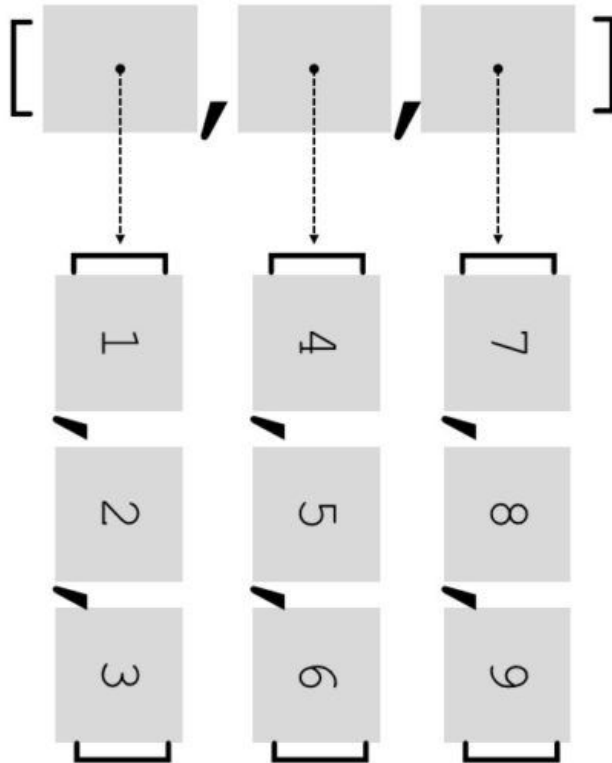


Algorithm for prediction on the results of the ICFES Saber Pro exam results

***Pedro Botero Aristizabal
Samuel Ceballos Posada
Medellín, June 2 / 2020***

CART decision tree algorithm

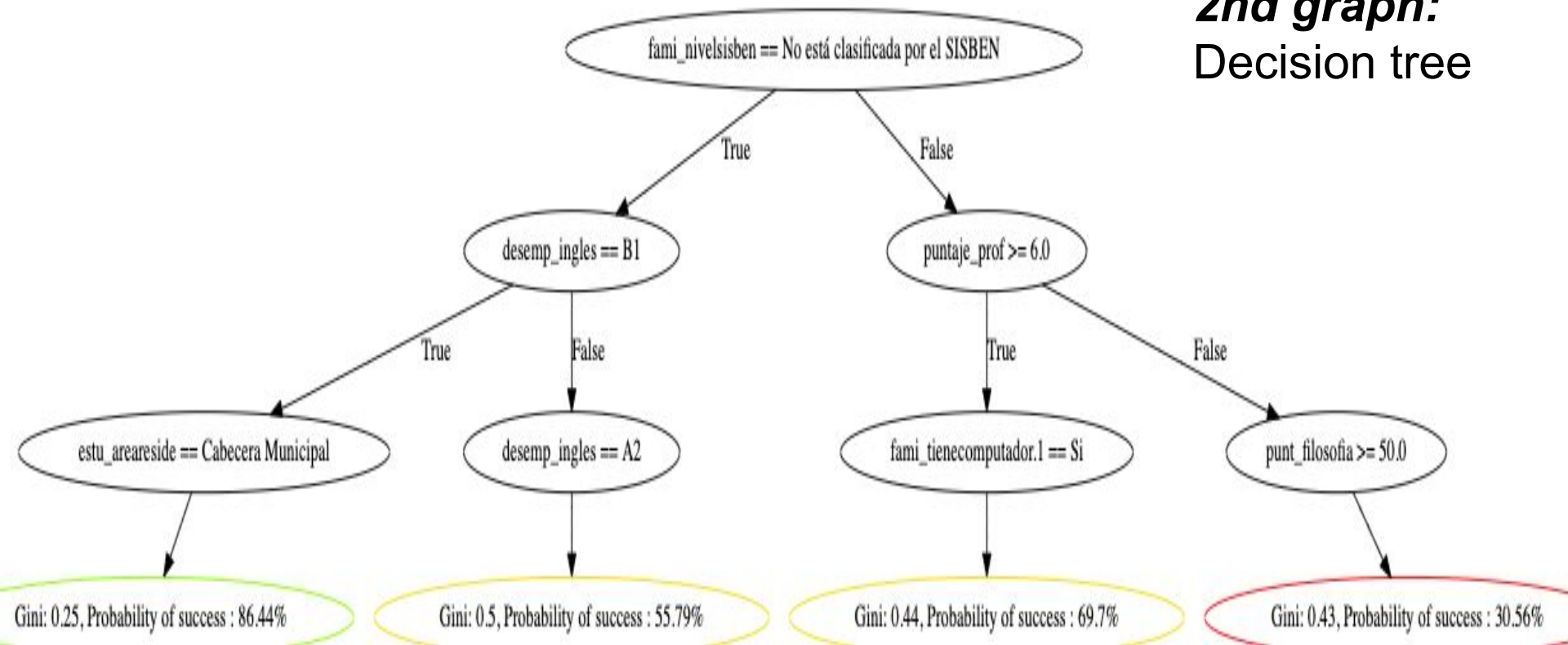


1st graph:

We used a list of lists to read and store the data.

