

Los Arboles que Predicen el Futuro



Presentación del Equipo



Brahyan
Rios



Jonathan
Julio



Miguel
Correa



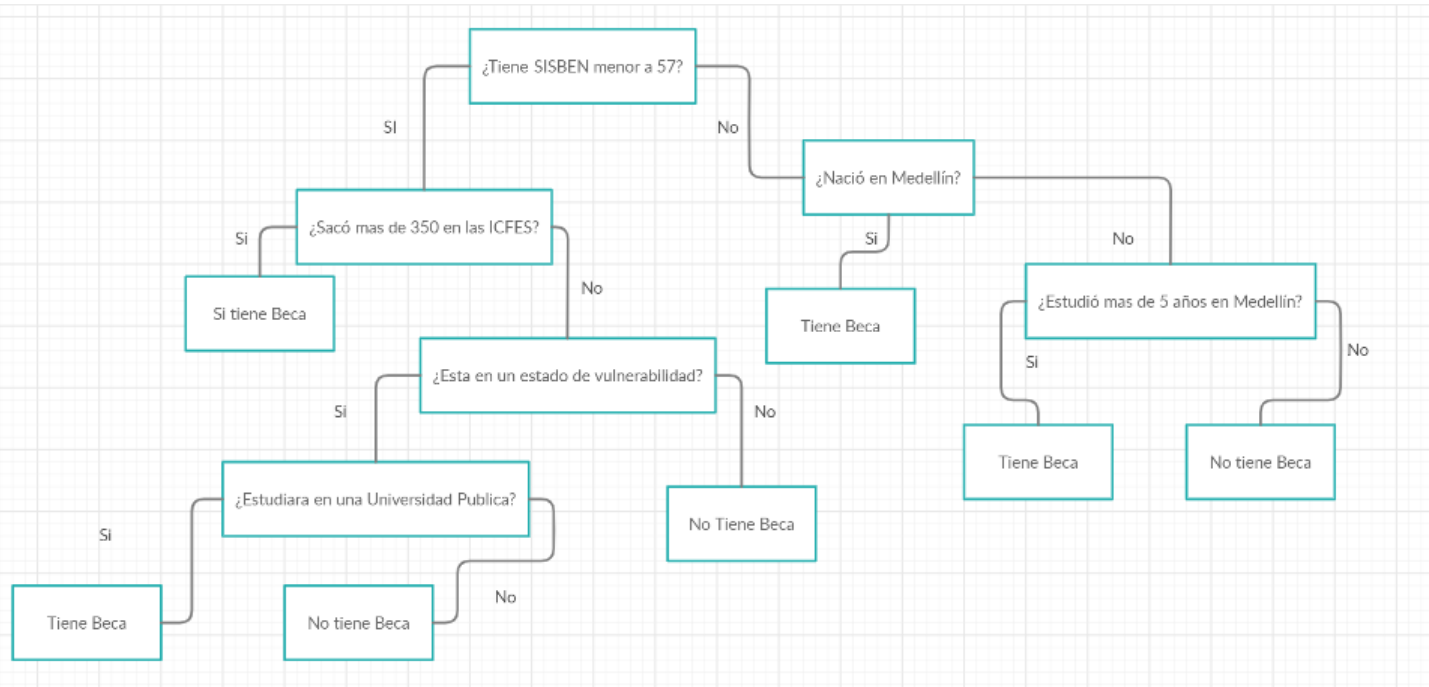
Mauricio
Toro



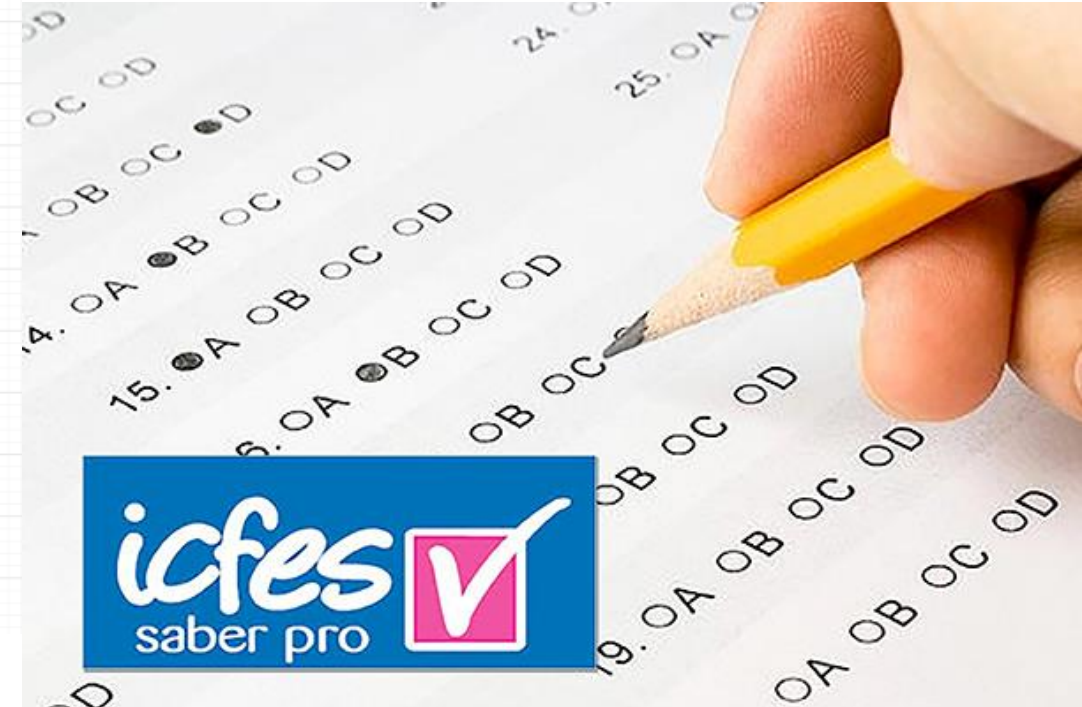
<http://github.com/berioss/proyecto/>

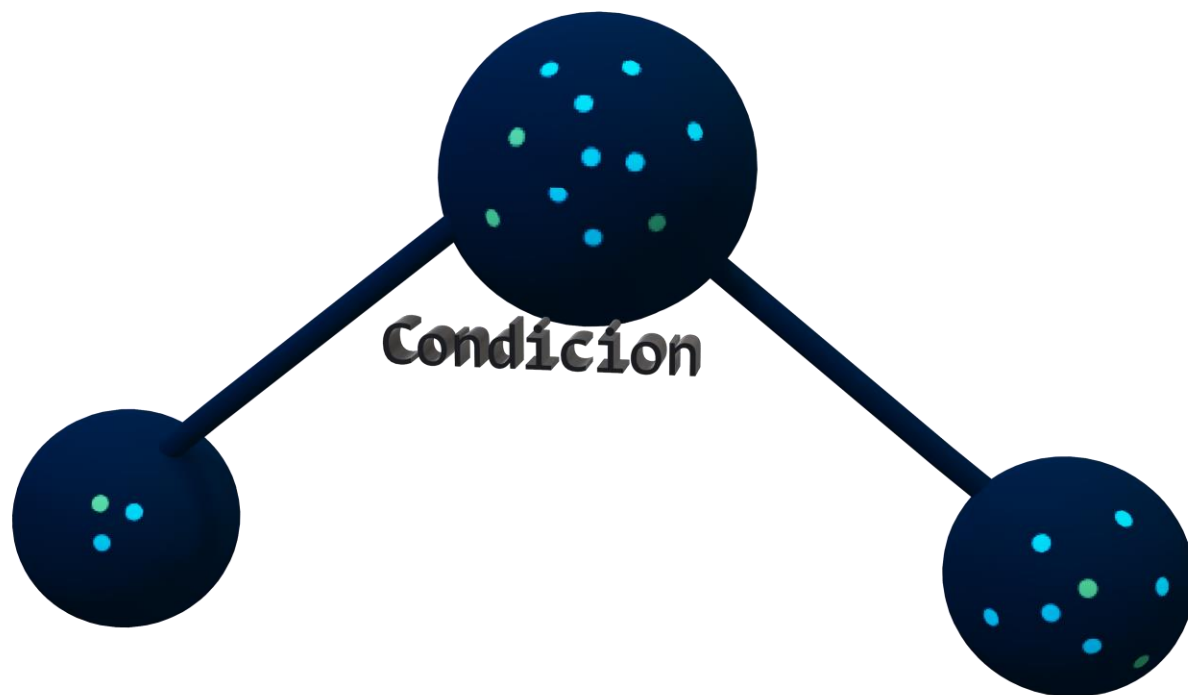


Diseño del Algoritmo

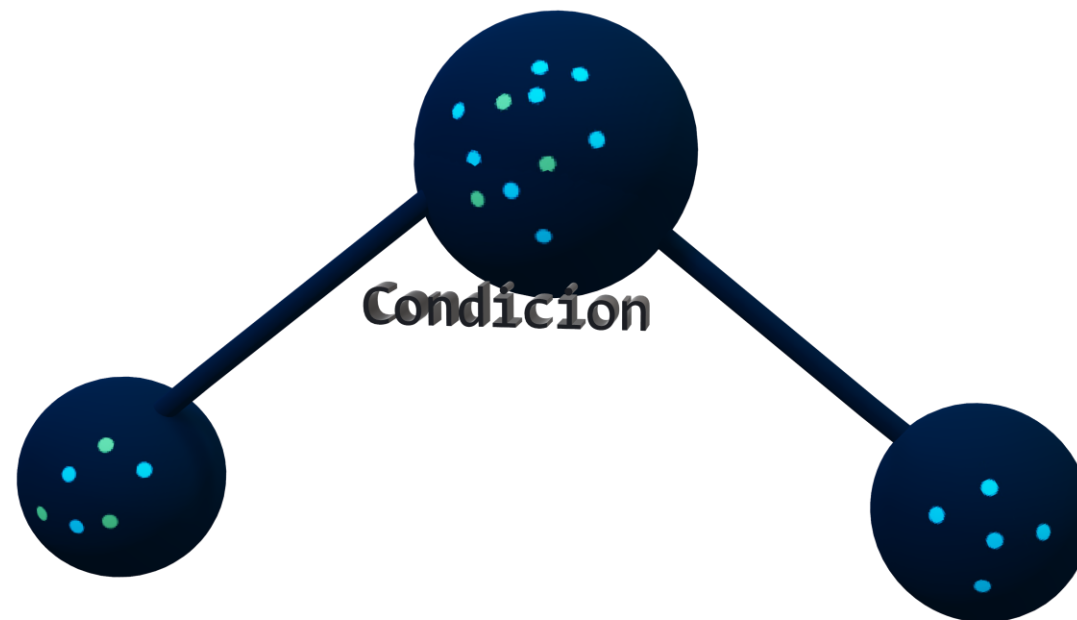


El entrenamiento consiste en que el algoritmo tomara todos los ejemplos y con estos se crearan los nodos lo cuales tendrán un orden dependiendo a la impureza de Gini, así las preguntas que menos impureza ponderada tengan serán las más aptas para el árbol de decisión, también se tiene que considerar un elemento ético del programador al escoger las preguntas.





Esta división es para saber si obtiene beca o no el estudiante, y está basada en la condición "Pisos de su casa ≥ 2 ". Para este caso, la impureza Gini de la izquierda es 1, la impureza Gini de la derecha es 0.53, y la impureza ponderada es de 0.66.



Esta división es para saber si obtiene beca o no el estudiante, y está basada en la condición "Estrato ≥ 4 ". Para este caso, la impureza Gini de la izquierda es 0.5, la impureza Gini de la derecha es 1, y la impureza ponderada es de 0.72.

