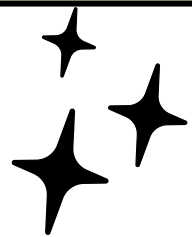


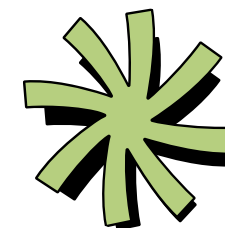
Desempeño Estudiantil



Materia: Inteligencia artificial

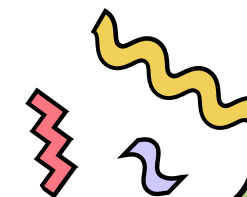
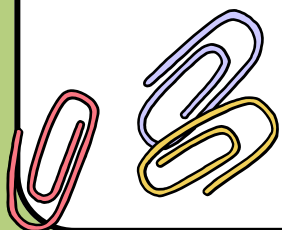


Equipo: CodeMonkeys



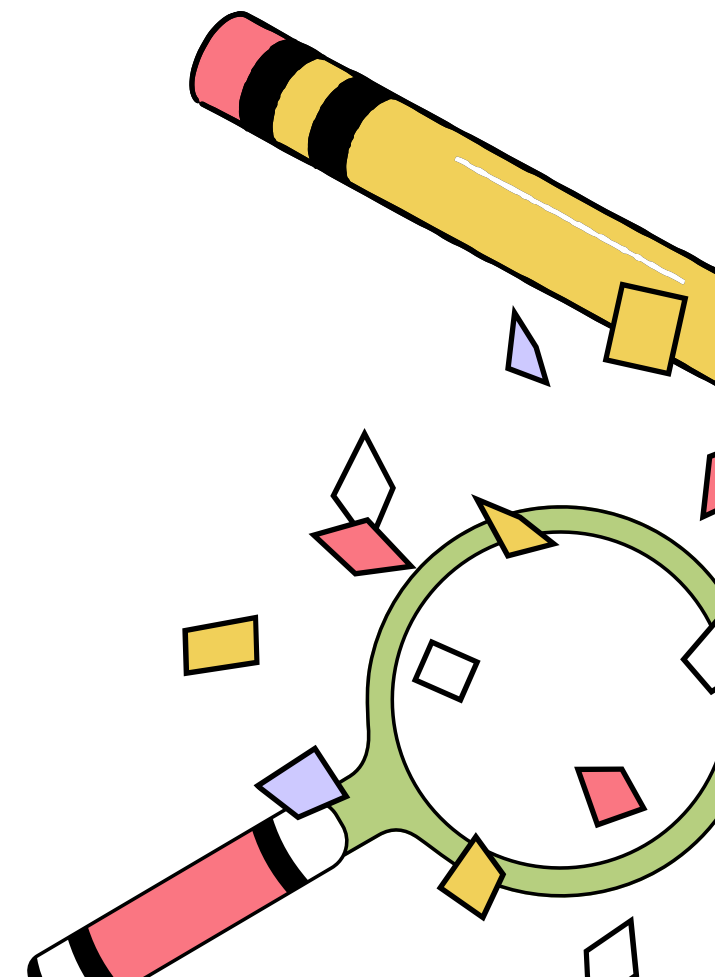
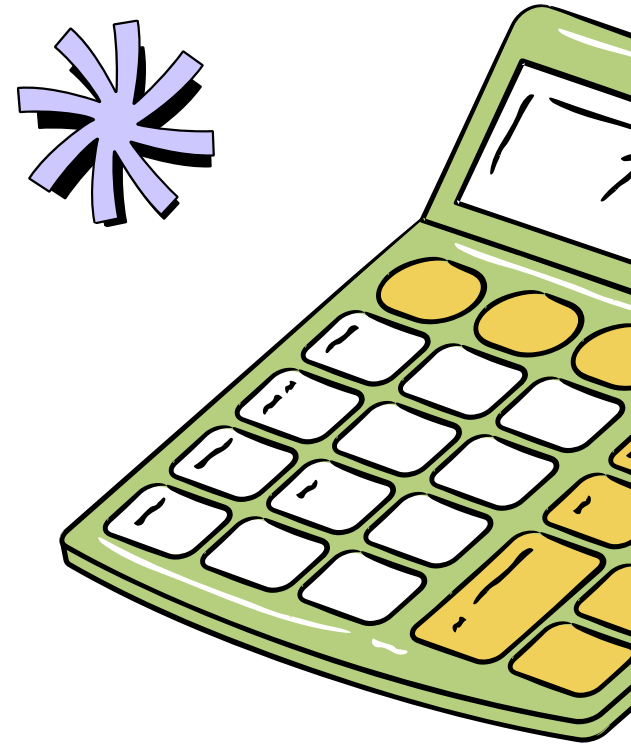
Integrantes:

Díaz Villegas Erick
Hernandez Garcia Carlos Zahid
Ordaz Rebollo Ximena



Índice de contenidos

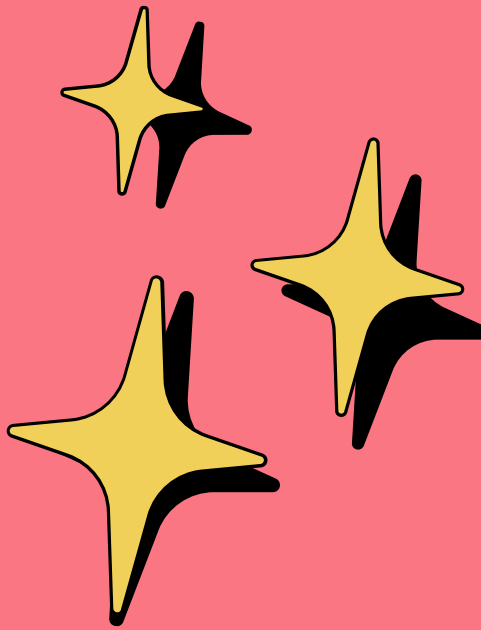
- ☐ Planteamiento del problema, objetivo y justificación
- ☐ Visualizaciones gráficas
- ☐ Tarea de Machine Learning a resolver
- ☐ Metodología de la experimentación
- ☐ Resultado y Conclusiones
- ☐ Referencias



Planteamiento del problema

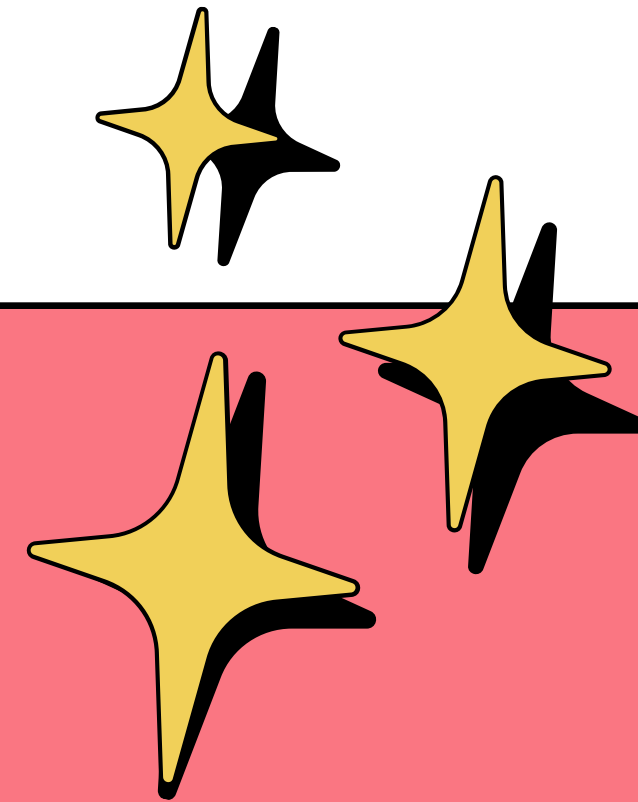
La educación es un pilar fundamental del desarrollo personal como de una nación. Existen áreas de oportunidad en el sector educativo, un ejemplo de ello en México son **problemas de equidad e inclusión, entornos de desarrollo deficientes, así como falta de cobertura educativa.**

Los resultados de los estudiantes y encuestas sobre su entorno son puntos por abordar para mejorar la calidad educativa. Por lo que sería adecuado considerar en qué casos los alumnos tienen mejor o peor desempeño.



