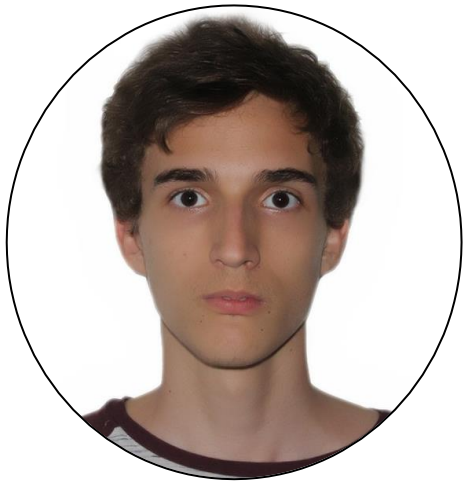
ALGORITHMS BASED ON TREES OF DECISION AND IN DATA OF KNOWLEDGE 11

UNIVERSIDAD EAFIT

Team Presentation

Miguel Ángel Martínez Flórez



Pablo Maya Villegas

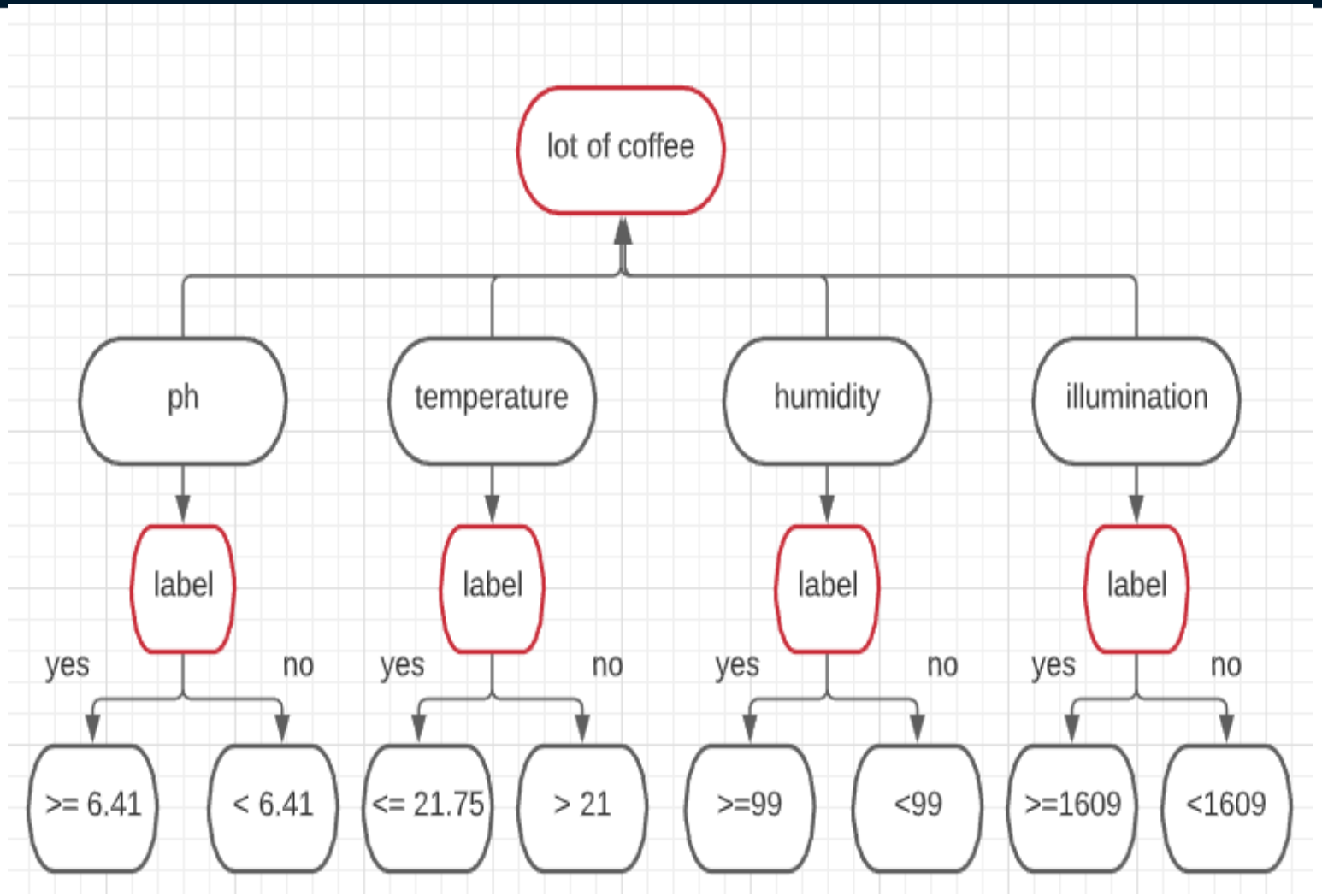
Miguel Correa

Mauricio Toro

https://github.com/pmayavi/ST0245-002/tree/master/proyecto

