

ÁRBOLES DE DECISION EN PRUEBAS DE ESTADO

Presentación del Equipo



Juan José
Zuluaga



Juan José
Wilches



Miguel
Correa



Mauricio
Toro



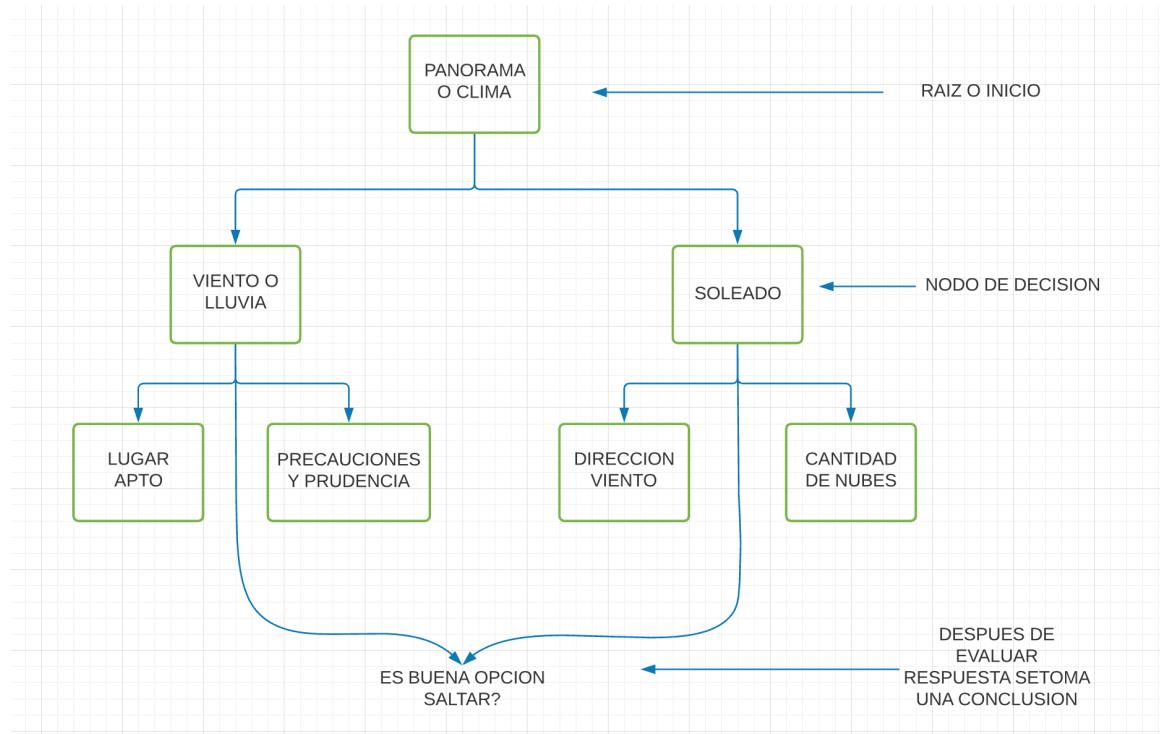
<http://github.com/>

JJZULUAGAB

/proyecto/

UNIVERSIDAD
EAFIT[®]

Diseño del Algoritmo



Este algoritmo fue utilizado para construir un árbol binario de decisión para definir que tan prudente y bueno puede ser un salto en paracaidas dependiendo de el clima, allí se evaluan las condiciones desde lo bueno hasta lo malo y se toma una decision.