Objetivo

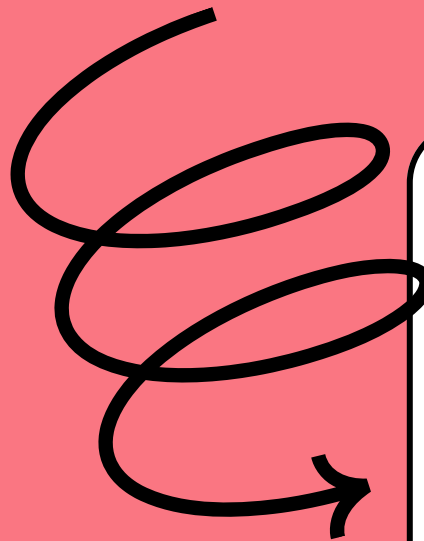
Utilizar un algoritmo clasificador (Naive Bayes) para realizar un modelo de entrenamiento en el que pueda predecir qué estudiantes aprobarán o reprobarán según las variables que lo rodean, mediante discretizaciones de variables y probabilidades de clases y características.



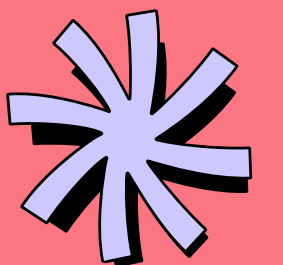
Justificación

La riqueza de las naciones *“ahora depende del conocimiento y las habilidades del capital humano como principales detonantes del crecimiento económico.”* (Gutiérrez A., 2007, p. 3)

Los resultados de la prueba PISA en México del año 2022 decrementaron a comparación de los resultados del año 2018.



En 2012 la proporción de calificaciones inferiores a un nivel 2 aumentó en 11 % en matemáticas.



Tarea de Machine Learning a resolver

Naive bayes

- Gaussiano
- Multinomial
- Mixto:
 - Gaussiano
 - Gumbel
 - Exponencial



Metodo de ensamble

Independencia de las variables

Combinar múltiples modelos para crear uno mas robusto

Tarea de Machine Learning a resolver

Fundamentos Teoricos

$$\underbrace{\operatorname{argmax}}_{k \in \{1, \dots, K\}} \left[P(C_k | \mathbf{x}) = \frac{P(C_k)P(\mathbf{x} | C_k)}{P(\mathbf{x})} \right]$$

$$P(C_k | \mathbf{x}) = P(C_k) \prod_{i=1}^n P(x_i | C_k)$$

Desbordamiento
Numerico

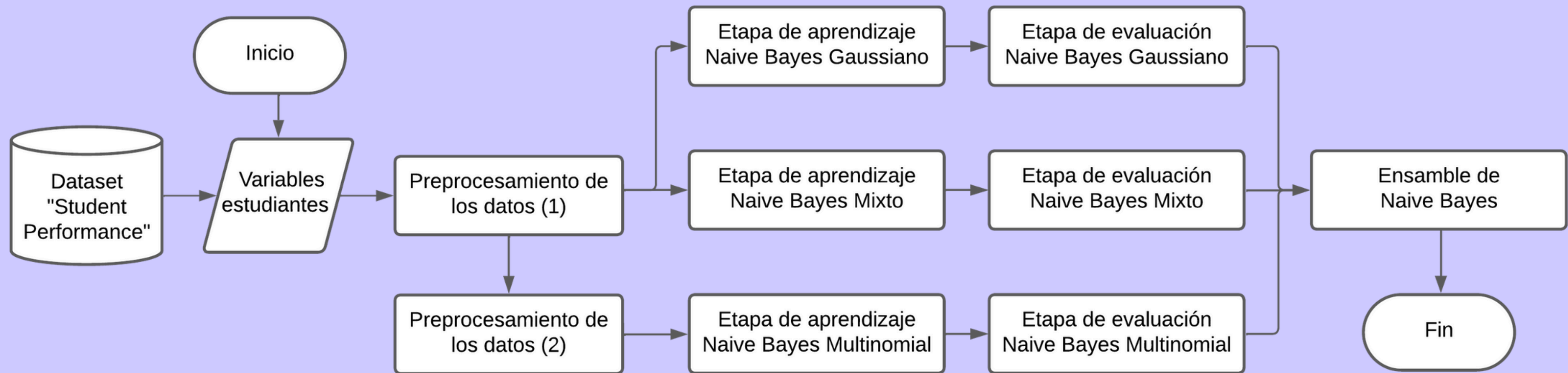
$$\log(P(C_k | \mathbf{x})) = \log \left(P(C_k) \prod_{i=1}^n P(x_i | C_k) \right)$$

