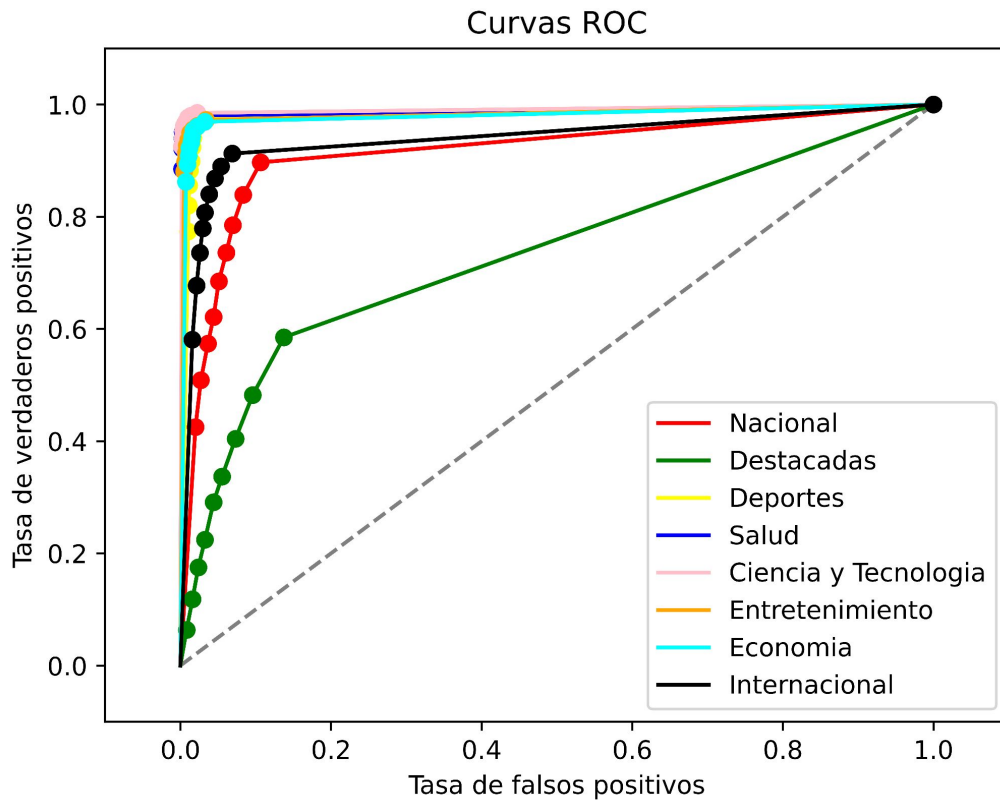
Internacional:

- TP: 1588
 - FP: 540
 - FN: 308
 - TN: 12975
 - Precisión: 0.75
 - Recall: 0.84
 - F1 Score: 0.79
 - Accuracy: 0.94
- 

Matriz de confusión



Curva ROC



Valores AUC

| | |
|-----------------------|-------|
| Nacional: | 0.910 |
| Destacadas: | 0.732 |
| Deportes: | 0.977 |
| Salud: | 0.988 |
| Ciencia y Tecnologia: | 0.990 |
| Entretenimiento: | 0.980 |
| Economia: | 0.976 |
| Internacional: | 0.934 |

A partir de un rango de umbrales:

