

*ANÁLISIS
DE DATOS
PARA LA
PREDICCIÓN
DEL ÉXITO
ACADÉMICO
MEDIANTE
ÁRBOLES DE
DECISIÓN*



Presentación del Equipo



Valentina
Moreno
Ramírez



Alejandra
Palacio
Jaramillo



Miguel
Correa

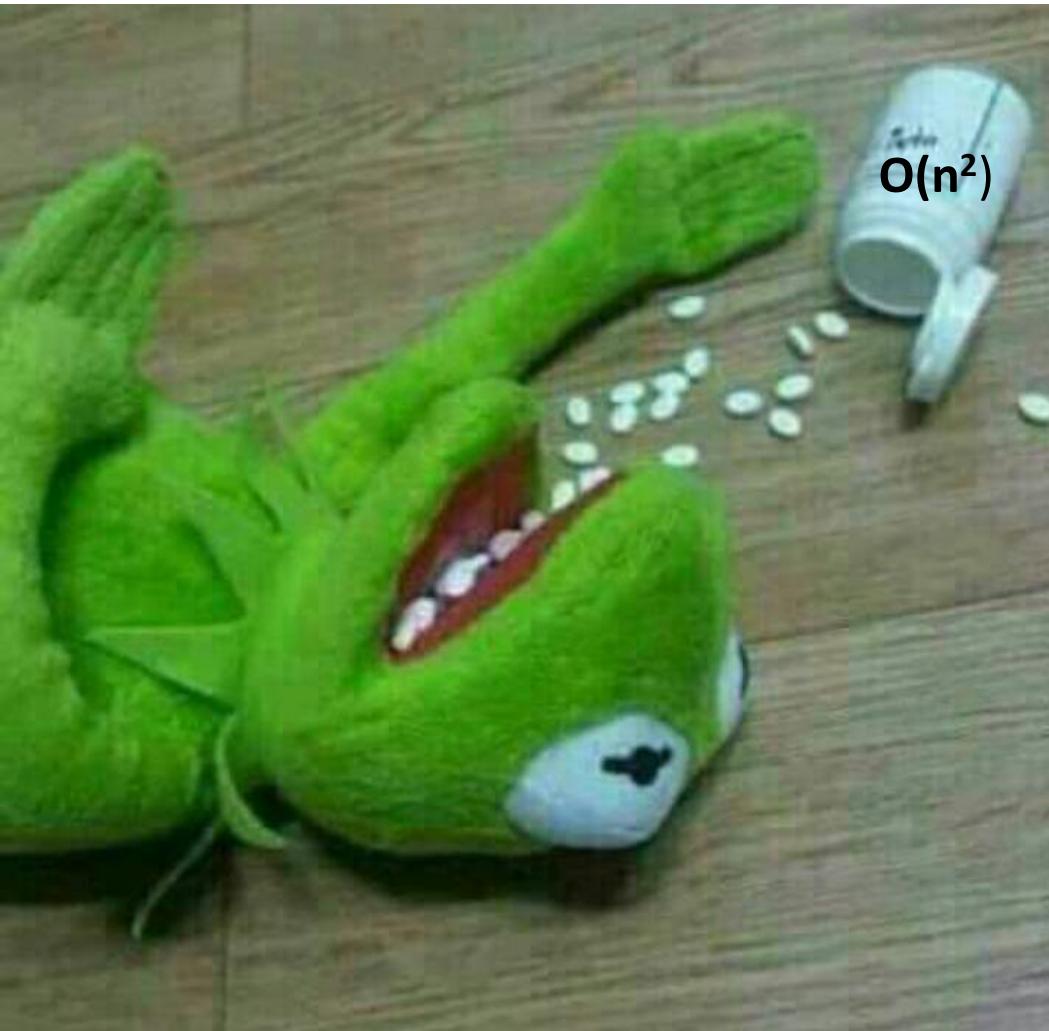


Mauricio
Toro



<http://github.com/apalacioj/proyecto>

¿Cómo nos sentimos durante el proyecto?



