

# Aplicaciones con árboles de decisión y bosques aleatorios en la predicción del éxito académico en pruebas Saber Pro



# Presentación del Equipo



Jacobo Rave  
Londoño



Juan Sebastian  
Guerra

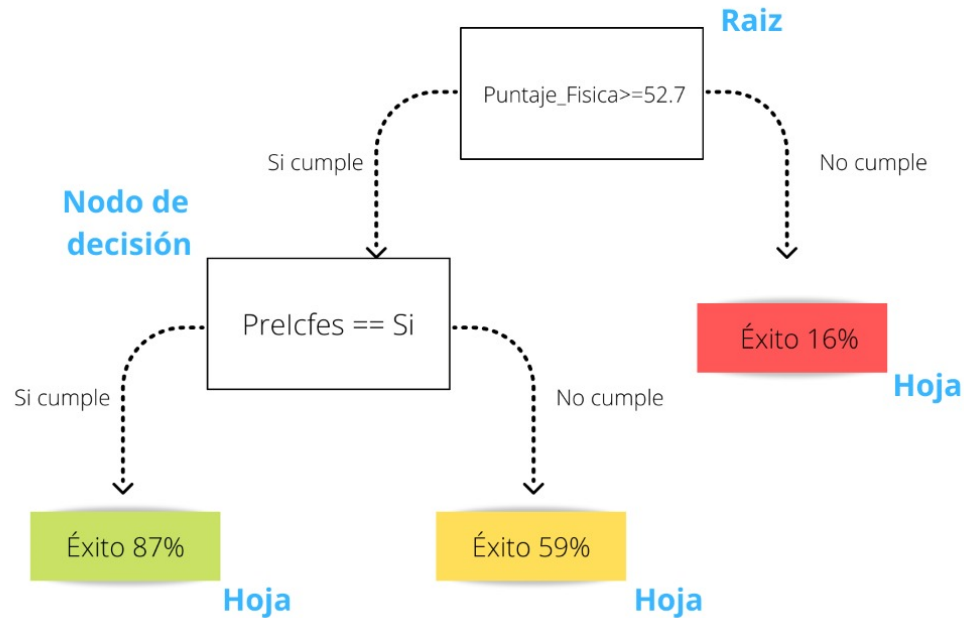


Miguel  
Correa



Mauricio  
Toro

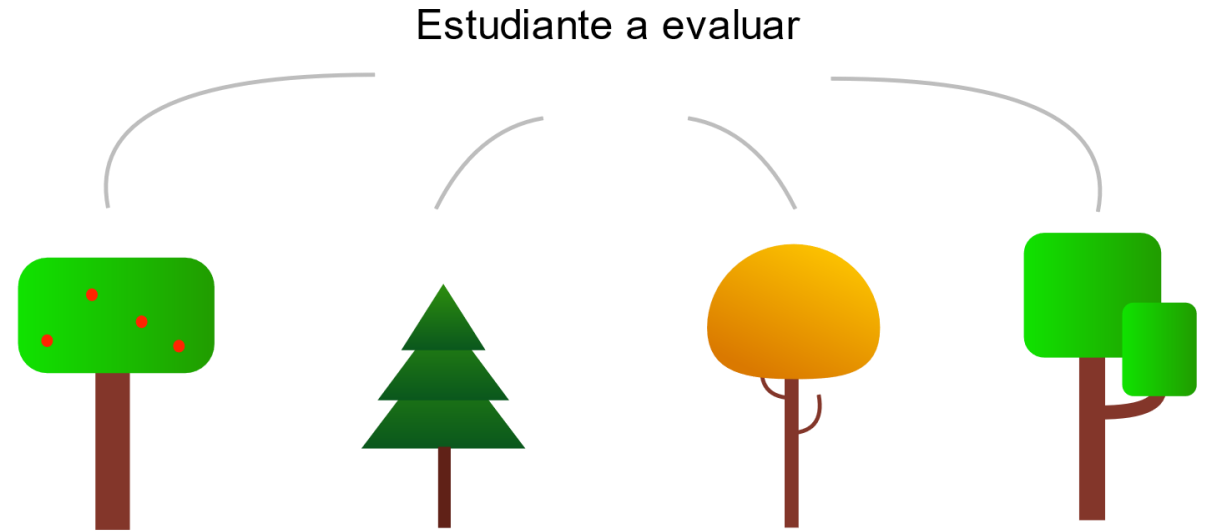
## Diagrama de árbol CART



Índice de Gini ponderado

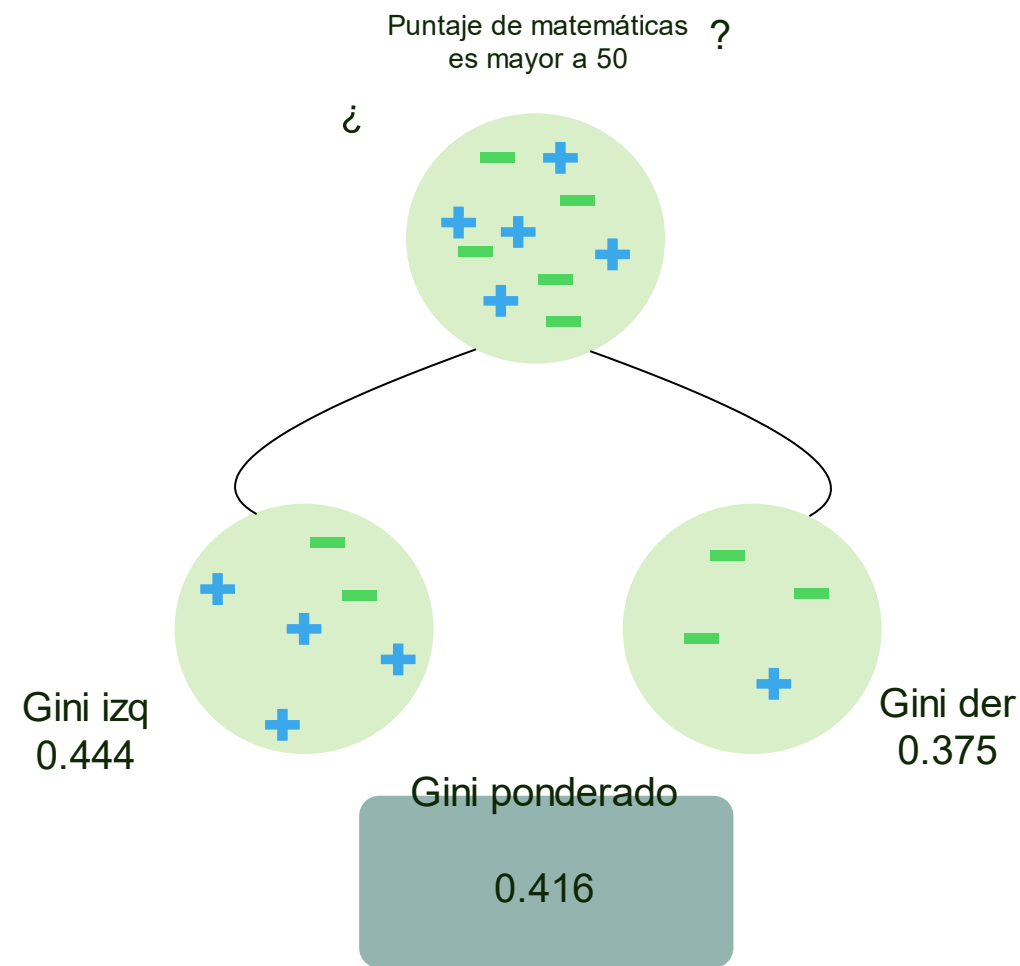
Este algoritmo predice el éxito académico a partir de atributos esenciales, por ejemplo, el puntaje de física y la presentación de preicfes dan un indicio de un posible éxito en el momento de presentar la prueba, si se cumple la raíz y un nodo de decisión se puede predecir el éxito del estudiante según los porcentajes.

La implementación de árboles busca construir arboles con muestras de los datos de entrenamiento, lo que hace que la construcción de cada árbol sea mas rápida. Después del entrenamiento, comprobamos una entrada por cada uno de los árboles del bosque y se llega a una concesión respecto a su éxito.