$U \in [0,1]$
con paso 0.1

Validación cruzada en k-partes

Split #1

Error training: 3088

Error test: 1033

Split #2

Error training: 3064

Error test: 998

Split #3

Error training: 3088

Error test: 979

Split #4

Error training: 3037

Error test: 1027

Split #5

Error training: 3047

Error test: 1065

K= 5

30871 entradas

→ Test: 6175

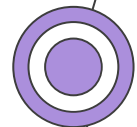
→ Training: 24696



03

Redes Bayesianas

Funcionamiento e implementación



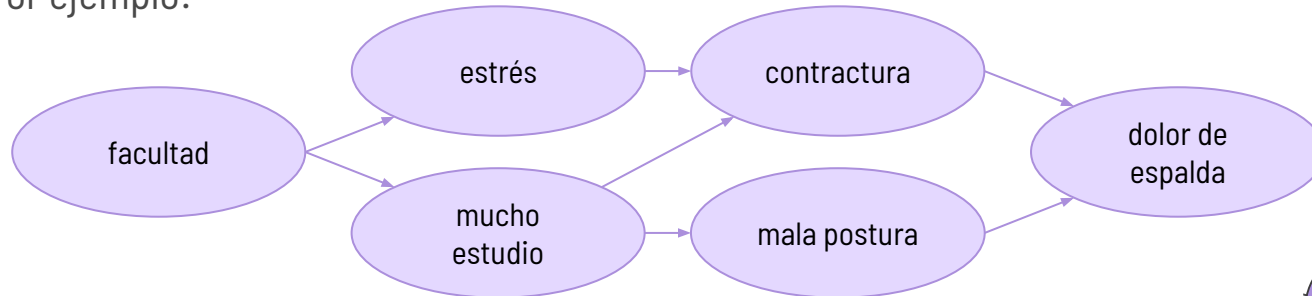
Redes Bayesianas

¿Qué pasa si dada una clase, los atributos no son independientes? → **Redes Bayesianas**

Intuición

- grafo acíclico dirigido
- cada nodo representa una variable aleatoria
- cada variable tiene asociada una probabilidad condicional
- la dependencia de las variables está determinada por la estructura del grafo

Por ejemplo:





Redes Bayesianas



Definición

Una red bayesiana para un conjunto de variables aleatorias $U = \{X_1, X_2, \dots, X_n\}$ consiste en un par $B = \langle G, T \rangle$ donde:

- G es un grafo acíclico dirigido en el cual cada nodo representa una de las variables X_1, X_2, \dots, X_n .
- El conjunto de arcos en el grafo representa relaciones de dependencia directa entre las variables.
- T es una tabla de probabilidades condicionales que está asociada a cada nodo en el grafo.

De esta forma, nuestro objetivo final es a partir de la tabla T , obtener las probabilidades que ocurra un evento dado un conjunto de variables → **Es posible realizar inferencias**





Redes Bayesianas



¿Cómo se realizan estas Inferencias?

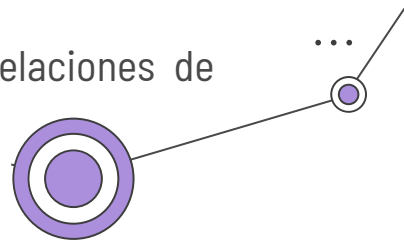
Se calcula la probabilidad conjunta utilizando el Teorema de la factorización de la probabilidad:

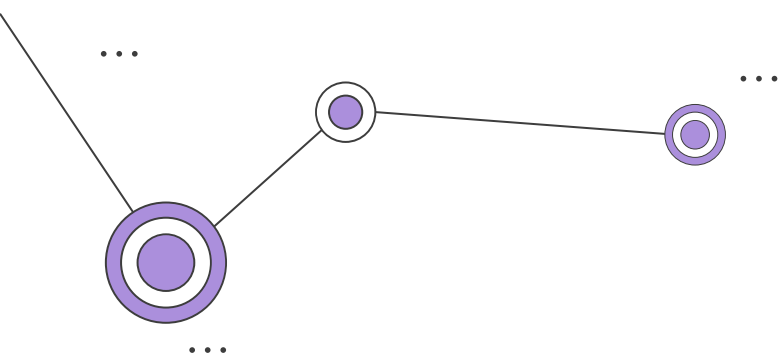
$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | Pa x_i)$$

$Pa x_i$ son todas las variables padres de x_i .

Aprendizaje

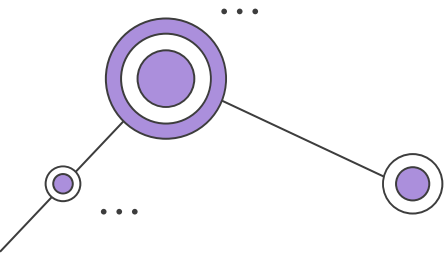
Obtener una red Bayesiana a partir de un conjunto de entrenamiento es un proceso de aprendizaje dividido en dos etapas:

- **Aprendizaje paramétrico:** usando los datos de entrenamiento y una estructura dada, se estiman probabilidades a priori y condicionales con frecuencias relativas ajustadas por la Corrección de Laplace para evitar ceros.
 - **Aprendizaje de la estructura:** Obtener la estructura de la red o las relaciones de dependencia e independencia entre las variables: **AlgK2**
- 