Valor fuera
del rango
representable

$$\log(P(C_k | \mathbf{x})) = \log(P(C_k)) + \log(P(x_1 | C_k)) + \dots + \log(P(x_n | C_k))$$

Metodología de la experimentación

Esquema de la experimentación



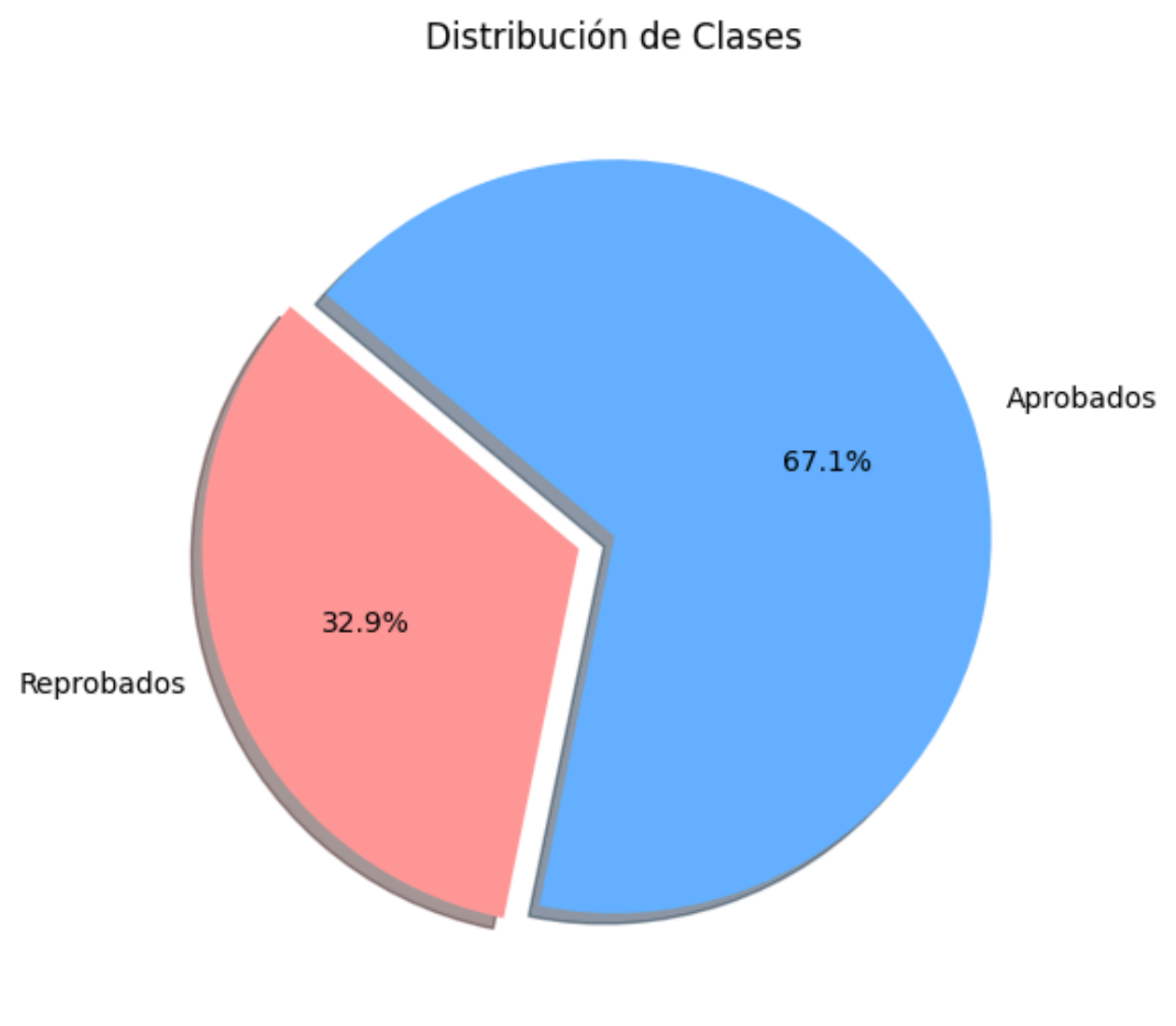
Descripción de los datos

- 395 Registros:
 - 130 Reprobados.
 - 265 Aprobados.