IMPUREZA DE GINI

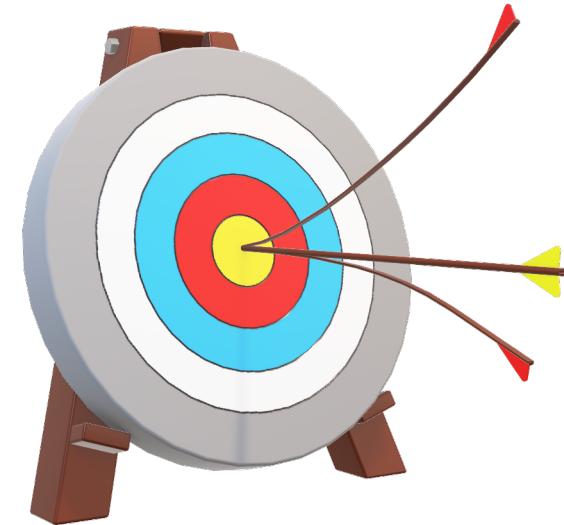


IMPURO, O NO FILTRADO



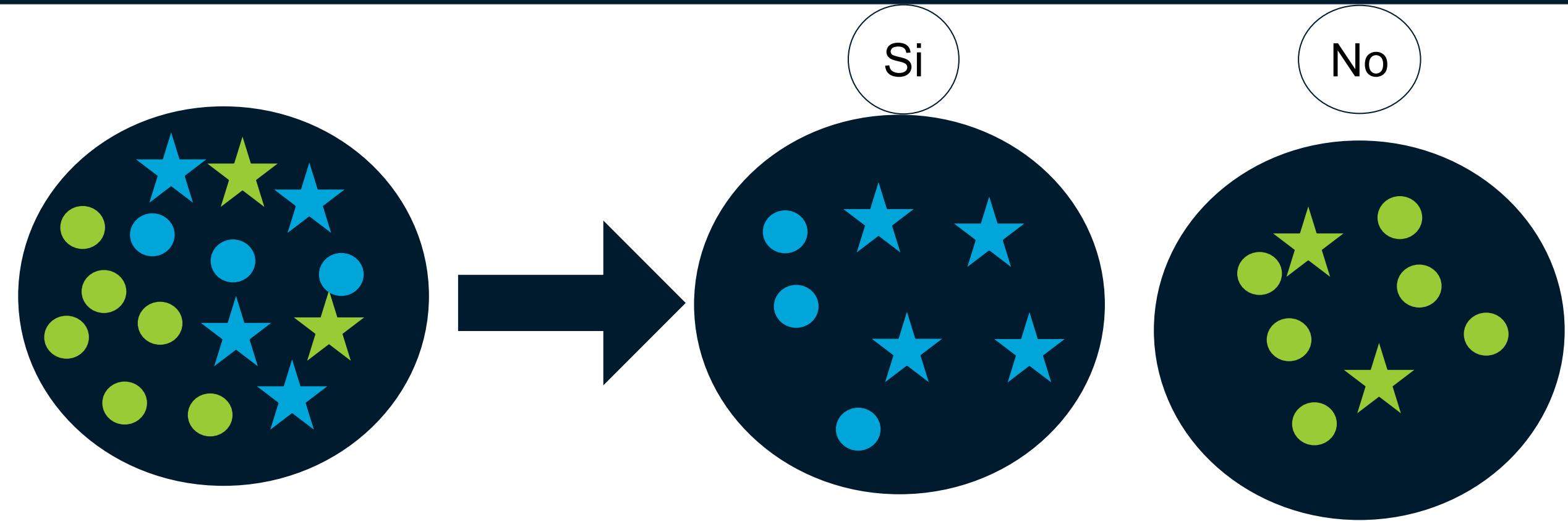
MAS PURO O FILTRADO

Como metrica de evaluacion, utilizando Gini, se hace una filtracion de la importancia de datos, y para saber que efectividad o exactitud tendran estos datos al final del proceso o predicion del resultado.

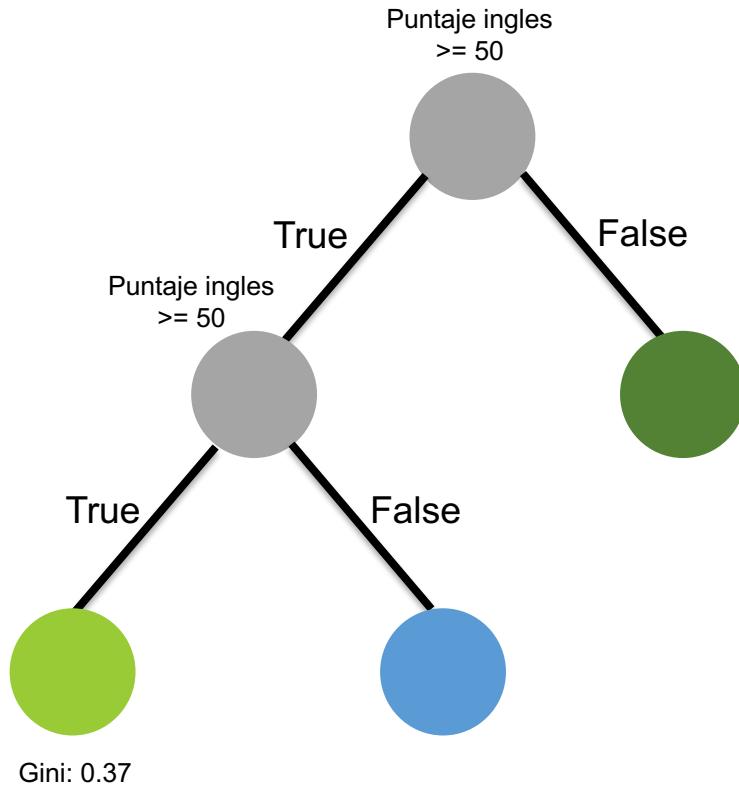


EXACTITUD Y PRECISION

División de un nodo



Este árbol busca separar la estrellas de los círculos, para la pregunta “es azul” tenemos una impureza Gini en el nodo Si de 0,48 y en el nodo No hay tenemos una impureza Gini de 0,375.



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

Características Más Relevantes



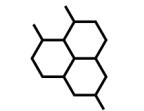
Ciencias Sociales



Inglés



Biología



Química



Física



Desempeño inglés



Filosofía

Complejidad del Algoritmo



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

Complejidad en tiempo y memoria del algoritmo, (CART), siendo n los nodos del arbol y m, las veces que lo utiliza, este Proyecto esta orientado a una predicción de resultados pero siempre buscando tener un equilibrio especialmente con la memoria y espacio que ocupa.