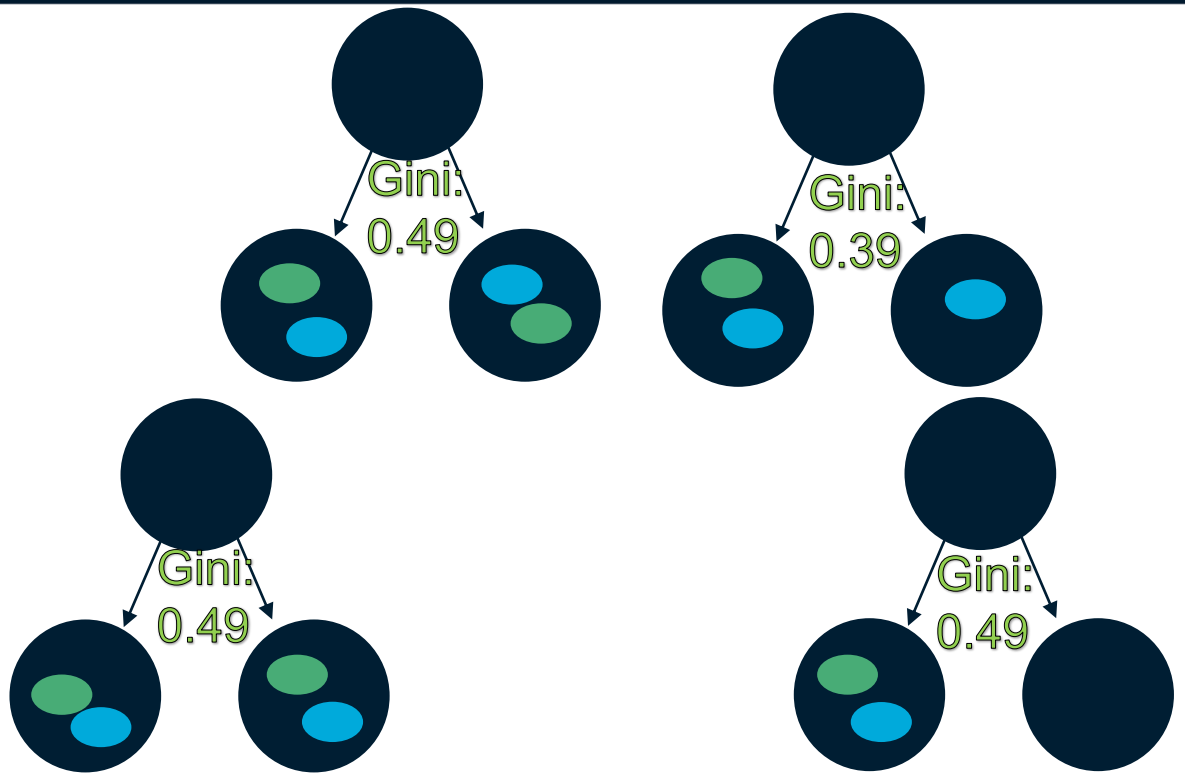
UNIVERSIDAD EAFIT

Algorithm Design

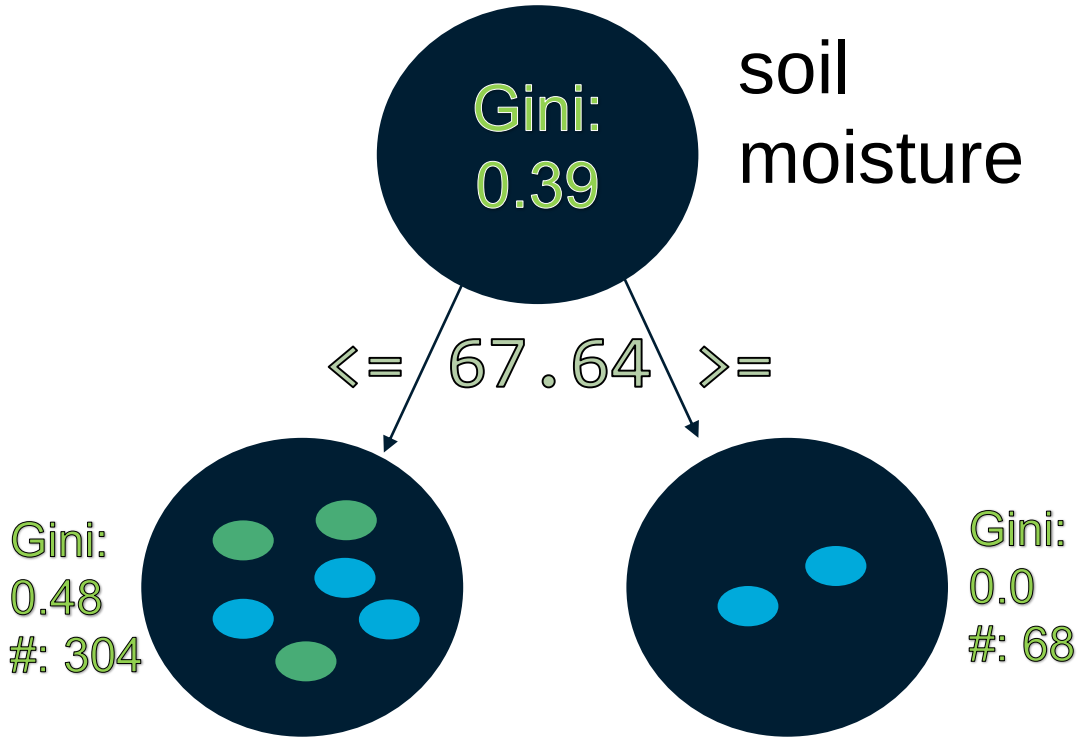


As we know, the main function of the tree is to decide if any of the variables considered that are fundamental for the root, such as temperature, humidity, etc. contain the virus in label and according to the standards of each variable it is indicated whether or not it has a label.
here we use the decision tree id3 to identify that the most influential values are the soil humidity and illuminance.

Division of a node



Testing in all or some values we check in what interval the weighted gini value is the smallest