Dos escuelas públicas de Brasil. (2005 – 2006).

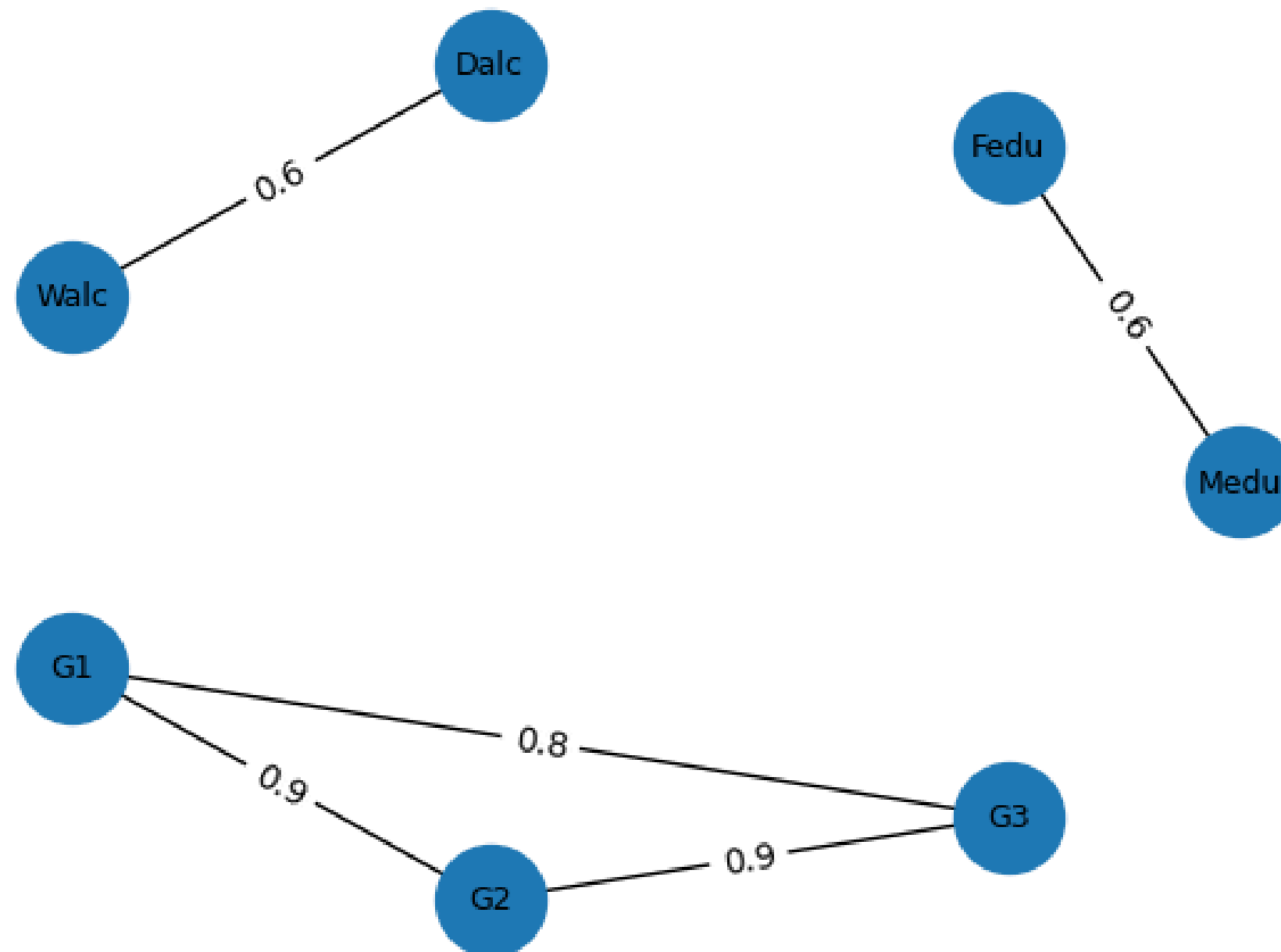
Nombre de la Variable	Tipo
age	Entero (15 a 22)
Medu	Categórica (0 a 4)
Fedu	Categórica (0 a 4)
traveltime	Categórica (1 a 4)
studytime	Categórica (1 a 4)
failures	Categórica (1 a 4)
famrel	Categórica (1 a 5)
freetime	Categórica (1 a 5)