Complejidad del Algoritmo

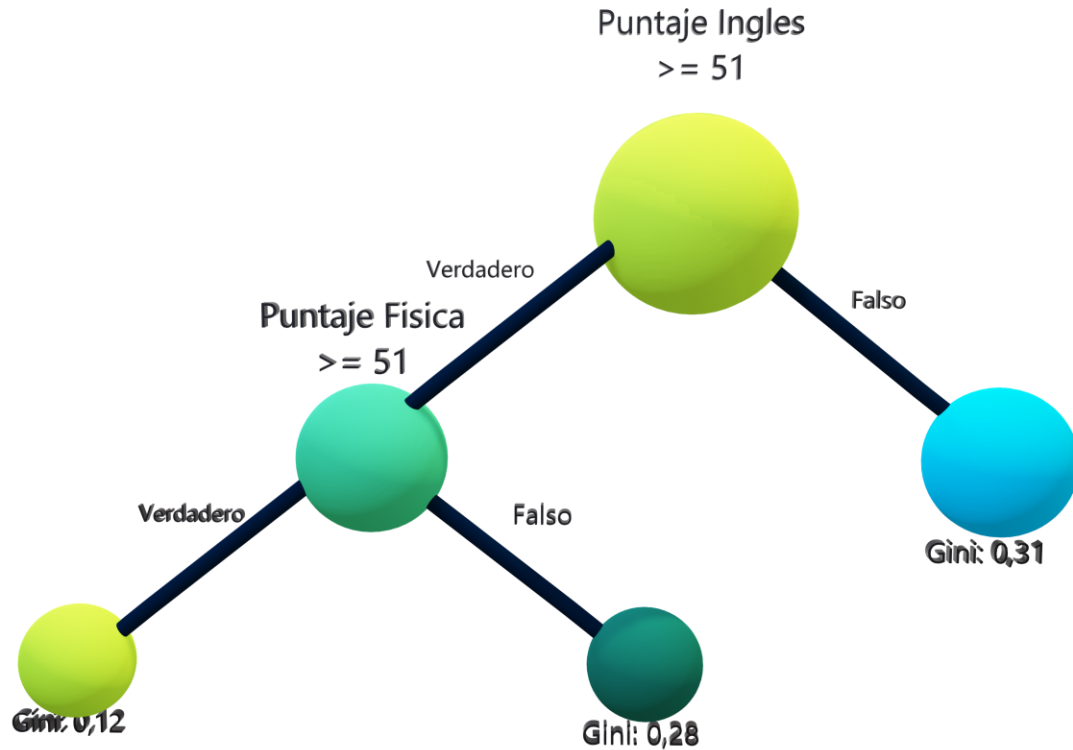


	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(M^2 \cdot N \cdot 2^P)$	$O(2^P)$
Validación del modelo	$O(P)$	$O(1)$

Nuestro algoritmo fue el CART, ya que nos interesa la forma en la que se divide el árbol binario, usando la impureza de Gini. En la complejidad, M se refiere a el número de filas, N el número de columnas, y P es la altura del árbol



Modelo de Árbol de Decisión

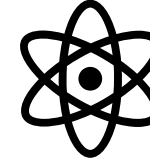


Un árbol de decisión para predecir el resultado del Saber Pro usando los resultados del Saber 11. Se usa la impuerza de Gini, obtenido al preguntar sobre los datos de los estudiantes para ir dividiendo a los estudiantes en dos lados, siendo el derecho donde los estudiantes no cumplen con la condicion, y el izquierdo si