CART decision tree algorithm

2nd graph:
Decision tree



Design Criteria of the Data Structure

The main elements of CART are:

1. Rules for splitting data at a node based on the value of one variable.
2. Stopping rules for deciding when a branch is terminal and can be split no more.
3. Finally, a prediction for the target variable in each terminal node.

Data Structure Operations

Function	Complexity	Function	Complexity
create_df	$O(n)$	information_gain	$O(n \times m)$
labelling	$O(n)$	best_option	$O(n^2 \times m)$
is_number	$O(1)$	create_question	$O(1)$
best_value	$O(n \times m)$	build_tree	$O(n^2 \times m)$
match	$O(1)$	print_tree	$O(n \times m)$
partition	$O(n)$	predict	$O(n)$
gini	$O(n)$	classify	$O(n \times m)$

Table 1: Time complexity for the data structure functions.

Total = $O(n^2 \times m)$

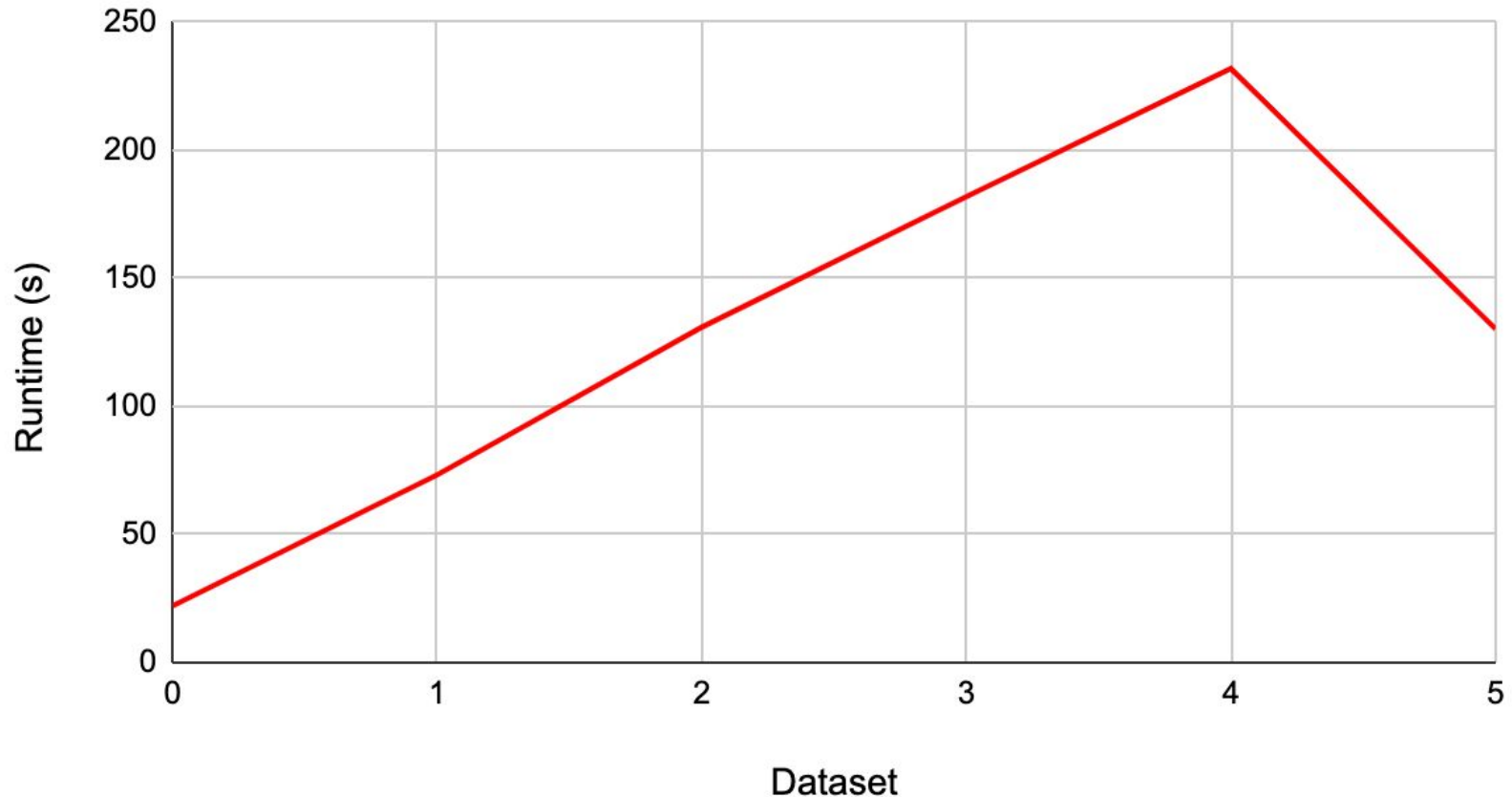
Time and Memory Consumption

Dataset	Runtime (S)	Memory Used (MB)	Accuracy (%)
0	22.08	74.20	71.22
1	73.29	194.74	71.86
2	130.95	314.88	70.74
3	181.91	435.09	72.26
4	231.89	555.46	70.85
5	130.31	244.06	70.91

Table 2: Time and memory consumed by datasets of different sizes.