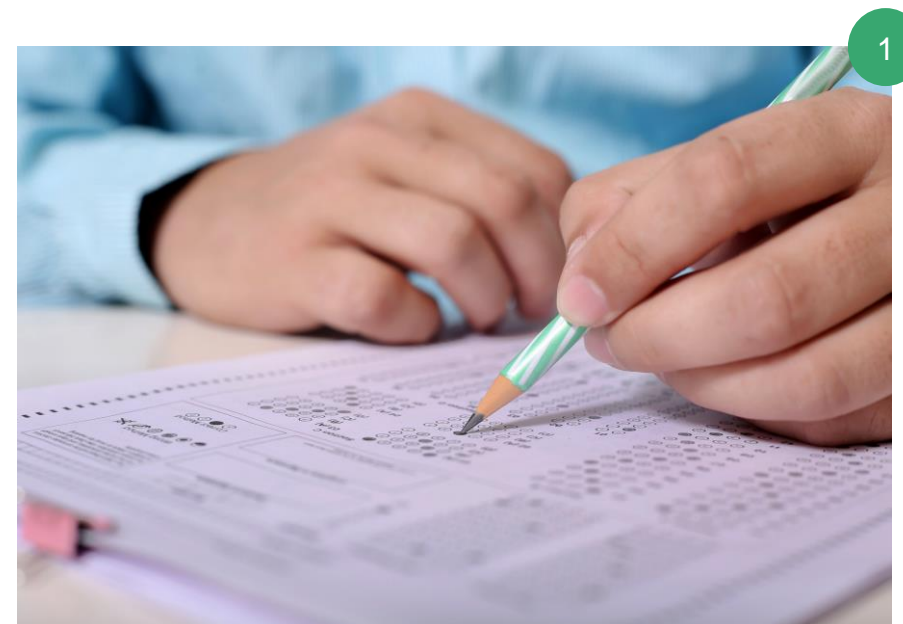


Usando el índice de Gini, la división esta dada según el nivel de impureza que nos brinde cada condición, siendo la menor la mas propicia para hacer la división. Se usa listas para evitar el exceso en consumo de memoria.



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

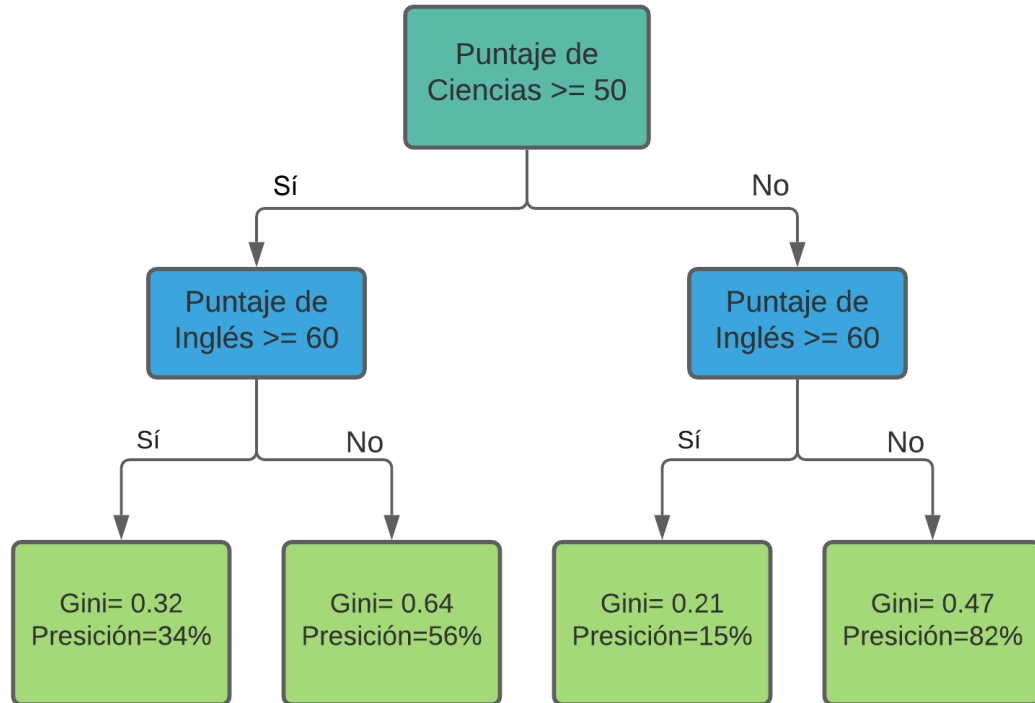
Para resolver el problema empleamos el algoritmo CART y bosques aleatorios. Al calcular la complejidad en tiempo y memoria del algoritmo se obtienen los datos de la tabla donde N es la cantidad de datos y M la cantidad de columnas de interés.



Problema a resolver

Predecir el éxito académico en las pruebas Saber Pro del estudiante. Para esto se toman variables que facilitan tal predicción con resultados pasados en pruebas Saber.

