

PREDICCIÓN
DE RESULTADOS
EN PRUEBAS
SABER PRO
USANDO
DATOS
ACADEMICOS
SOBRE EL
SABER 11



Presentación del Equipo



Miguel
Manzur



José Aníbal
Pinto



Jorge Juan
Araujo



Miguel
Correa



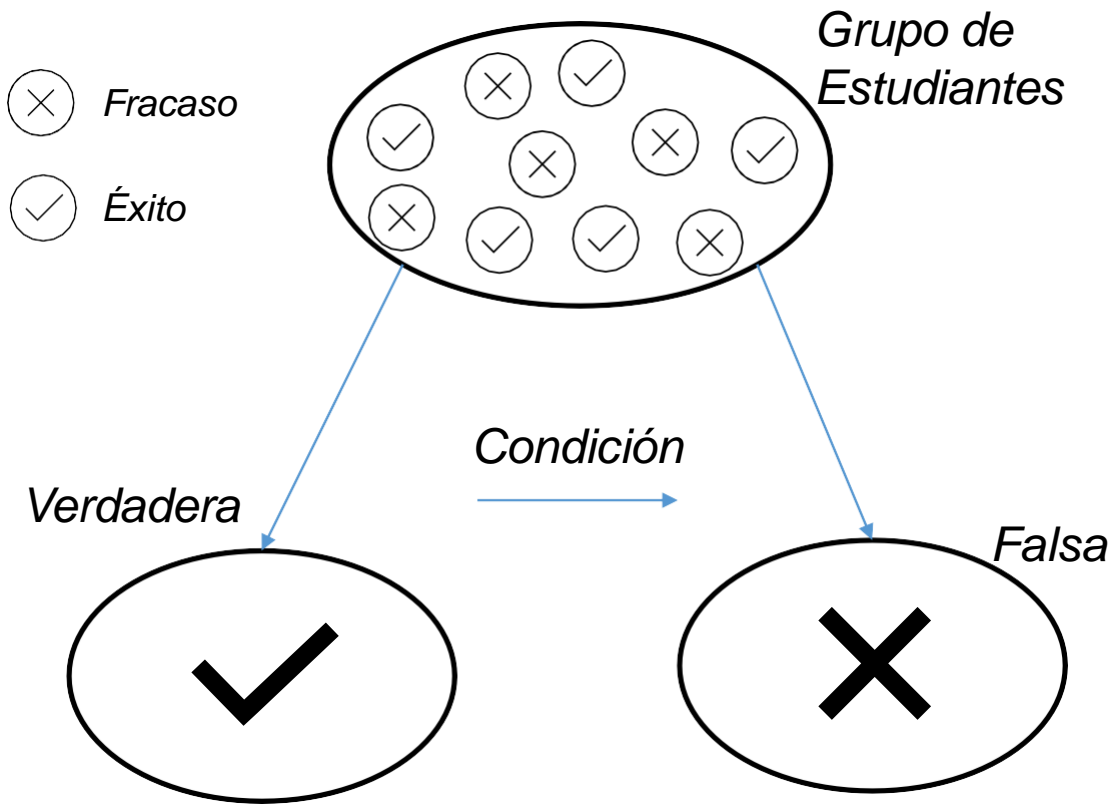
Mauricio
Toro



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



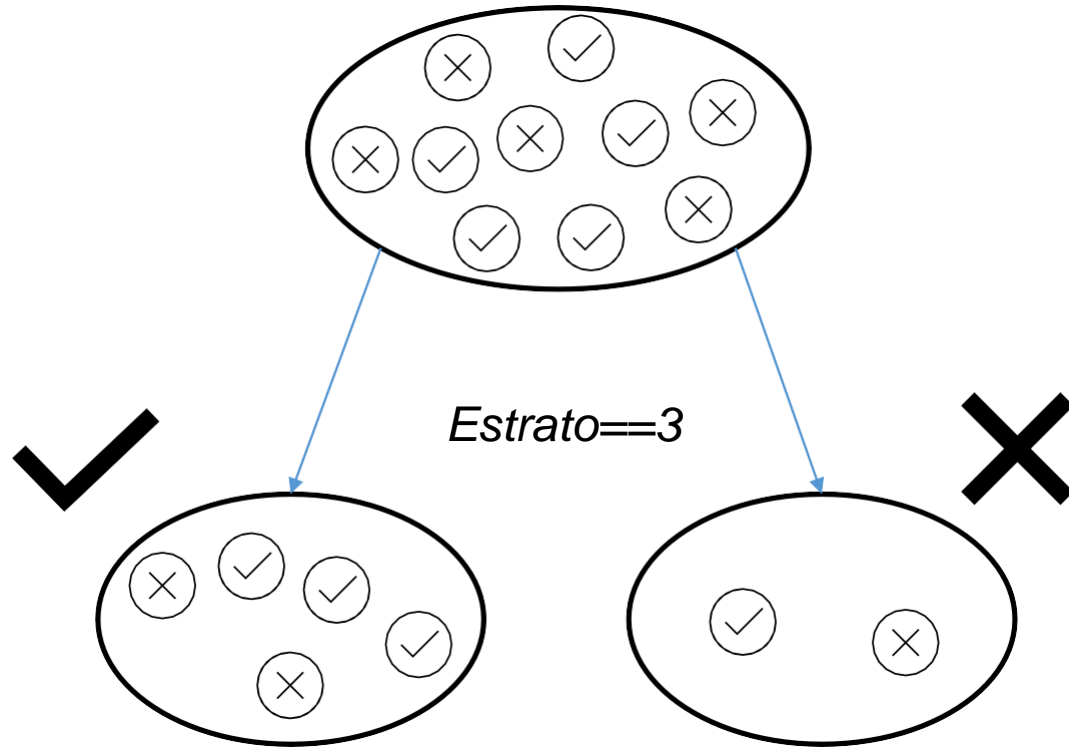