Nombre de la Variable	Tipo
goout	Categórica (1 a 5)
Dalc	Categórica (1 a 5)
Walc	Categórica (1 a 5)
health	Categórica (1 a 5)
absences	Entero (0 a 93)
G1	Entero (0 a 20)
G2	Entero (0 a 20)
G3	Entero (0 a 20)



Preprocesamiento de los datos (1)

Grafo de dependencias

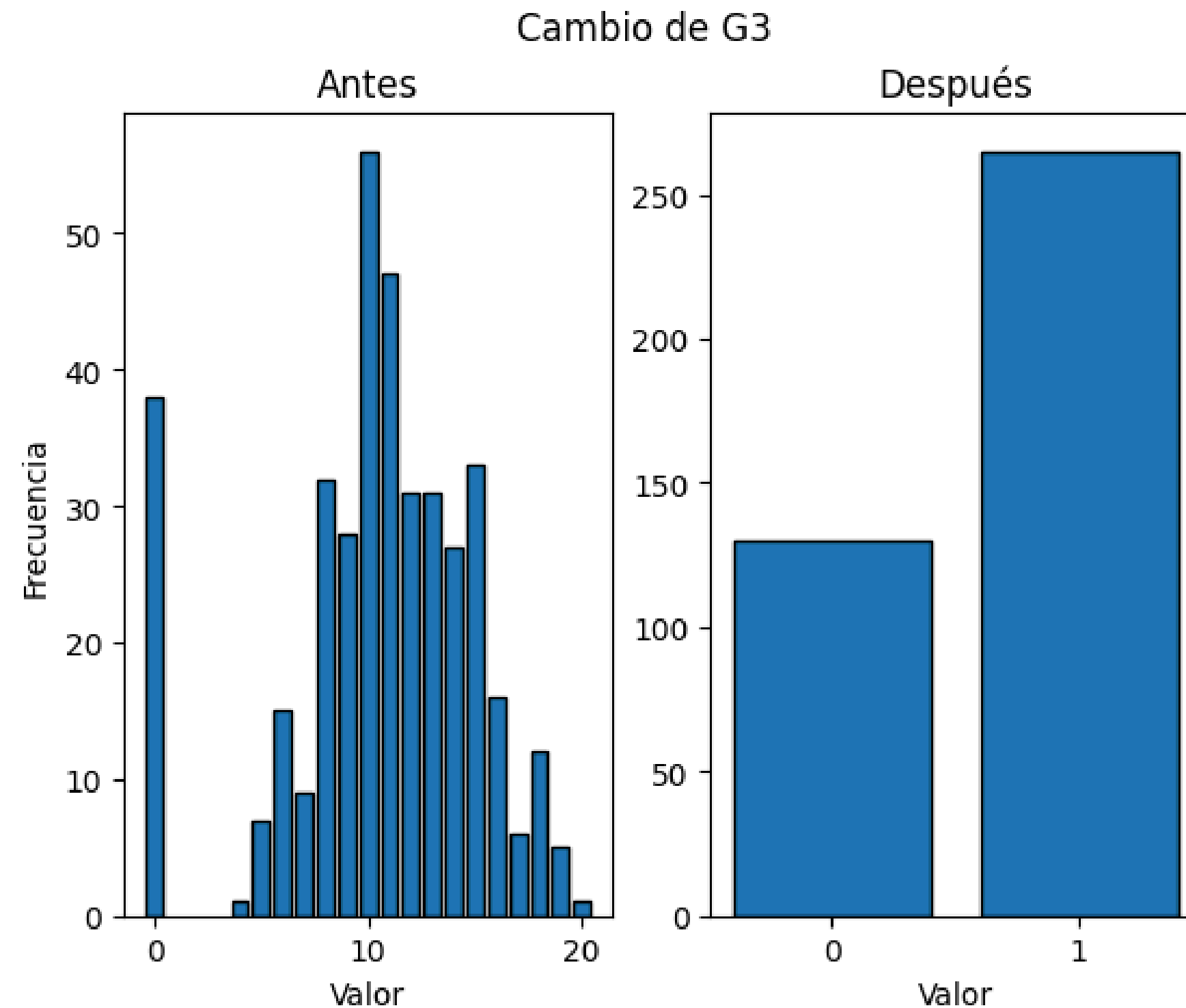


Valor crítico: 0.429 (16 variables)

Descarte de las variables:

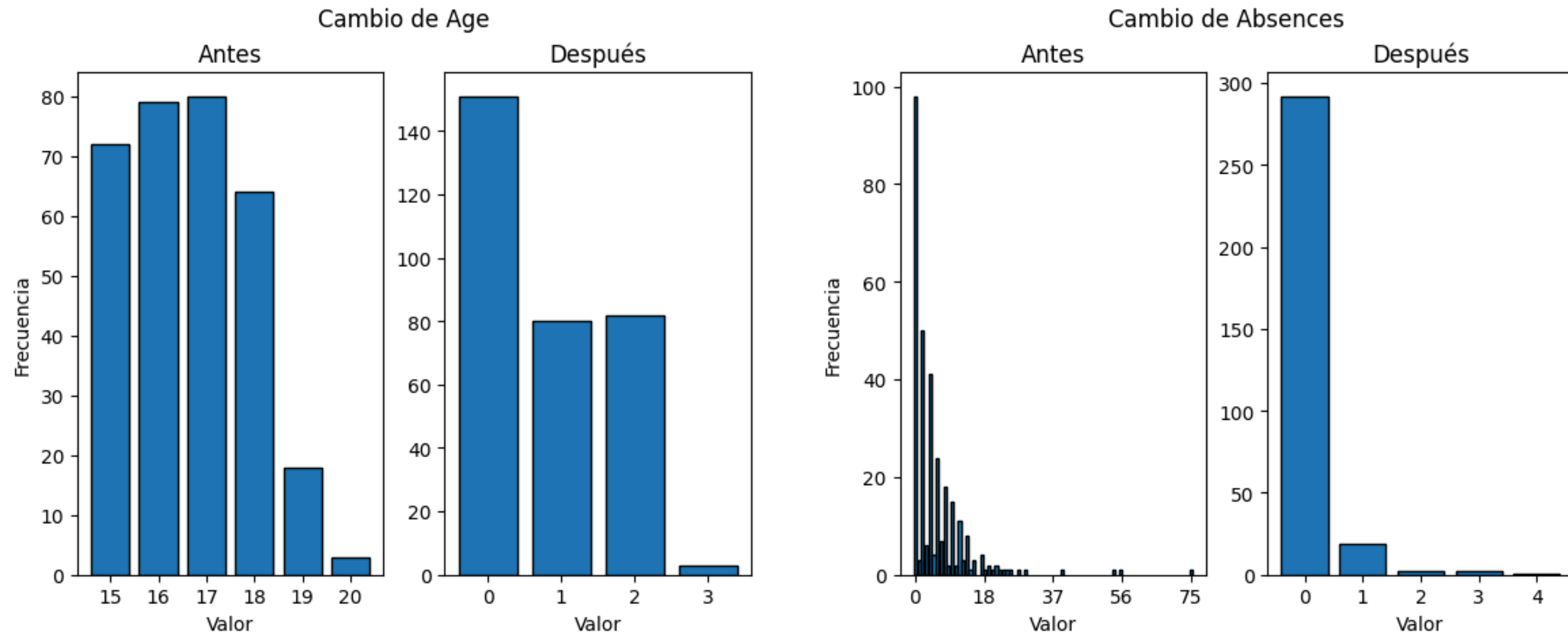
- Medu
- Walc
- G1
- G2

Preprocesamiento de los datos (1)



Discretización de la etiqueta (G3).