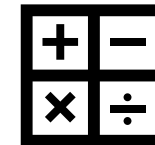
Características Más Relevantes



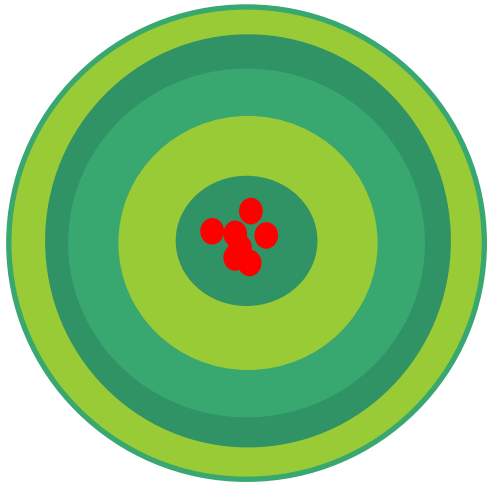
Puntaje Inglés



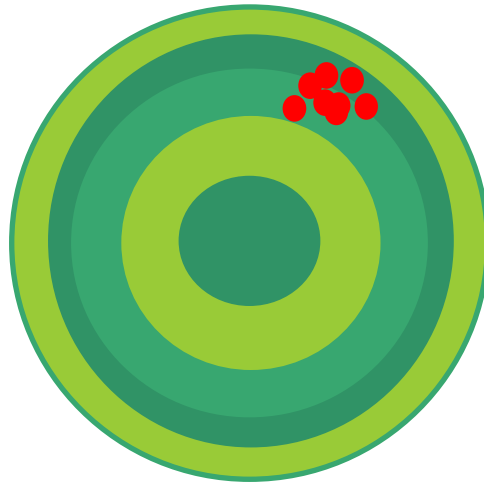
Puntaje Fisica



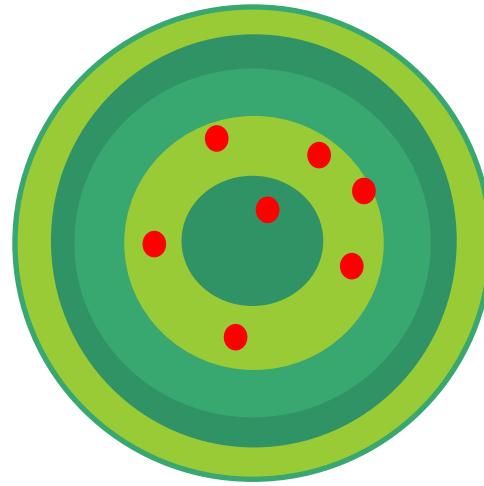
Puntaje Matematicas



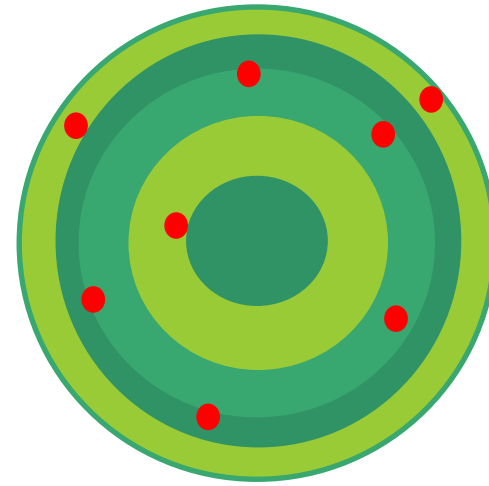
Alta exactitude
y alta precision



Baja exactitude