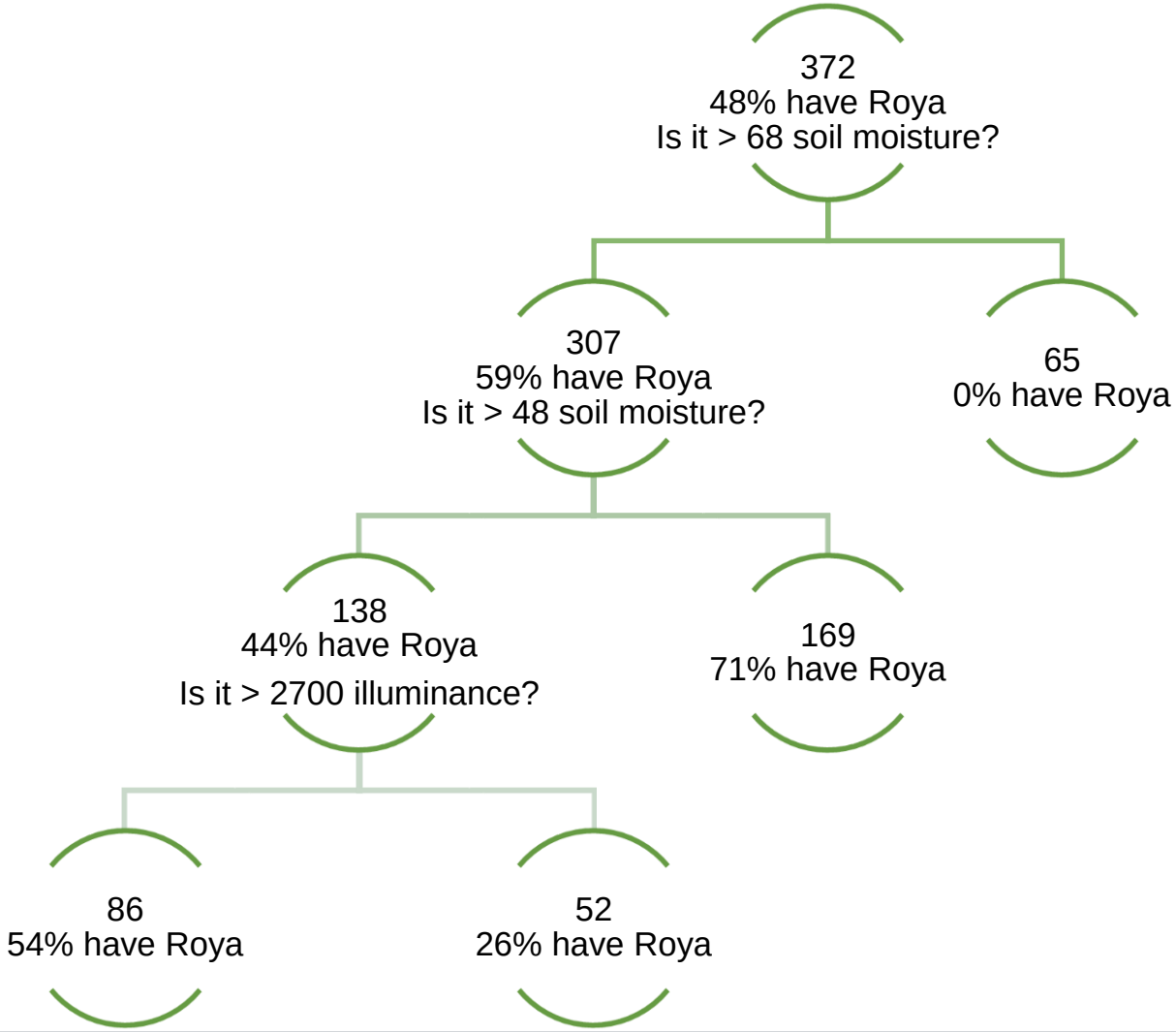
Here we can confirm that when moisture is higher than 68 then its guaranteed that the plant doesn't have the Roya disease and we can separate the tree

Complexity of the Algorithm

	Time complexity	Memory complexity
Train the decision tree	O(N*M)	O(N*M)
Validate the decision tree	O(N)	O(1)
Reader.reader()	O(N)	
Trainer.main()	O(N*M)	
Trainer.trainer()	O(N*M)	
Tree.main()	O(N)	
Tree.tree()	O(N)	

The time and memory complexity of our algorithm is performed with CART. The n and m terms are representations of the complexity of the algorithm.
. N is the amount of data in the file.
. M is the number of times the cycle of searching for the best division number is repeated.