Terminamos el proyecto y seguimos frescas como lechuga

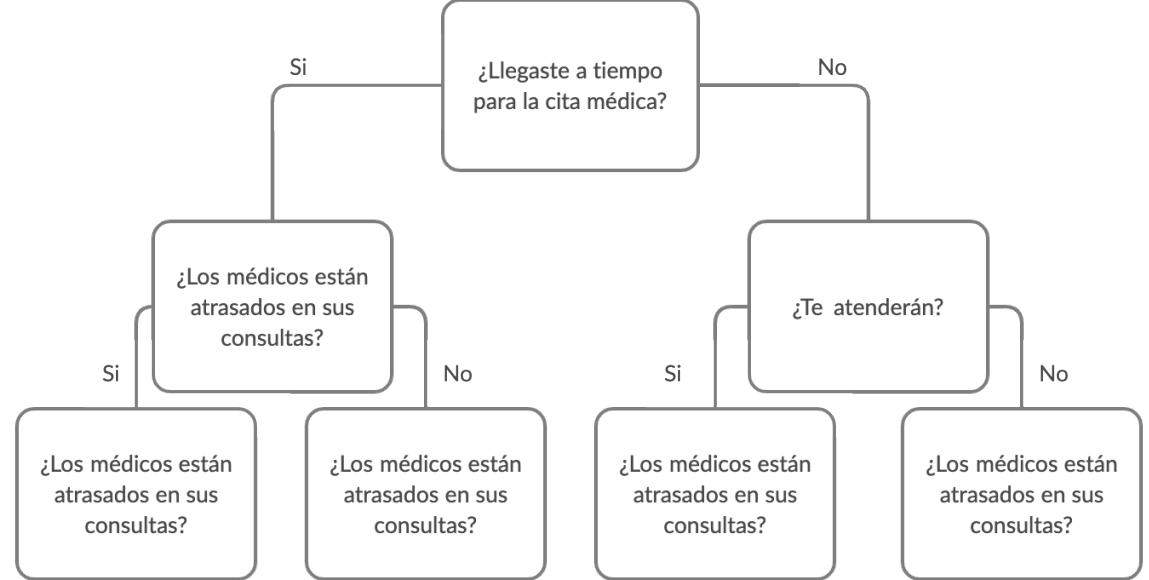


Coctél para remediar el insomnio

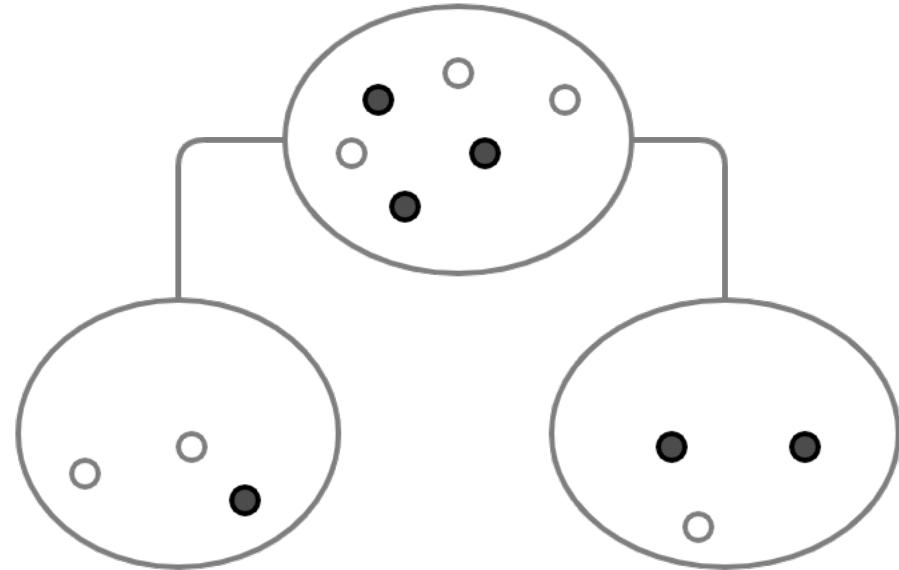


- 20 ml de vodka
- 20 ml de ginebra
- 20 ml de ron
- 20 ml de tequila
- 15 ml de licor de naranja
- 15 ml de zumo de limón o lima
- 70 ml de zumo de naranja
- Hielo
- Azúcar
- Rodajas de limón

