

Estructuras
de predicción
de éxito
Académico
en prueba
saber pro.



Presentación del Equipo



Tomás
Marin



Juan
Vera



Miguel
Correa



Mauricio
Toro



<https://github.com/Vera3588/ST0245-001/tree/master/proyecto>



TOM COLLINS

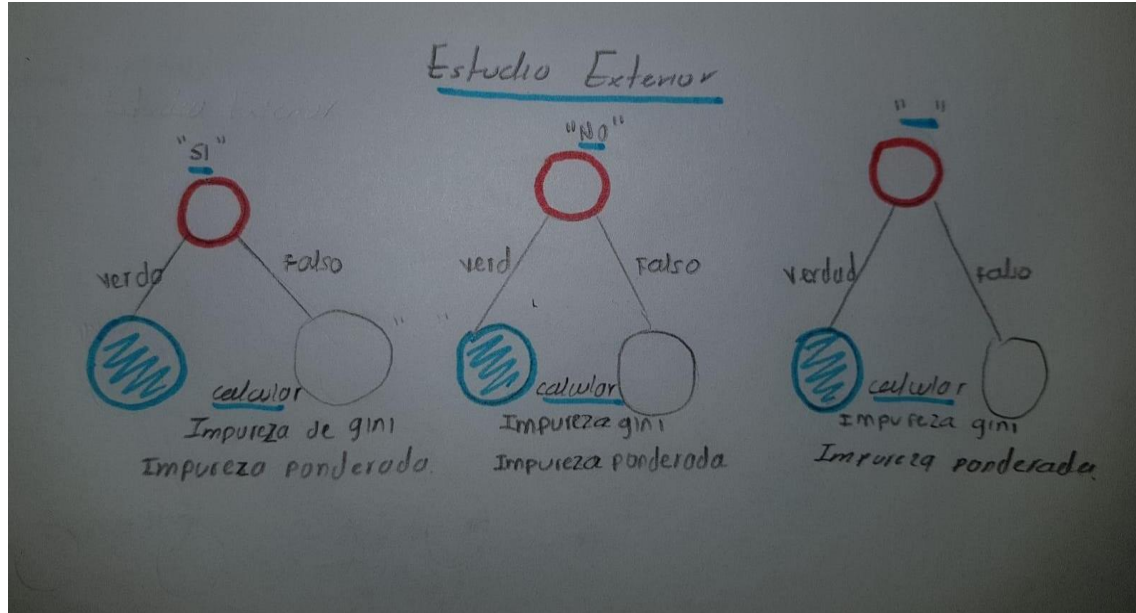


Esta bebida es de antes de la invención de los refrescos y en este caso se sugiere emplear Old Tom Gin, ya que brinda más sabor.

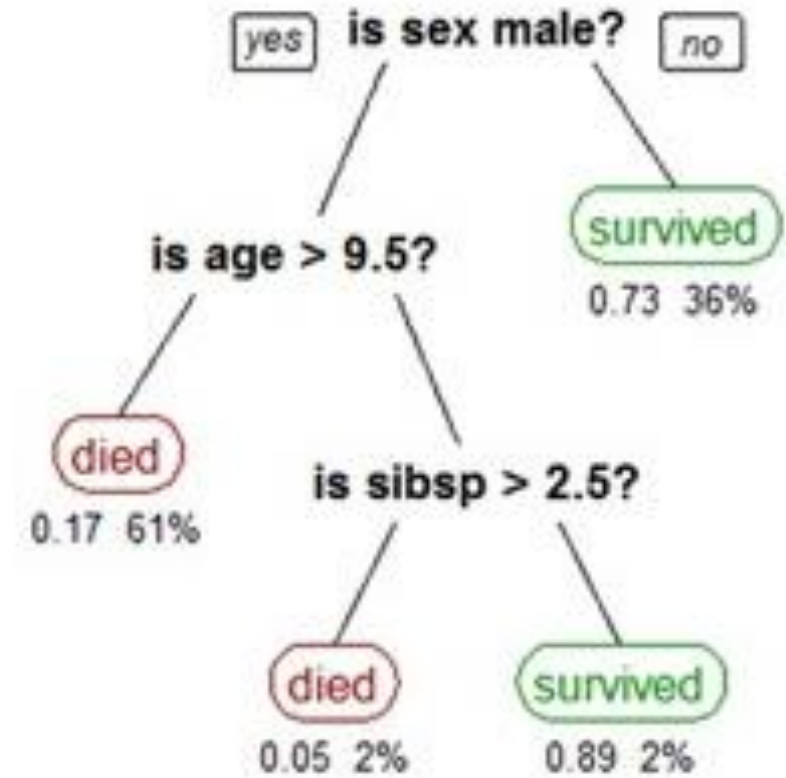
Ingredientes:

- 60 mililitros de ginebra
- 30 mililitros de jugo de limón
- 1 cucharadita de azúcar
- soda (agua carbonatada, sifón o gaseosa)
- rodajas de limón y hierbabuena para decorar.

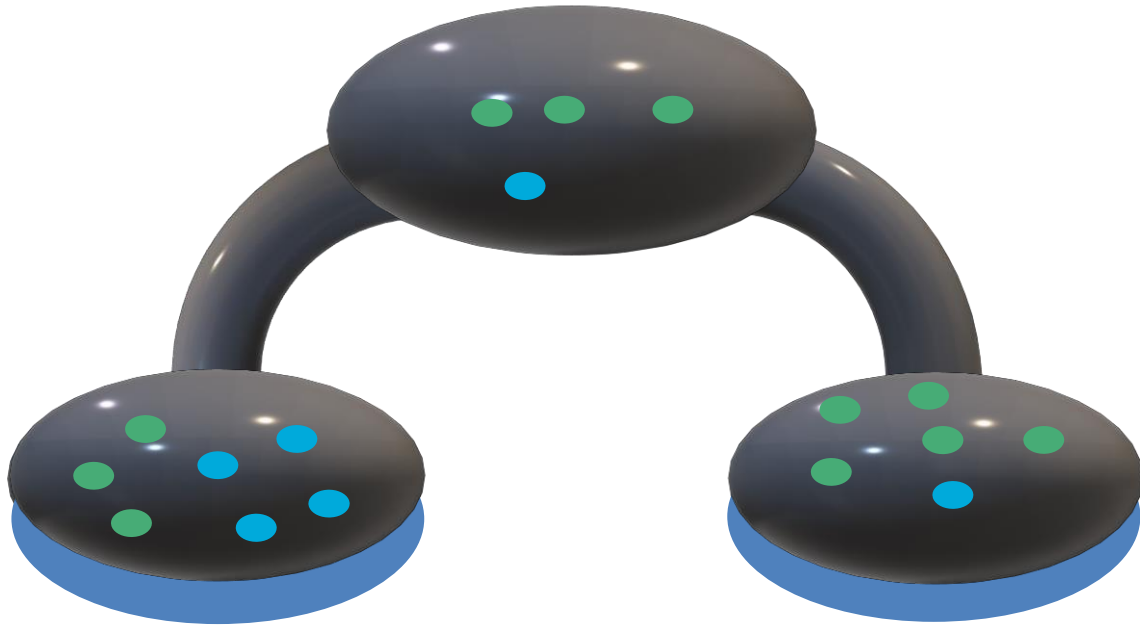
Diseño del Algoritmo



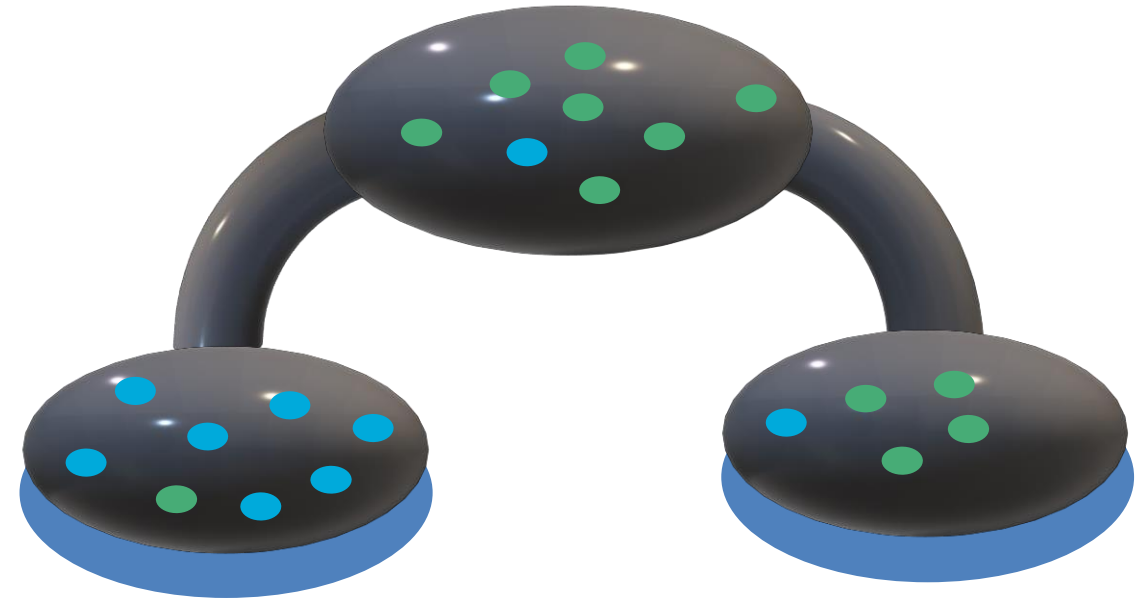
La estructura de datos que vamos a usar para este proyecto es un árbol de decisión CART el cual nos ayudara a predecir los resultados de los exámenes tomando varias variables que afectan a los estudiantes.



División de un nodo



Esta división está basada en “Estrato == 3” en este caso, la impureza Gini de la izquierda equivale a 0.74, la impureza Gini de la derecha equivale a 0.27 y la impureza ponderada es de 0.39.