Decision Tree Model



Relevant characteristics:

Soil moisture

Illuminance

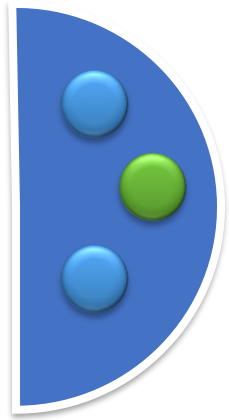
The CART decision tree, it tries multiple values and selects the one that leaves the purest node possible, then it divides the impure one

UNIVERSIDAD EAFIT

Evaluation Metrics

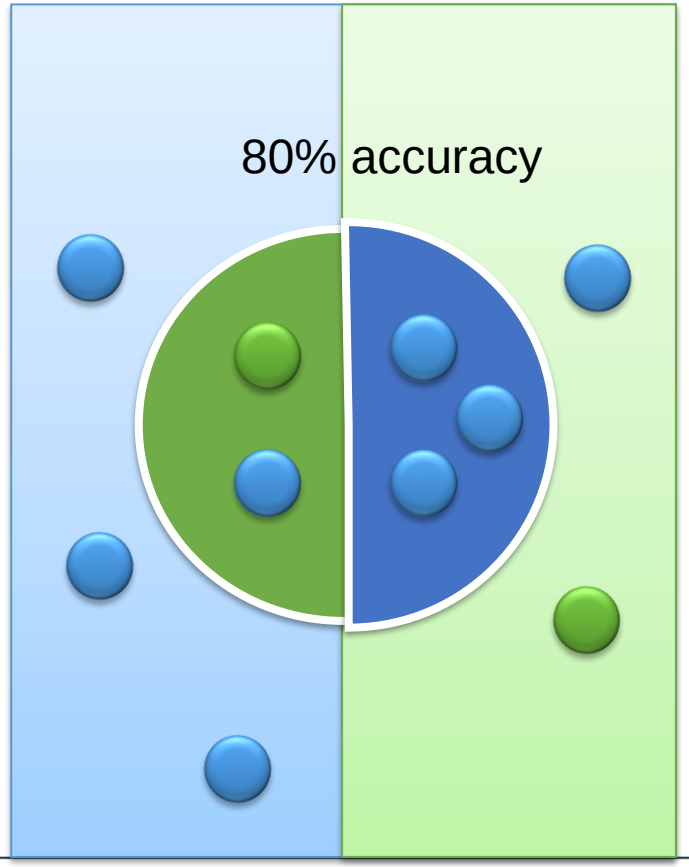
Precision, the percentage of values classified correctly in the field

66% precision



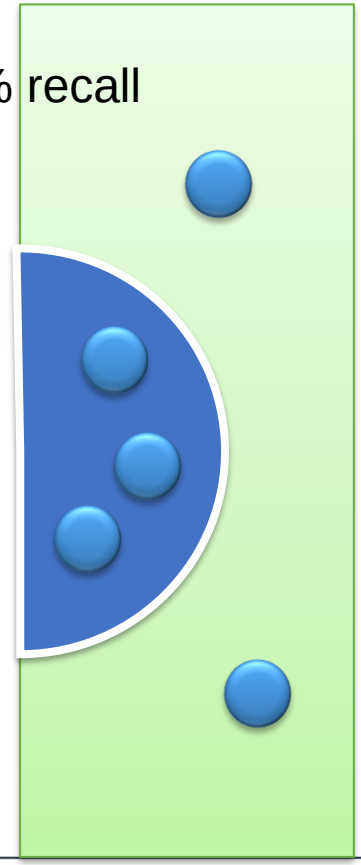
Accuracy, the percentage of values classified accordingly