# Modelo de Árbol de Decisión



Un árbol de decisión para predecir el resultado del Saber Pro usando los resultados del Saber 11.

## Características Más Relevantes



Puntaje Español

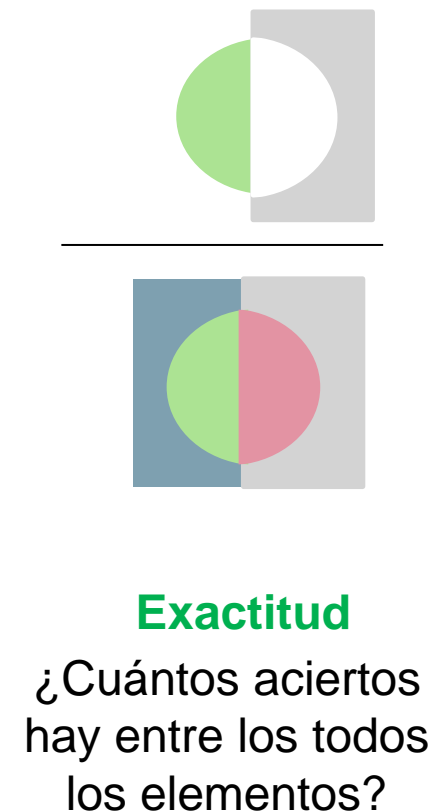
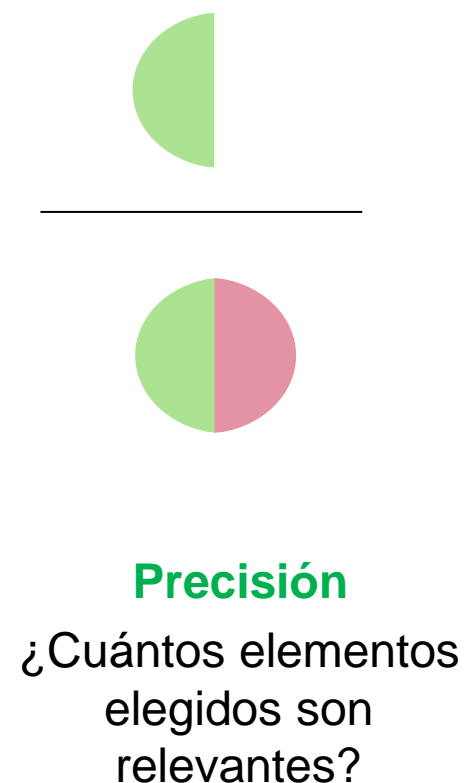
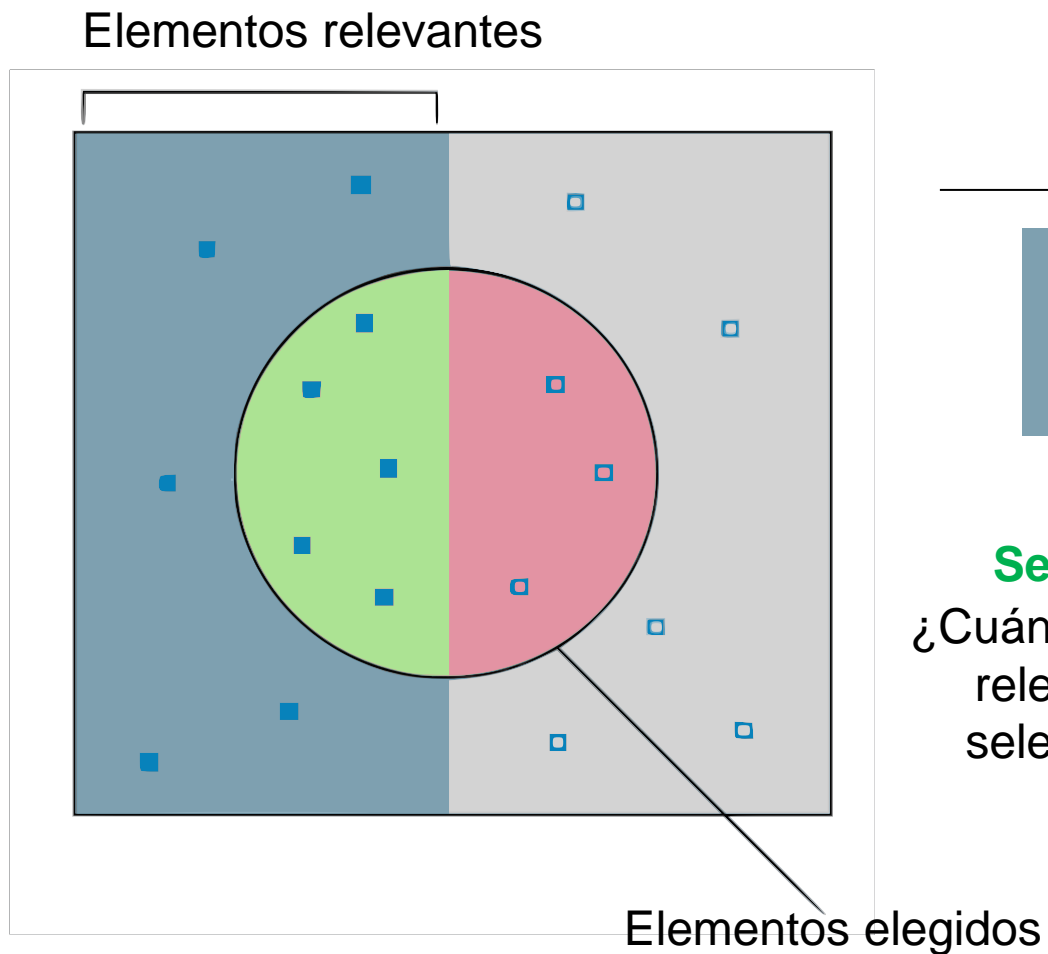