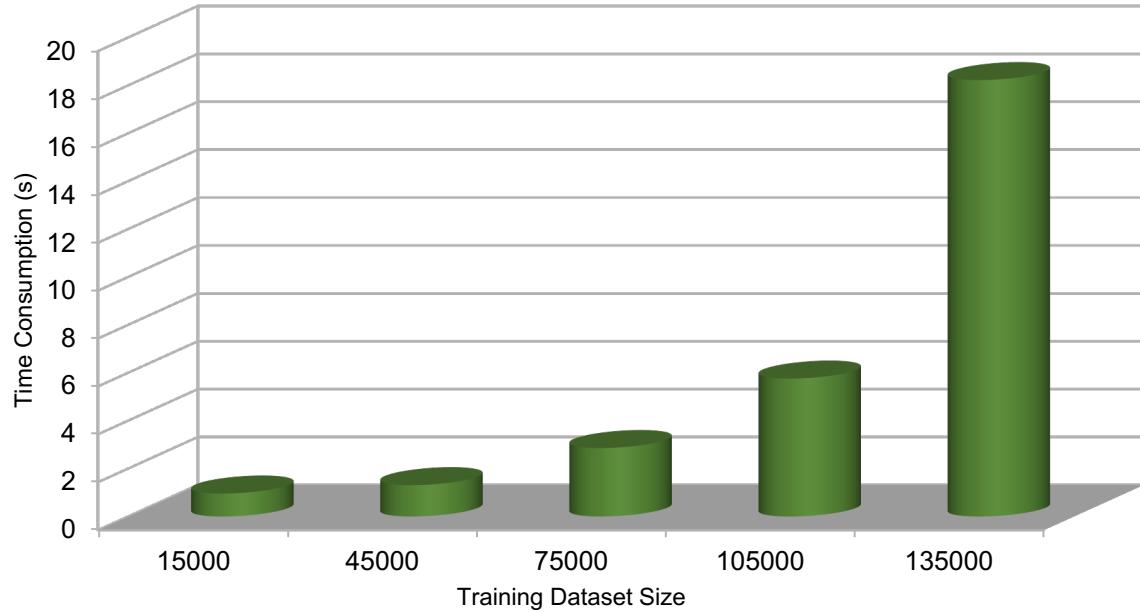


	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.76	0.76
Precisión	0.81	0.8
Sensibilidad	0.7	0.7

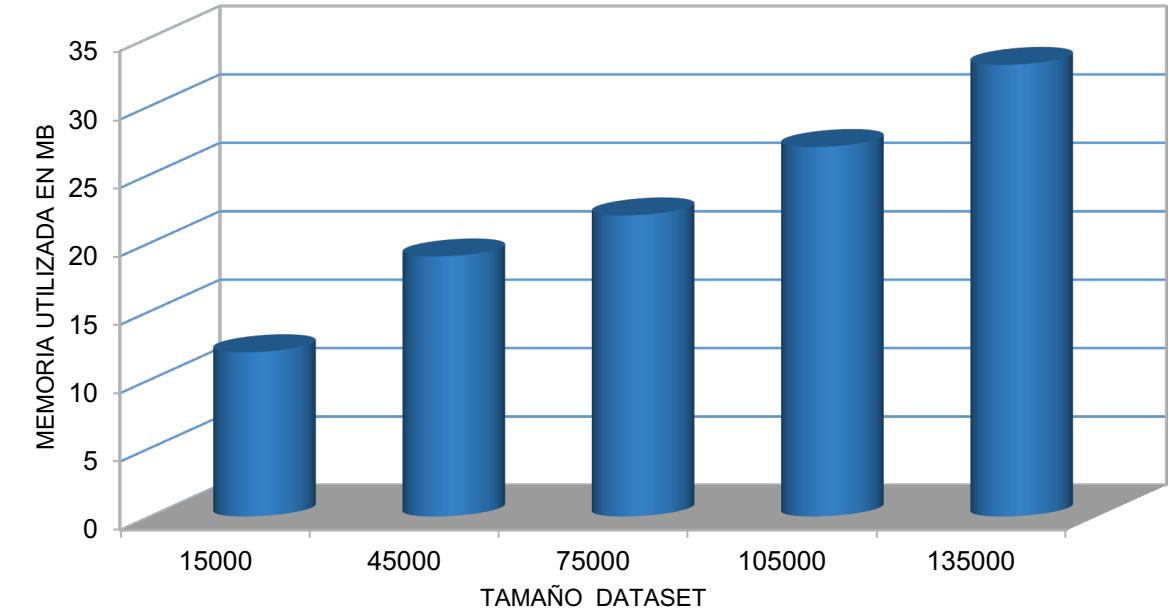
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes no aciertan 10.773.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria

Result.



Muchas
gracias por su
atencion