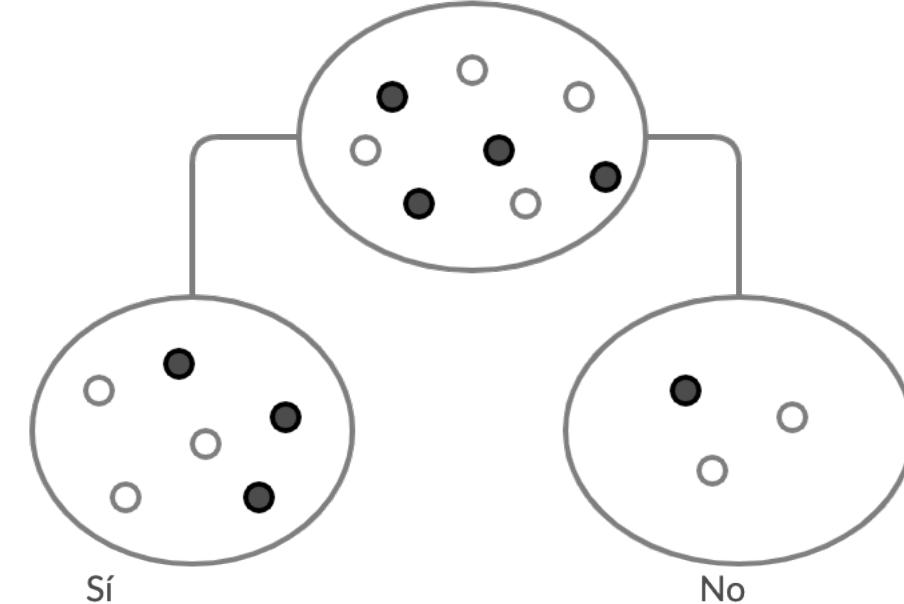
Diseño del Algoritmo



Algoritmo para construir un árbol binario de decisión usando el algoritmo CART. En este ejemplo, mostramos un modelo para predecir si una persona tardará más de lo habitual en una cita médica.



Esta división está basada en la condición “puntaje en matemáticas ≥ 58 .” Para este caso, la impureza Gini de la izquierda es 0.44, la impureza Gini de la derecha es 0.44 y la impureza ponderada es de 0.44.