Preprocesamiento de los datos (2)



Discretización de:

- Age
- Absences

Etapa de aprendizaje

Naive Bayes Gaussiano				
	0		1	
Age	16.98	1.58	16.47	1.44
Fedu	2.41	1.13	2.62	1.19

Media

Varianza

Naive Bayes Mixto						
	0			1		
Age	16.98	17	1.58	16.47	16	1.44
Fedu	2.41	3	1.13	2.62	2	1.19

Media

Moda

Varianza

Naive Bayes Multinomial										
	0					1				
Age	0.358	0.292	0.368	0.019	0.009	0.548	0.243	0.214	0.014	0.005
Fedu	0.009	0.264	0.283	0.292	0.198	0.01	0.176	0.314	0.229	0.295

0

1

2

3

4

Priori	
0	1
0.3354	0.6645

Suavización de Laplace.
Se sumó +1 a las
frecuencias.

Etapa de consulta

Para cada Naive

$$\log(P(C_k|x)) = \log(P(C_k)) + \log(P(x_1|C_k)) + \dots + \log(P(x_n|C_k))$$

	NBG
1	0
2	1
3	0
4	1
5	1
6	0
7	0
8	1
9	1
10	1

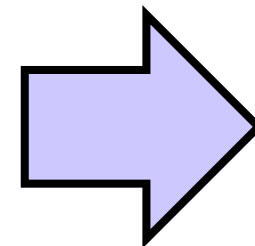
	NBM
1	1
2	1
3	0
4	0
5	1
6	1
7	0
8	1
9	1
10	1

	NBMix
1	0
2	1
3	0
4	1
5	1
6	1
7	0
8	1
9	1
10	1

Etapa de consulta

Ensamble de Naive Bayes

	Ensamble		
1	0	1	0
2	1	1	1
3	0	0	0
4	1	0	1
5	1	1	1
6	0	1	1
7	0	0	0
8	1	1	1
9	1	1	1
10	1	1	1



	Moda
1	0
2	1
3	0
4	1
5	1
6	1
7	0
8	1
9	1
10	1

Resultados

Resultados	
Gaussiano	81.01%
Multinomial	69.62%
Mixto	81.01%
Ensamble	82.28%

¿Se cumplió el objetivo?

Se obtuvo un rendimiento del 82%

El objetivo no se cumplió dado que se esperaba un rendimiento de al menos 85%.

El bajo rendimiento del clasificador puede ser a causa de:

- Falta de datos

Trabajos a futuro

- Recolectar información de estudiantes en México.
- Aumentar la cantidad de datos que se analizan.
- Implementar otros modelos de clasificación.

Contacto



Zahid

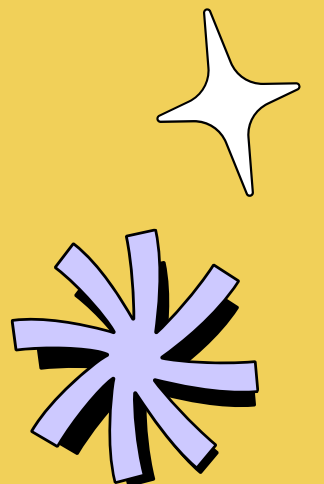
zahidhdez@outlook.com

Ximena

ordazr.ximena@gmail.com

Erick

erickdiazvillegas@gmail.com



Referencias



Cortez, Paulo. (2014). Student Performance. UCI Machine Learning Repository. <https://doi.org/10.24432/C5TG7T>.

Cortez, P., & Silva, A.M. (2008). Using data mining to predict secondary school student performance.

Gutiérrez, A. (2007). EDUCATION AND ECONOMIC GROWTH [Tesis doctoral, Atlantic International University]. Repositorio institucional de Atlantic International University. <https://www.aiu.edu/spanish/education-and-economic-growth.html>

OECD. (2023). PISA 2022 Results: Country Notes - Mexico. OECD Publishing. <https://www.oecd.org/publication/pisa-2022-results/country-notes/mexico-519eaf88#chapter-d1e11>

Ramsey, P. H. (1989). Critical Values for Spearman's Rank Order Correlation. Journal of Educational Statistics, 14(3), 245-253. doi:10.3102/10769986014003245

Gracias por su atención.