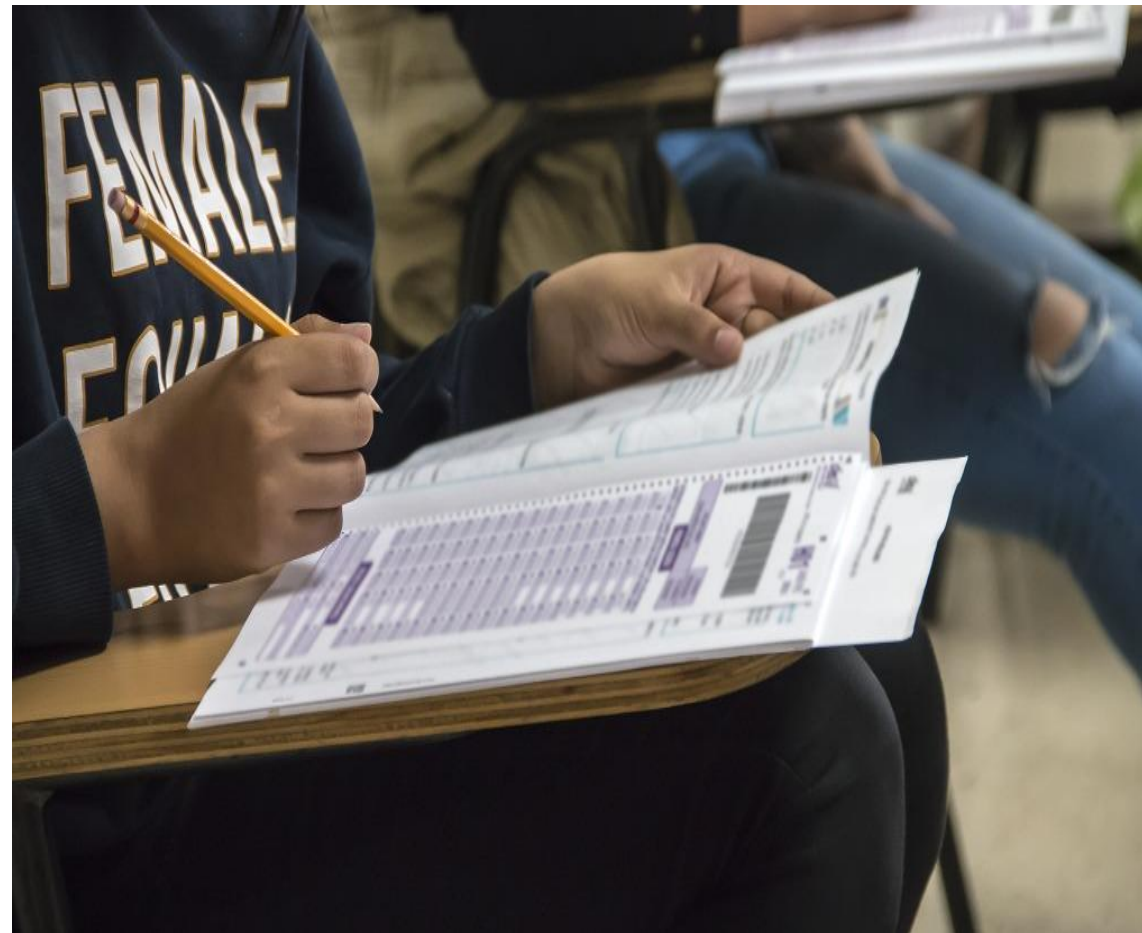
Esta división está basada en “simulacrotipoicfes == SI” en este caso, la impureza Gini de la izquierda equivale a 0.9, la impureza Gini de la derecha equivale a 0,08 y la impureza ponderada es de 0.16.