Diseño del Algoritmo



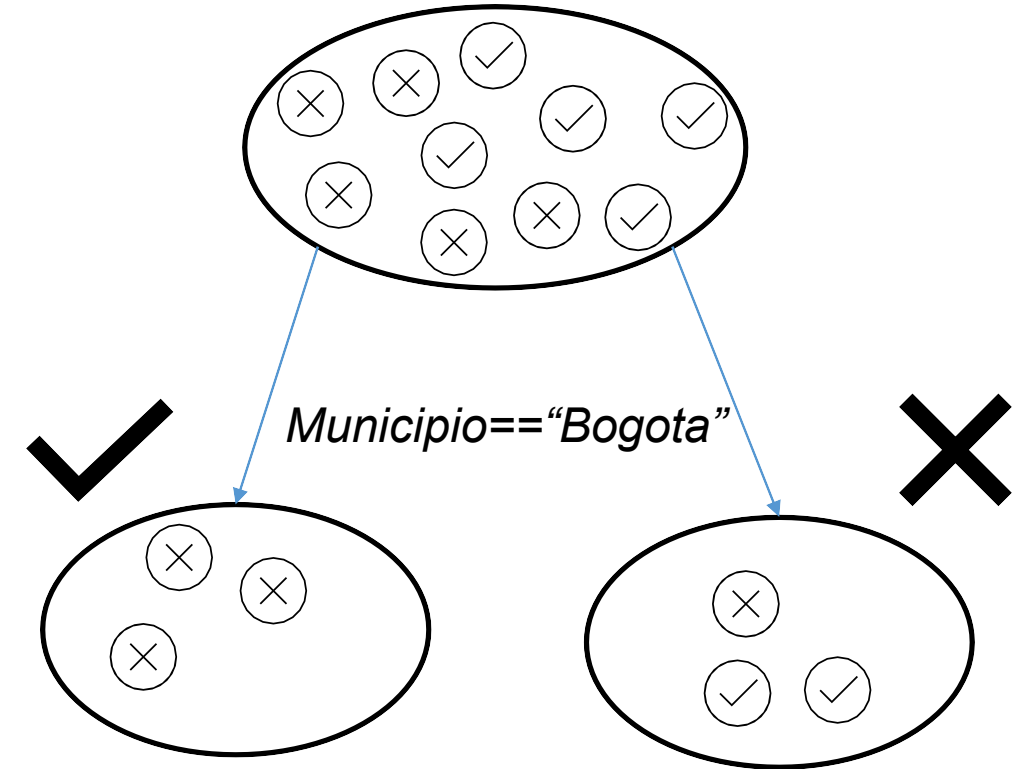
Algoritmo para construir un árbol binario de decisión usando CART. En este ejemplo, mostramos un modelo para predecir si un estudiante ha tenido buenos resultados (Éxito o Fracaso) por lo que buscaremos las condiciones que tengan en común estos estudiantes