Puntaje Inglés



Puntaje Ciencias

# Métricas de Evaluación





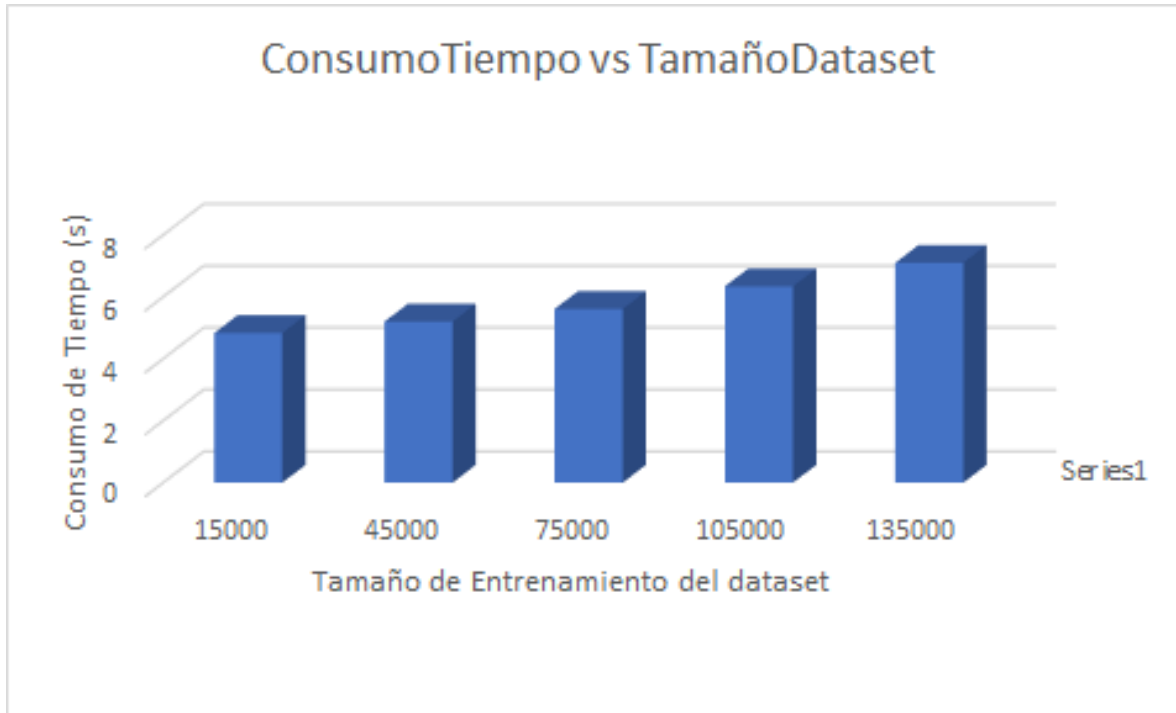
	Conjunto de entrenamiento	Conjunto de validación
Exactitud	0.74	0.73
Precisión	0,75	0.74
Sensibilidad	0.7	0.7

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. El bosque aleatorio probó ser útil para la generalización de una predicción. Sin embargo, la precisión que se logra se mantiene en un nivel estándar.