80% accuracy



Recall, the percentage of values classified correctly from the whole data

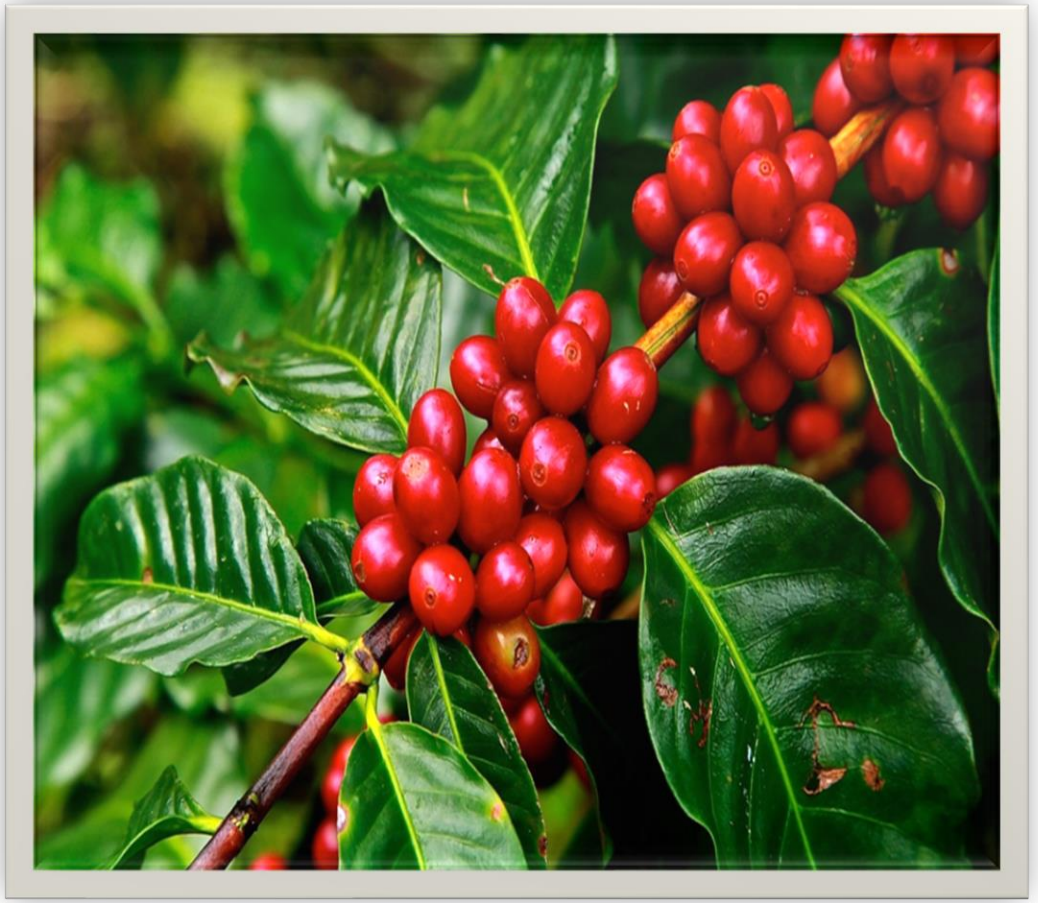
60% recall



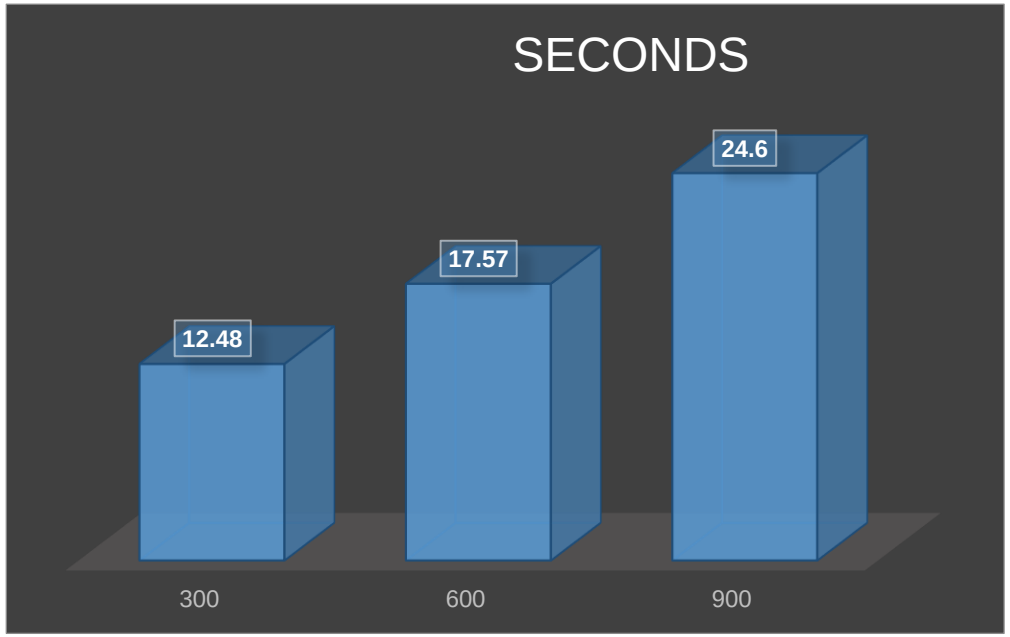
Evaluation Metrics

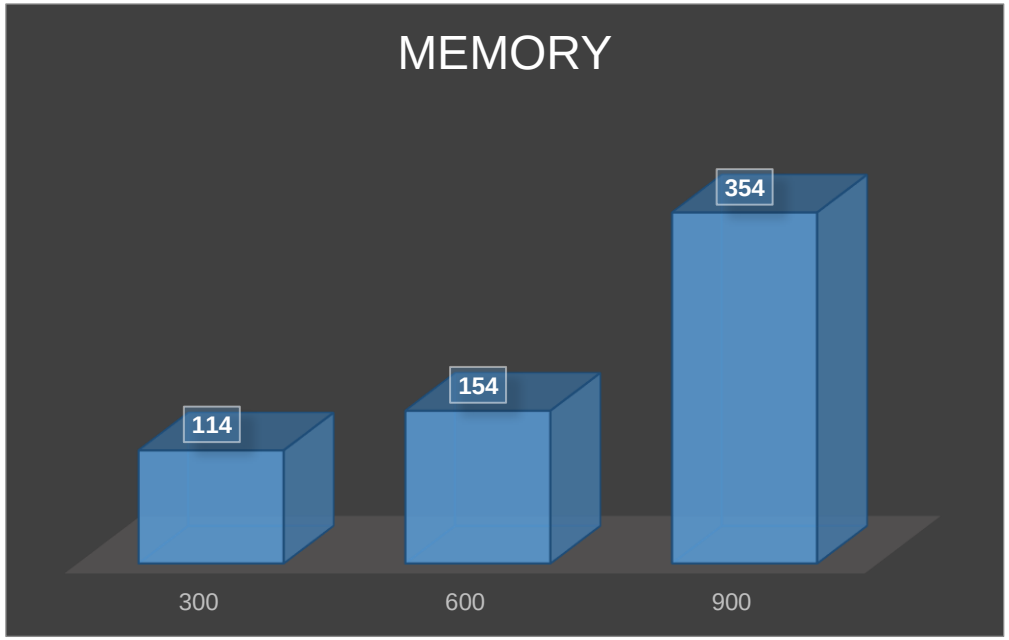
	Data set 1	*Data set 2*	*Data set 3*
Accuracy	85%	81%	62%
Precision	90%	66%	52%
Recall	96%	80%	84%

Evaluation metrics obtained with the training data sets of size 300, 600 and 900



Time and memory consumption





Time consumption

Memory consumption

THANK YOU FOR LISTENING TO US!