División de un nodo



Esta división está basada en la condición “Estrato==3.” Para este caso, la impureza Gini de la izquierda es 0.48, la impureza Gini de la derecha es 0.5 y la impureza ponderada es de 0.49.



Esta división está basada en la condición “Municipio==“Bogotá”.” Para este caso, la impureza Gini de la izquierda es 0, la impureza Gini de la derecha es 0.44 y la impureza ponderada es 0.22.

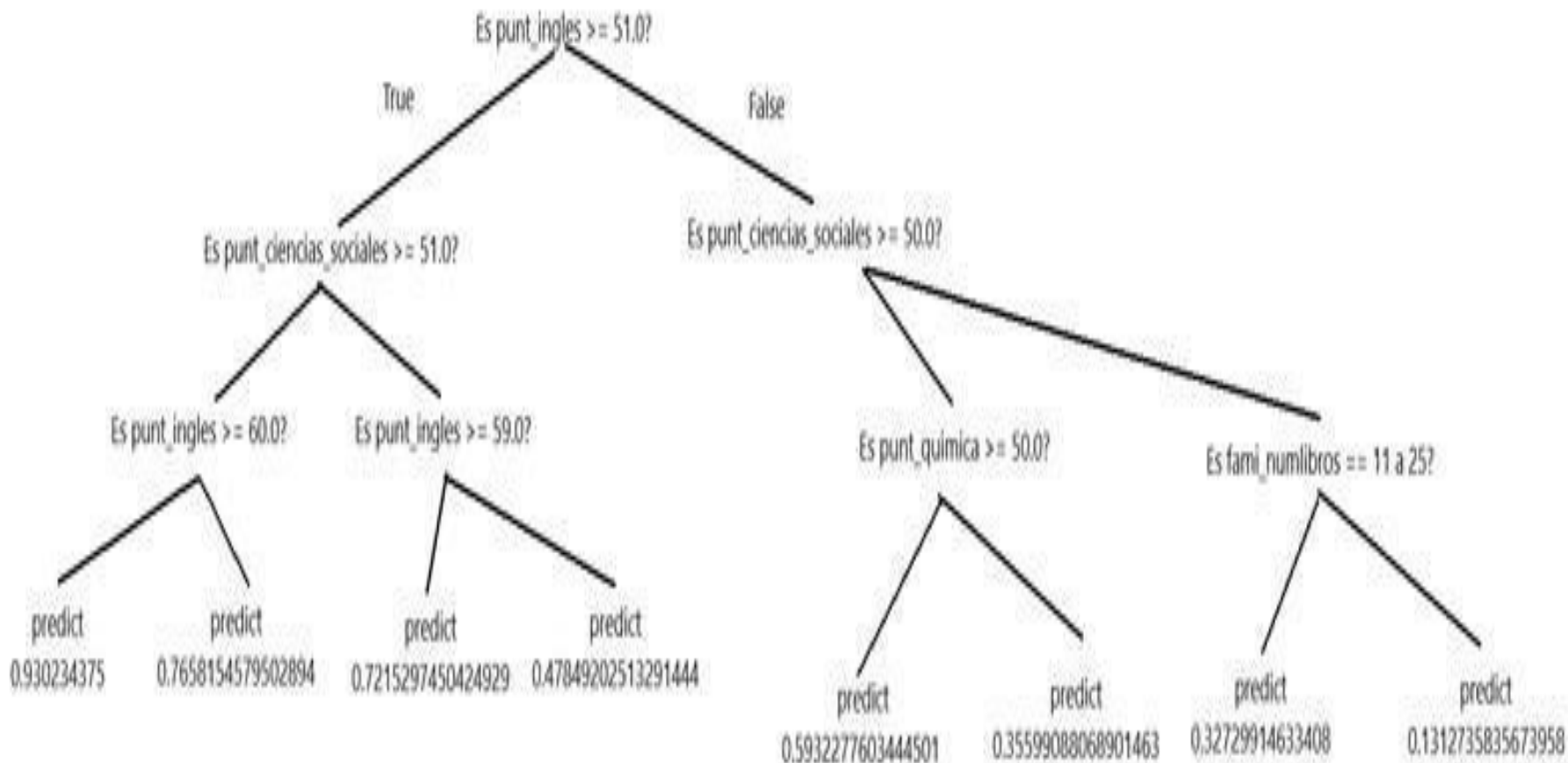
Complejidad del Algoritmo



	Complejidad en tiempo	Complejidad en memoria
Entrenamiento del modelo	$O(2^x * n * m)$	$O(n * m * x)$
Validación del modelo	$O(\log_2 n)$	$O(n)$