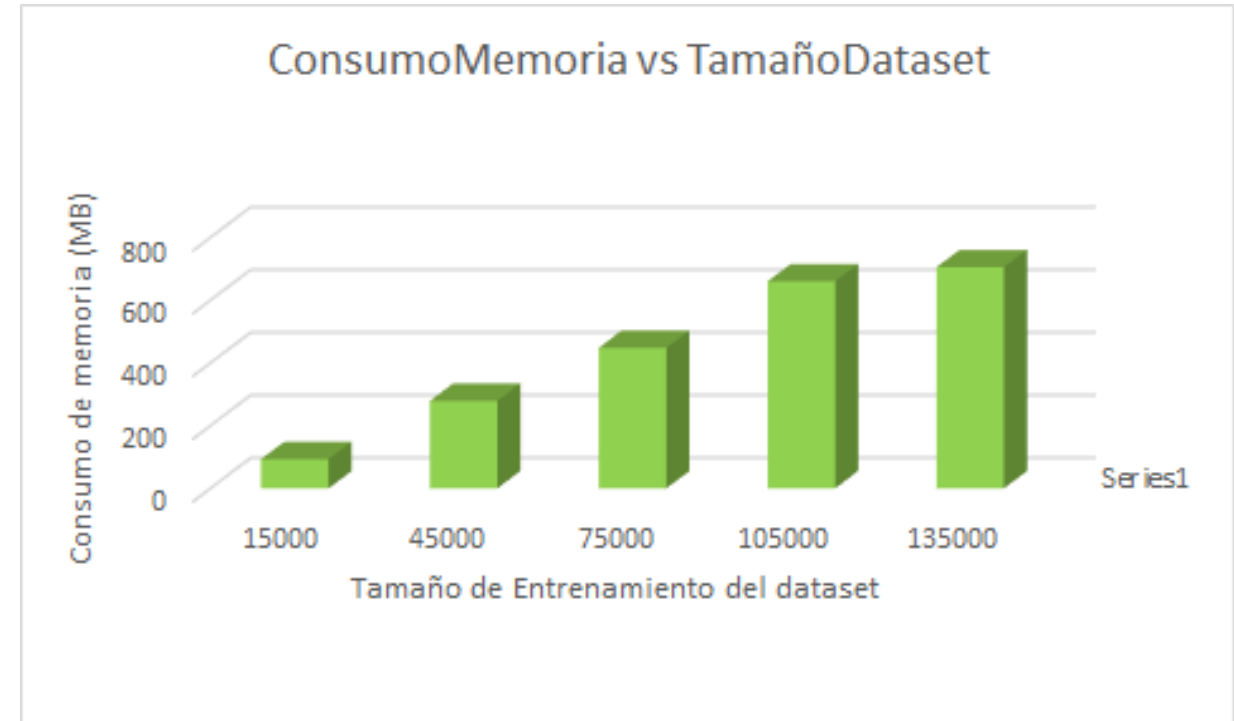


2

# Consumo de tiempo y memoria



Consumo de tiempo\*



Consumo de memoria\*

\*Consumo para la creación un bosque con mil árboles



## Aplicaciones con árboles de decisión y bosques aleatorios en la predicción del éxito académico en Pruebas Saber Pro

320.9KB

Make Private

Public

0



Contributors: [Juan Sebastián Guerra](#), [Jacobó Rave](#), [Mauricio Toro](#), [MIGUEL ANGEL CORREA MANRIQUE](#)

Identifier: DOI 10.17605/OSF.IO/PZXWQ

Category:  Project

Description: El siguiente texto trata sobre la formulación de un algoritmo para predecir el éxito académico en los resultados de las pruebas saber Pro en la educación superior, teniendo en cuenta variables de gran peso, como lo pueden ser resultados de exámenes anteriores, así como carencias tecnológicas que afectan parcialmente los resultados y su éxito, de manera que la información obtenida se recopile para generar posibles soluciones a futuro.

*Juan S Guerra, Jacobo Rave, Mauricio Toro, and Miguel A Correa.  
2020. Aplicaciones con árboles de decisión y bosques aleatorios  
en la predicción del éxito académico en Pruebas Saber Pro.  
DOI:<https://doi.org/10.17605/OSF.IO/PZXWQ>*



¡GRACIAS!