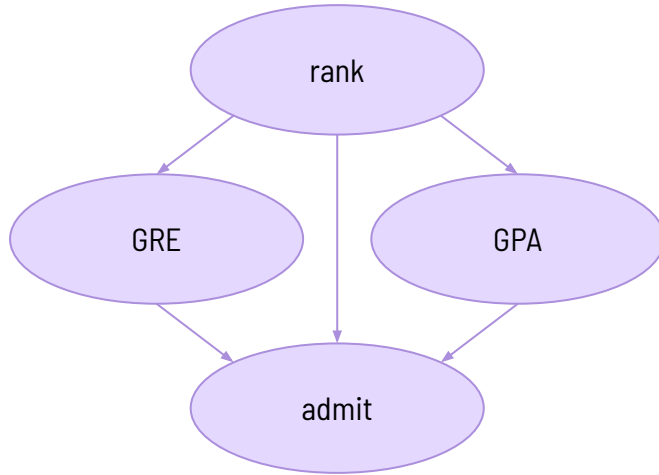


Ejercicio 3:

Admisión universitaria



Ejercicio a Resolver



En función de la información y la estructura presentadas, se deben calcular las siguientes probabilidades:

- $P(\text{admit} = 0 \mid \text{rank} = 1)$
- $P(\text{admit} = 1, \text{GRE} = 450, \text{GPA} = 3.5 \mid \text{rank} = 2)$

Utilizando una red Bayesiana.

Ejercicio a Resolver

| GRE | GPA | admit | 0 | 1 |
|-----------|---------|-------|-------|-------|
| | | rank | | |
| GRE < 500 | GPA < 3 | 1 | 0,714 | 0,286 |
| | | 2 | 0,545 | 0,455 |
| | | 3 | 0,929 | 0,071 |
| | | 4 | 0,900 | 0,100 |
| | GPA ≥ 3 | 1 | 0,500 | 0,500 |
| | | 2 | 0,792 | 0,208 |
| | | 3 | 0,792 | 0,208 |
| | | 4 | 0,824 | 0,176 |
| GRE ≥ 500 | GPA < 3 | 1 | 0,200 | 0,800 |
| | | 2 | 0,762 | 0,238 |
| | | 3 | 0,545 | 0,455 |
| | | 4 | 0,625 | 0,375 |
| | GPA ≥ 3 | 1 | 0,449 | 0,551 |
| | | 2 | 0,583 | 0,417 |
| | | 3 | 0,738 | 0,263 |
| | | 4 | 0,775 | 0,225 |

Se discretizaron las variables GRE y GPA de la siguiente manera:

- $GRE \geq 500$ o $GRE < 500$
- $GPA \geq 3$ o $GPA < 3$.

Luego, en base a la estructura definida con el grafo, se obtuvo la tabla de probabilidades condicionales asociadas a cada nodo del grafo.

Ejercicio a Resolver

Para calcular $P(\text{admit} = 0 \mid \text{rank} = 1)$, nos quedamos con los valores de la tabla anterior: rank = 1.

| | | admit | 0 | 1 |
|-----------|---------|-------|-------|-------|
| GRE | GPA | | | |
| GRE < 500 | GPA < 3 | | 0,714 | 0,286 |
| | GPA ≥ 3 | | 0,500 | 0,500 |
| GRE ≥ 500 | GPA < 3 | | 0,200 | 0,800 |
| | GPA ≥ 3 | | 0,449 | 0,551 |

$$P(\text{admit} = 0 \mid \text{rank} = 1) = \frac{P(\text{admit} = 0, \text{rank} = 1)}{P(\text{rank} = 1)}$$

$$= \frac{\sum P(\text{admit} = 0, \text{GRE}, \text{GPA}, \text{rank} = 1)}{\sum P(\text{rank} = 1, \text{GRE}, \text{GPA}, \text{admit})}$$

Considerando, en las variables que no fijadas, la sumatoria de todos los posibles valores que pueden tomar.

Finalmente, $P(\text{admit} = 0 \mid \text{rank} = 1)$: 0,46 → 46%.