Complejidad en tiempo y memoria del algoritmo (En este semestre, una opción puede ser CART. Complejidad temporal de los algoritmos de entrenamiento y prueba. Donde x es el número de nodos en el árbol, n el número de estudiantes y m el número de las preguntas utilizadas.





Características Más Relevantes



Ingles



Ciencias Sociales

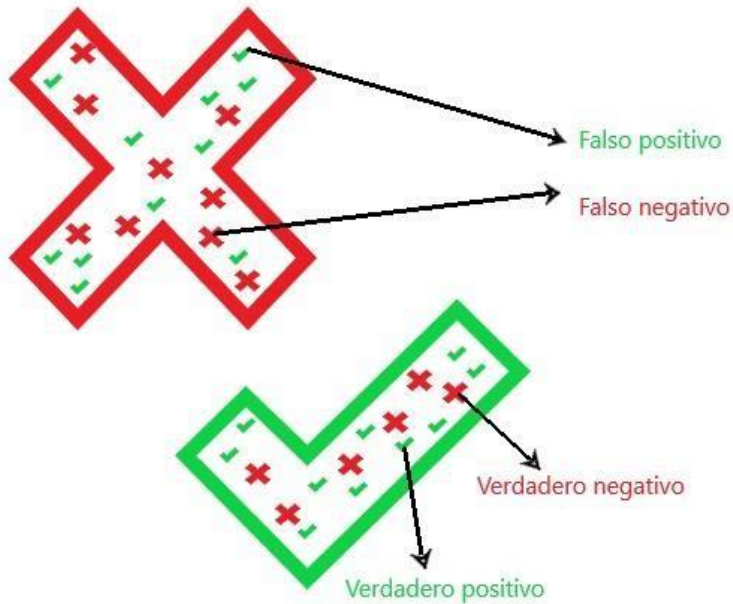


Quimica



de libros

Un árbol de decision para predecir el resultado del Saber Pro usando los resultados del Saber 11. Es un árbol de decisión binario donde los nodos son las preguntas, y cada pregunta divide el marco de datos en dos marcos de datos más pequeños que separan los objetos que cumplen la condición y los que no.