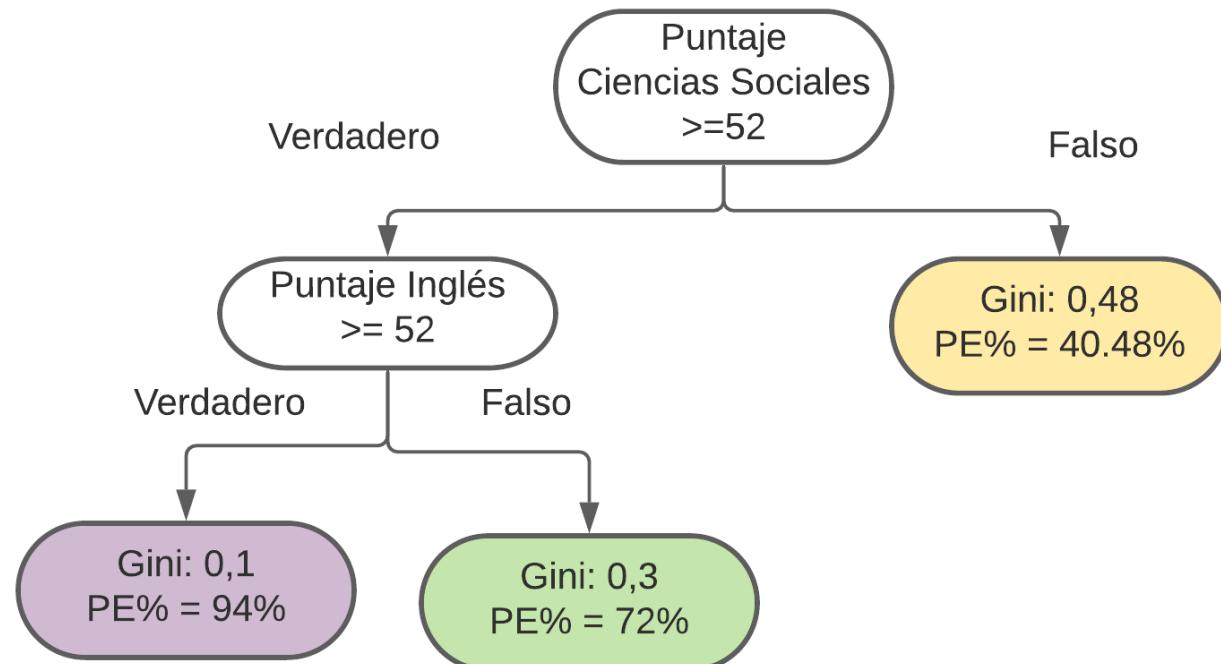


Esta división está basada en la condición “puntaje en lenguaje ≥ 48 .” Para este caso, la impureza Gini de la izquierda es 0.5, la impureza Gini de la derecha es 0.44 y la impureza ponderada es 0.46.

	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(N^2 \cdot M \cdot 2^M)$	$O(N \cdot M \cdot 2^M)$
Validación del modelo	$O(N \cdot M)$	$O(1)$

Complejidad en tiempo y memoria del algoritmo (Se usó el Algoritmo CART Para realizar el análisis) N será el número de filas y M el número de columnas.