Complejidad del Algoritmo

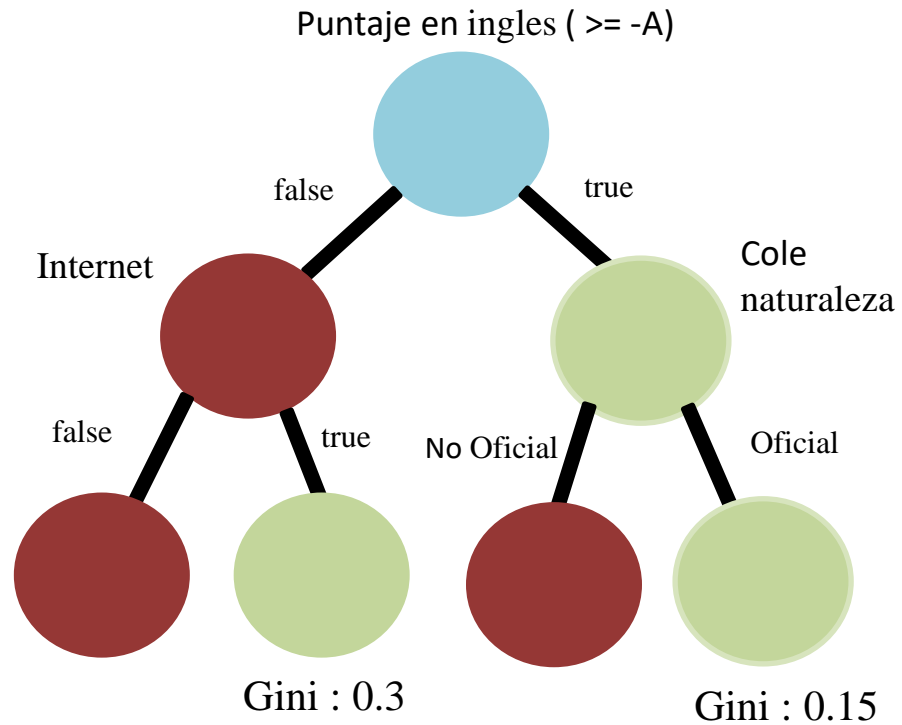


	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(n \cdot 2^M)$	$O(N \cdot M \cdot 2^M)$
Validación del modelo	$O(N \cdot M)$	$O(N)$