Precisión

$$\frac{\text{Verdadero } P}{(\text{Verdadero } P + \text{Falso } P)}$$

Sensibilidad

$$\frac{\text{Verdadero } P}{(\text{verdadero } P + \text{Falso } N)}$$

Exactitud

$$\frac{(\text{Verdadero } P + \text{Verdadero } N)}{(\text{Verdadero } P + \text{Verdadero } N + \text{Falso } N + \text{Falso } P)}$$

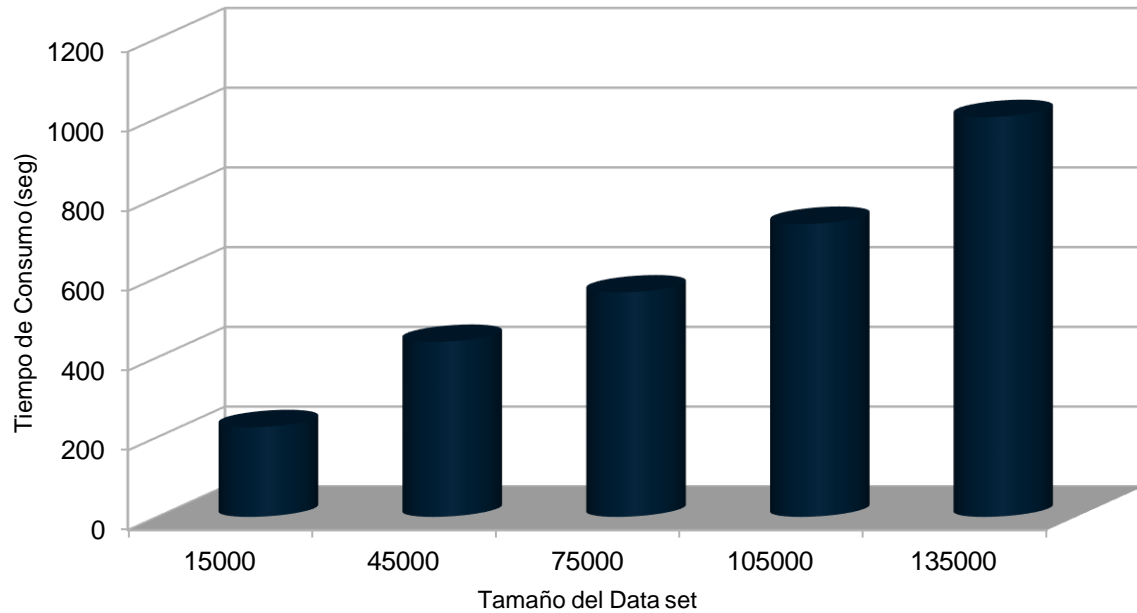


	Conjunto de entrenamiento	Conjunto de validación
Exactitud	80,5%	82%
Precisión	80,3%	80%
Sensibilidad	80,9%	84,6%

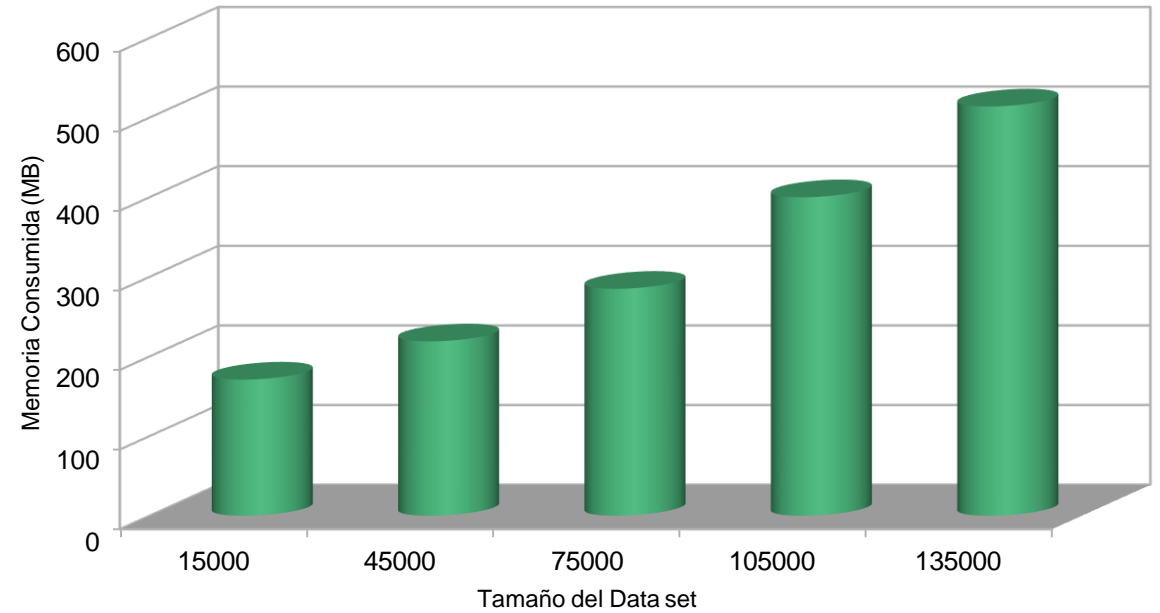
Métricas de evaluación obtenidas con el conjunto de datos de entrenamiento de 135,000 estudiantes y el conjunto de datos de validación de 45,000 estudiantes. Este fue el porcentaje que obtenimos según la proporción.