y alta precision



Alta exactitude
y baja precision



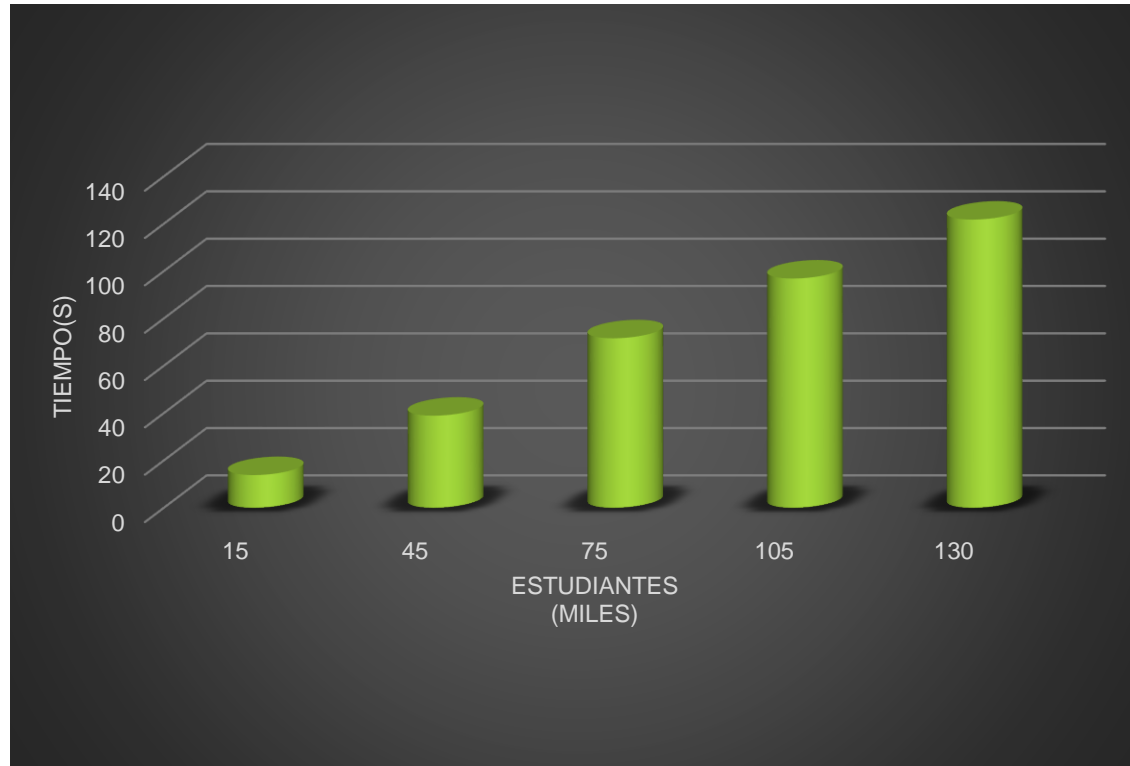
Baja exactitude
y baja precision

	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.81	0.72
Precisión	0.86	0.75
Sensibilidad	0.85	0.70

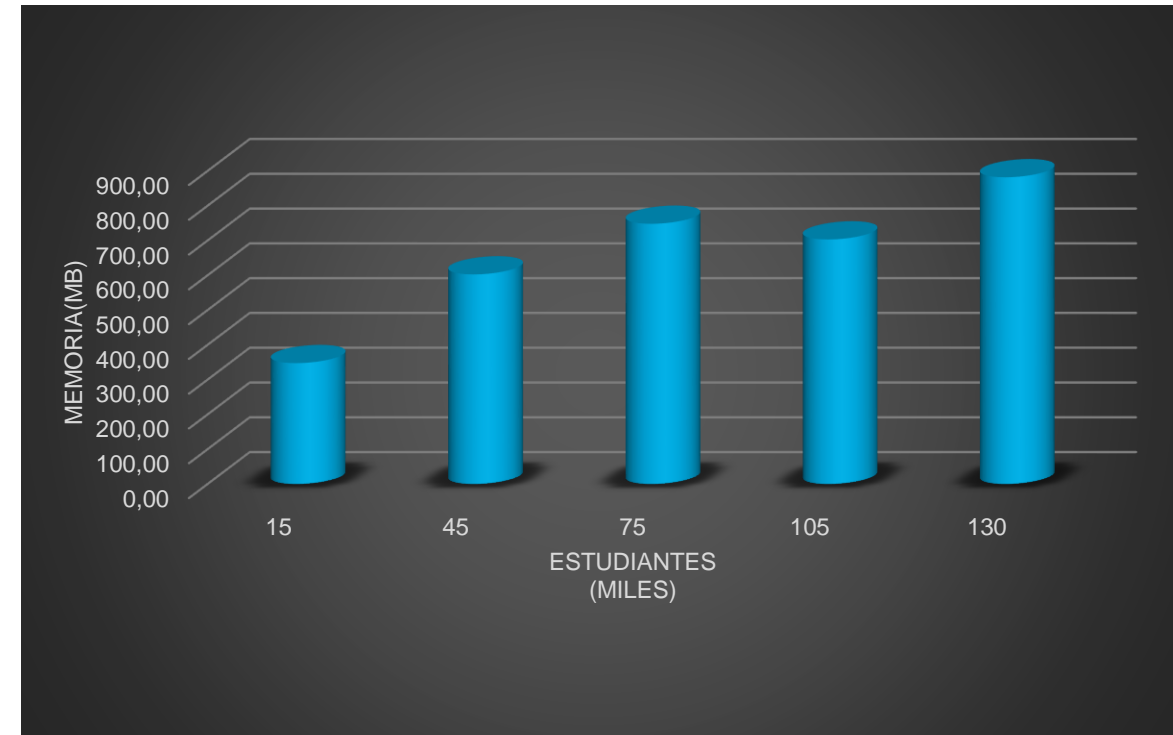
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria



¡GRACIAS POR
SU ATENCION!