- La complejidad es esta debido a que sería el número de filas por columnas y como cada piso cuenta con 2 nodos mínimo entonces por eso sería 2 a la m.
- N siendo las filas.
- M siendo las columnas.



Modelo de Árbol de Decisión



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

Características Más Relevantes



Inglés

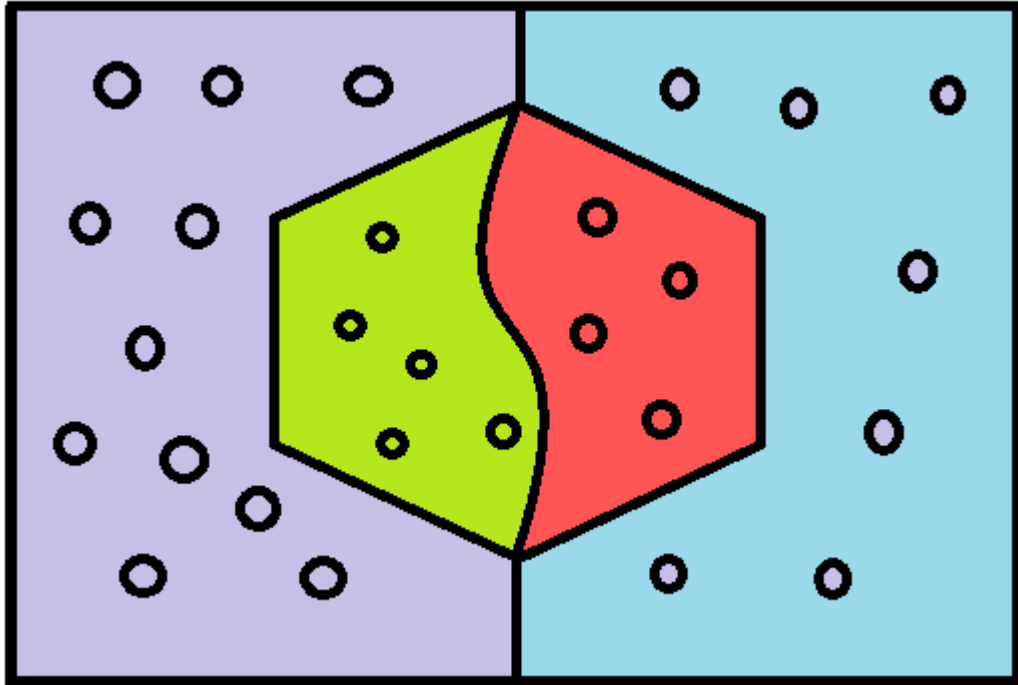


Género



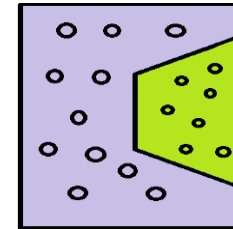
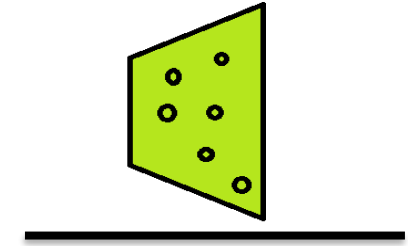
Residencia

Métricas de Evaluación

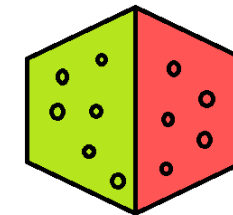
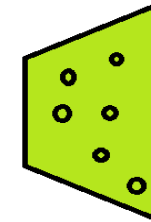