Ejercicio a Resolver

Para calcular $P(\text{admit} = 1, \text{GRE} = 450, \text{GPA} = 3.5 \mid \text{rank} = 2)$, nos quedamos con los valores de $\text{rank} = 2$.

| | | admit | 0 | 1 |
|-----------|---------|-------|-------|-------|
| GRE | GPA | | | |
| GRE < 500 | GPA < 3 | | 0,545 | 0,455 |
| | GPA ≥ 3 | | 0,792 | 0,208 |
| GRE ≥ 500 | GPA < 3 | | 0,762 | 0,238 |
| | GPA ≥ 3 | | 0,583 | 0,417 |

$$P(\text{admit} = 1, \text{GRE} = 450, \text{GPA} = 3.5 \mid \text{rank} = 2)$$

$$= P(\text{admit} = 1, \text{GRE} < 500, \text{GPA} \geq 3 \mid \text{rank} = 2)$$

$$= \frac{\sum P(\text{admit} = 1, \text{GRE} < 500, \text{GPA} \geq 3, \text{rank} = 2)}{\sum P(\text{admit}, \text{GRE}, \text{GPA}, \text{rank} = 2)}$$

Considerando, en las variables que no fijadas, la sumatoria de todos los posibles valores que pueden tomar.

Finalmente, $P(\text{admit} = 1, \text{GRE} < 500, \text{GPA} \geq 3 \mid \text{rank} = 2)$: 0.052 → 5.2%.