Consumo de tiempo y memoria



Consumo de tiempo



Consumo de memoria



J. Pinto, M. Manzur y J. Araujo.
PREDICCIÓN DE
RESULTADOS EN PRUEBAS
SABER PRO USANDO DATOS
ACADEMICOS SOBRE EL
SABER 11. OSF, Nov. 2020.
Available at: <https://osf.io/k6v2p/>

PREDICCIÓN DE RESULTADOS EN PRUE...

FilesWikiAnalyticsRegistrationsContributorsAdd-onsSettings

198.1KBMake PrivatePublicP 0...

PREDICCIÓN DE RESULTADOS EN PRUEBAS
SABER PRO USANDO DATOS ACADEMICOS
SOBRE EL SABER 11

Contributors: [Jorge Juan Araujo](#)

Date created: 2020-11-20 07:42 PM | Last Updated: 2020-11-20 07:52 PM

Create DOI

Category: Project

Description: Add a brief description to your project

License: Add a license

Wiki

Add important information, links, or images here to describe your project.

Files

Click on a storage provider or drag and drop to upload

Filteri

Name ^ v	Modified ^ v
PREDICCIÓN DE RESULTADOS EN PRUEBAS SABER PRO USANDO DA...	
OSF Storage (United States)	
Codigo Final Proyecto EDDYA.py	2020-11-20 07:50 PM
Proyecto entrega Final EDDYA.docx	2020-11-20 07:51 PM

Citation

Cite as:

[1] Williams, Laurie. "White-Box Testing" (PDF): 60-61, 69. Recuperado el 11 de octubre de 2020. [2]. Gao, J., Tsao, H.-S. J., & Wu, Y. (2003). Testing and quality assurance for component-based software. Boston, MA: Artech House. [3]. Garcia, J. y Sanchez, P. Extracción de Conocimiento para la Predicción y Análisis de los Resultados de la Prueba de Calidad de la Educación Superior en Colombia. Scielo Analytics. 2019. [4]. Ospina, H y Quintana, L. Predicción de accidentes viales en Cartagena, Colombia, con árboles de decisión y reglas de asociación. Universidad Javeriana. 2019. [5]. Timaran, R. Calderon, A. Hidalgo, A. Aplicación de los árboles de decisión en la identificación de patrones de lesiones fatales por causa externa en el municipio de Pasto, Colombia. ResearchGate. 2017. [6]. Camborda, M. Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la carrera de ingeniería civil de la universidad Continental. Universidad Nacional del Centro de Perú. 2014. [7] Nguyen. Comparative Study of C5.0 and CART algorithms. Retrieved from <http://mercury.webster.edu/aleshunassupport%20Materials/C4.5/Nguyen-Presentation%20Data%20mining.pdf>

Edit

UNIVERSIDAD
EAFIT[®]



GRACIAS POR SU TIEMPO