Características Más Relevantes



Mátemática

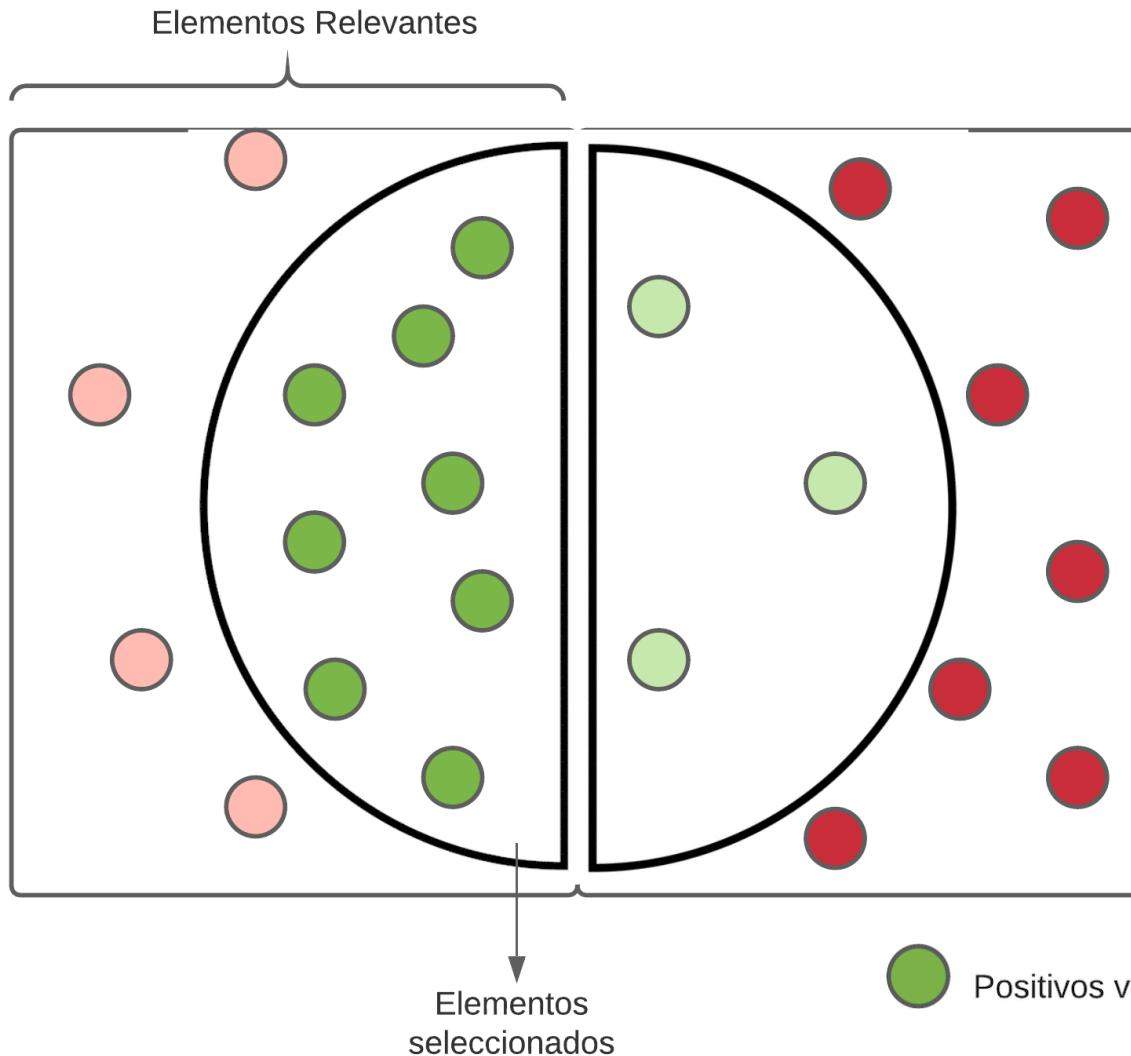


Inglés



Ciencias Sociales

Un árbol de decisión para predecir el resultado del Saber Pro usando los resultados del Saber 11. Violeta representa nodos con alta probabilidad de éxito; verde media probabilidad; y rojo baja probabilidad.



$$\text{Exactitud} = \frac{\text{Positivo verdadero} + \text{Negativo verdadero}}{\text{Total}}$$

$$\text{Precisión} = \frac{\text{Positivo verdadero}}{\text{Total de positivos predichos}}$$

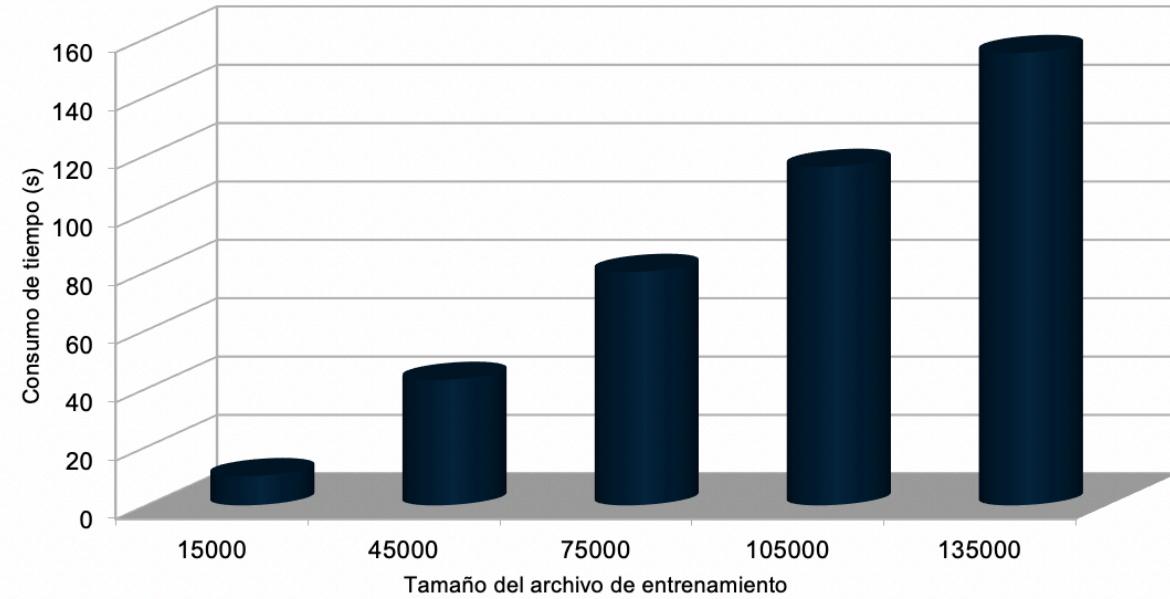
$$\text{Sensibilidad} = \frac{\text{Positivo verdadero}}{\text{Positivos Reales}}$$

	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.79	0.71
Precisión	0.8	0.79
Sensibilidad	0.76	0.72