Time and Memory Consumption

Runtime per dataset

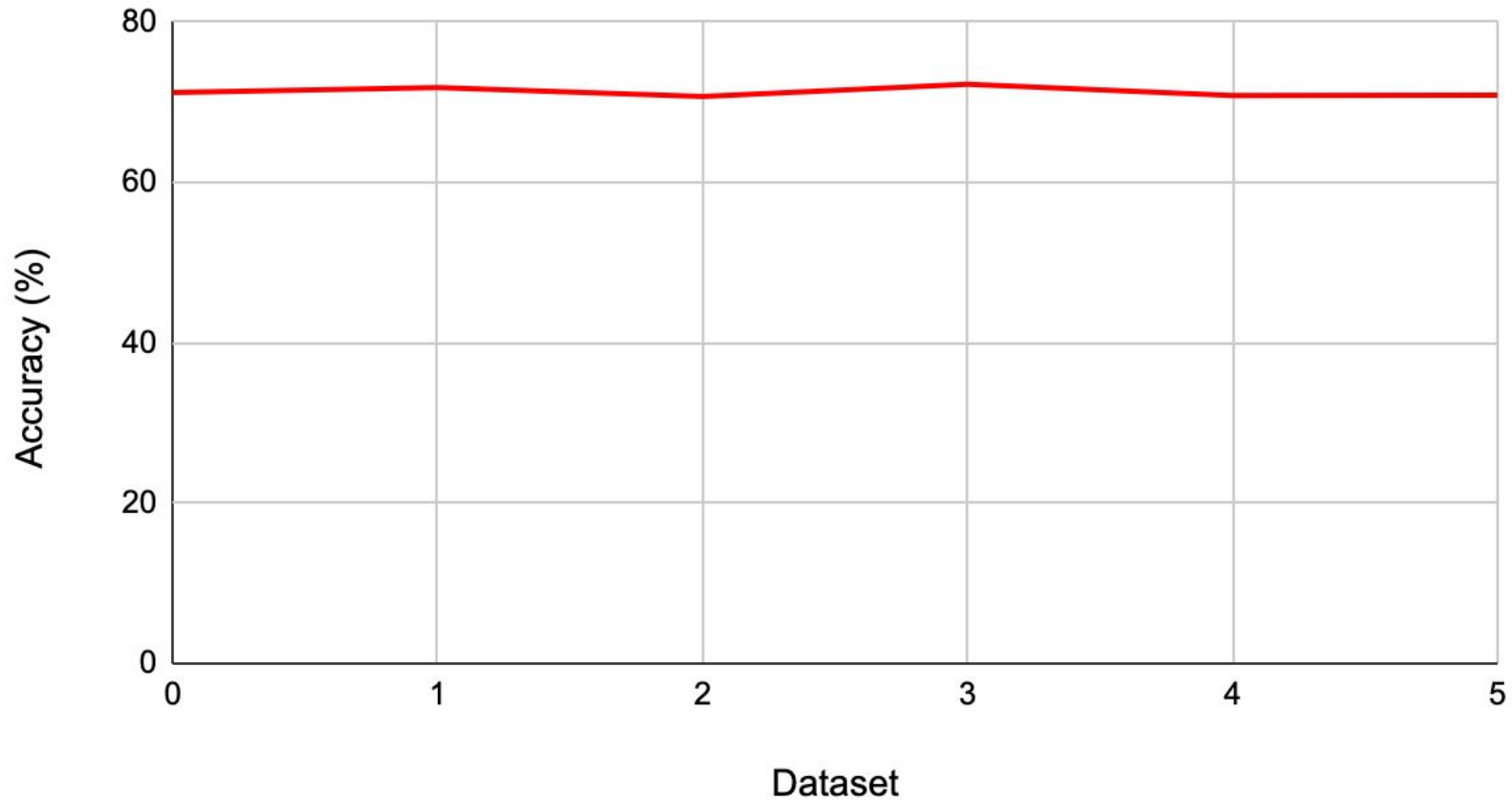


Time and Memory Consumption



Prediction accuracy

Accuracy per dataset



Implementation

		Predicted by the algorithm	
		0	1
Actual Result	0	18733	3749
	1	9358	13160

Table 3: Confusion matrix

Real Life Uses

- Universities and Institutions
- What next for a student
- Decision making