-  Falsos negativos
-  verdaderos negativos
-  Verdaderos positivos
-  Falsos positivos

Sensibilidad =



Precisión =





	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.5	0.6
Precisión	55.5%	54%
Sensibilidad	0.4	0.4

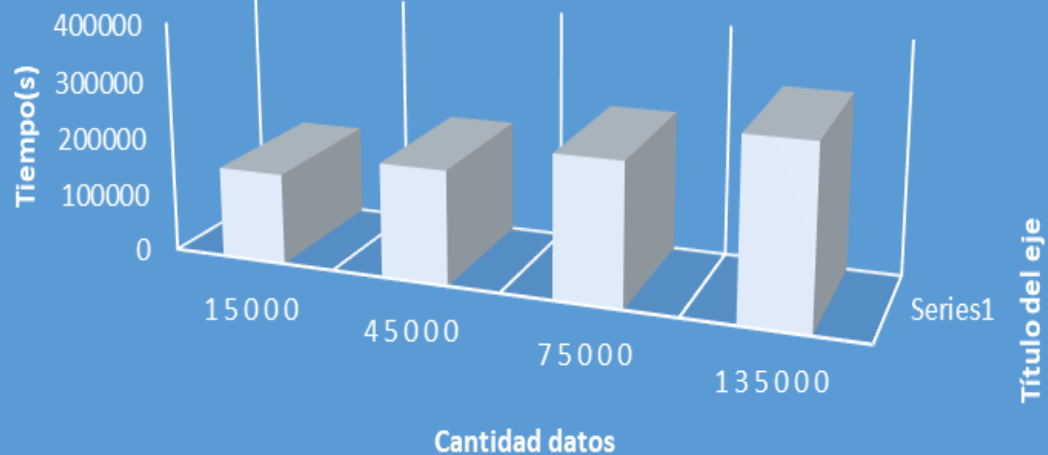
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.



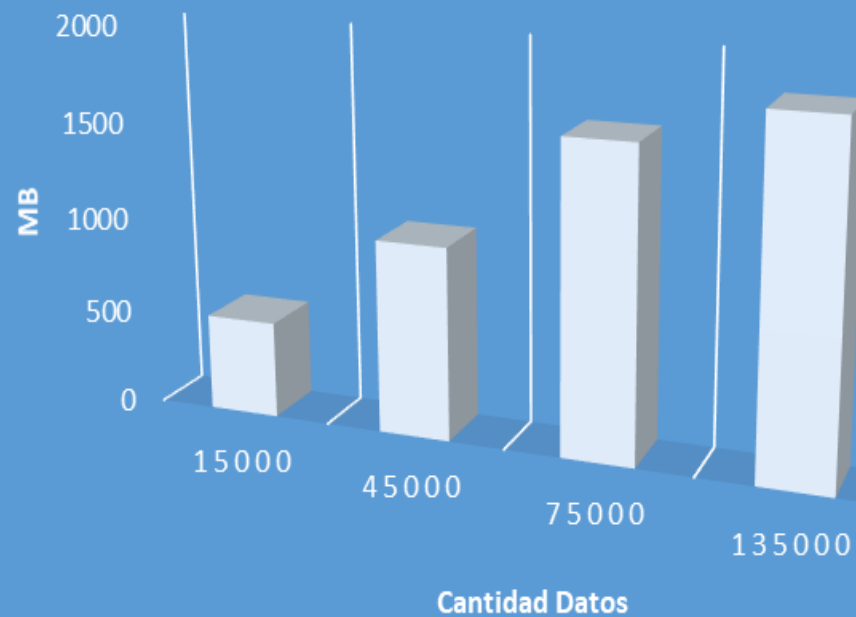
Consumo de tiempo y memoria



CONSUMO TIEMPO



CONSUMO MEMORIA





Por su atención |