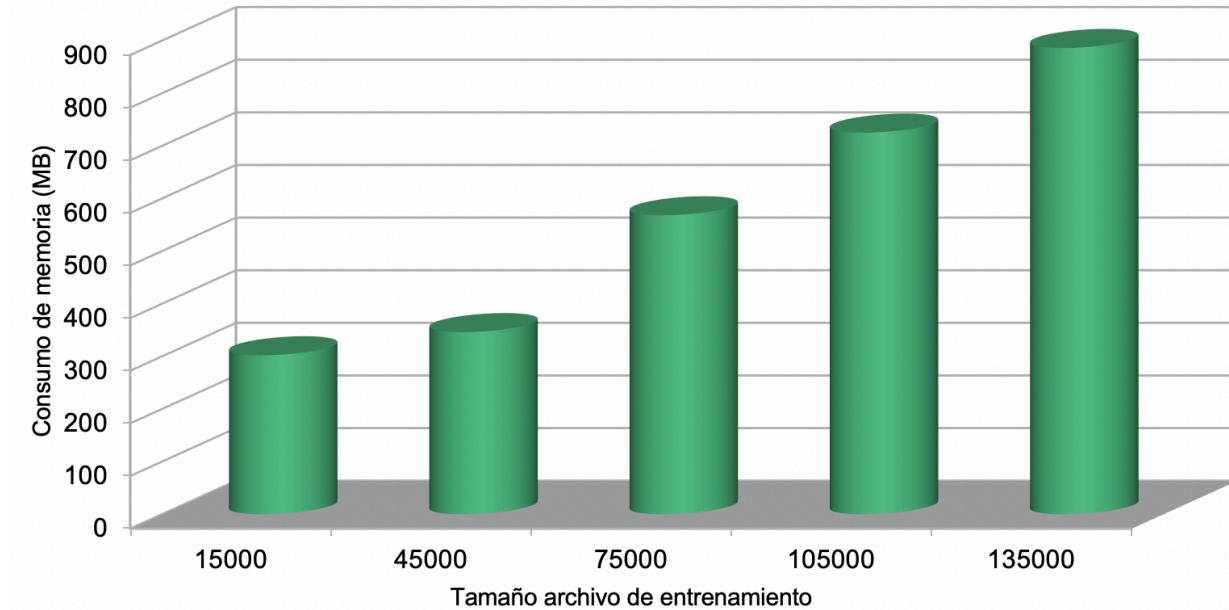
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria



Palacio and V. Moreno. Análisis de datos para la predicción del éxito académico mediante árboles de decisión . ArXiv e-prints, Nov. 2020.

Available
at:<https://arxiv.org/submit/3475741/view>

arXiv.org > user

Your arXiv.org account: vmorenor16

E-mail: vmorenor@eafit.edu.co Affiliation: EAFIT
Name: Valentina Moreno Ramírez URL:
Default Category: cs.DS Country: Colombia
Groups: cs and math Career Status: Other

[Change User Information](#) | [Change Password](#) | [Change Email](#) | [Disable MathJax](#) ([What is MathJax?](#))

Article Submissions					Update	Delete	Unsubmit
Identifier	Type	Title	Status	Actions	Expires		
submit/3475741	New	Data analysis for the prediction of academic success through decision trees	submitted				

Are there articles you are an author of on arXiv that are not listed here?

If you have the paper password use [the Claim Ownership with a password form](#)

If you do not have the paper password or are claiming multiple papers use [the Claim Authorship form](#).

For more information please see our help page on [authority records](#).

Reunión en "General" ended: 6h 57m

Reply

Reunión en "General" started

3 replies from you

Meeting ended: 4h 25m

Reply

General ended: 1h 48m

Reply



Así se queda después de una llamada de 7 horas





Trabajo en equipo:

vmorenor16 committed on 11 Oct

Add files via upload
vmorenor16 committed on 11 Oct

Verified d799b45 <>

Add files via upload
vmorenor16 committed on 11 Oct

Verified bc83c35 <>

➤ Commits on Oct 6, 2020

Add files via upload
vmorenor16 committed on 6 Oct

Verified 15c521c <>

Add files via upload
vmorenor16 committed on 6 Oct

Verified a75fca9 <>

Update ReadCSV.java
apalacioj committed on 6 Oct

Verified e3a1669 <>

➤ Commits on Aug 16, 2020

Add files via upload
apalacioj committed on 16 Aug

Verified b9260ca <>

Add files via upload
vmorenor16 committed on 16 Aug

Verified 0a9c894 <>

Add files via upload
vmorenor16 committed on 16 Aug

Verified a99d877 <>



Estamos listas para el tercer semestre



¡GRACIAS!