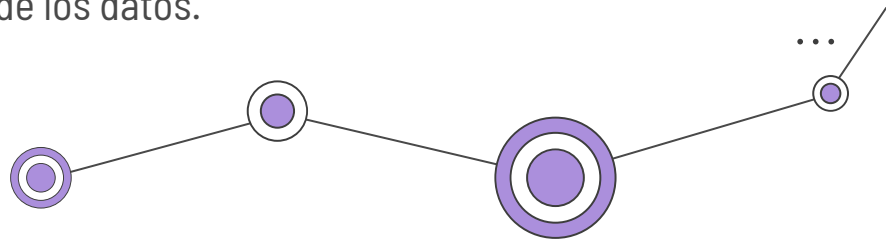
Ejercicio a Resolver

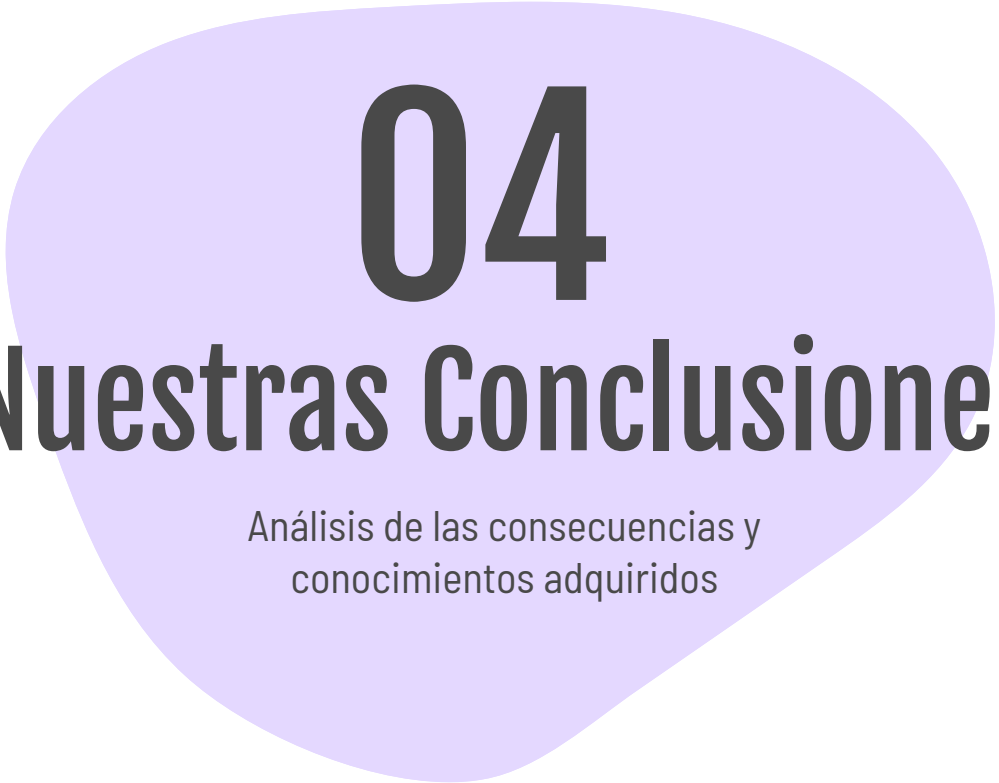



En caso que fuera **$P(\text{admit} = 1 \mid \text{rank} = 2, \text{GRE} = 450, \text{GPA} = 3.5)$** , calculamos $P(\text{admit} = 1 \mid \text{rank} = 2, \text{GRE} < 500, \text{GPA} \geq 3)$ que sale directamente con los valores de la tabla y da 0.208 \rightarrow **21%**

¿Cuál es el proceso de aprendizaje?

- **Aprendizaje paramétrico:**
 - Se estima la tabla de probabilidades de 'admit' que nos permite resolver los problemas planteados para esta estructura.
 - Esta tabla permite clasificar nuevos registros
- **Aprendizaje de la estructura:** no hizo falta en nuestro caso. Las relaciones y la estructura ya estaban dadas por el grafo de la red y la naturaleza de los datos.






04

Nuestras Conclusiones

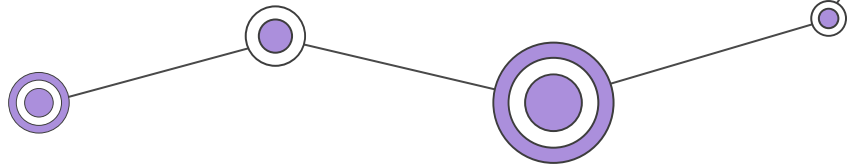
Análisis de las consecuencias y
conocimientos adquiridos





Nuestras Conclusiones



- Aún utilizando **Laplacian Smoothing**, un conjunto de train no balanceado (i.e que la probabilidad para un atributo sea nula) generará una tendencia muy marcada hacia una predicción en Naive Bayes.
 - Entrenar un Naive Bayes con un cierto conjunto no te asegura que lo vaya a **aprender** (Accuracy = 1.0, Error rate = 0.0)
 - Las Curvas ROC obtenidas de un Naive Bayes con 2 posibles etiquetas de salida (clases) son iguales pero con los **ejes invertidos**
- 



Nuestras Conclusiones



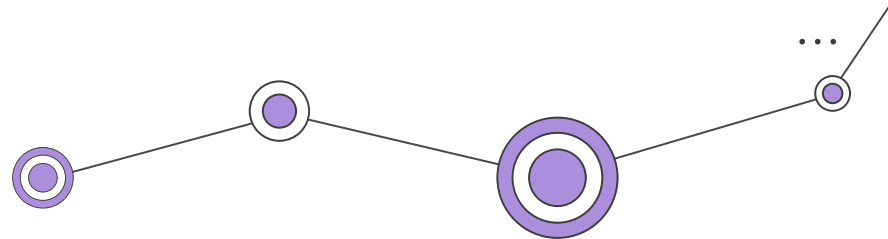
- El **pre-procesamiento** adecuado de datos es esencial para garantizar que el algoritmo de clasificación bayesiana trabaje con características **limpias y significativas**.
 - La **generalización** en los Clasificadores Bayesianos depende del conjunto que usa para entrenamiento. Un **sobreajuste** se produce cuando hay muchos datos en ese conjunto. **Validación k-cruzada** es una forma de evitarlo.
 - Al alterar el **umbral de decisión** en un Clasificador Bayesiano, se **alteran las métricas** asociadas al mismo (TP, TN, FP, FN, Accuracy, Recall, Precision, F1-score, etc.)
 - La **precisión alta** en las categorías "Ciencia y Tecnología" y "Salud" sugiere que sus palabras distintivas y bien separadas permite explotar mejor el modelo para hacer predicciones precisas.
- 



Nuestras Conclusiones



- La dirección de las relaciones entre variables en una red Bayesiana es fundamental
- Tanto Naive Bayes como las Redes Bayesianas proponen una probabilidad de ocurrencia de un evento pero no es absoluta
- En Redes Bayesianas
- Es posible realizar inferencias porque:
 - La estructura de la red está dada
 - Todas las variables de la red son observables y pueden calcularse todas las probabilidades condicionales existentes en la red



¡Gracias!
¿Preguntas?

