Coffee Rust Disease Identification Using Decision Tree Algorithms

Ana Sofia Gutiérrez Tejada
Santiago Hidalgo Ocampo
Medellín, 31/October

Data Structures Designed

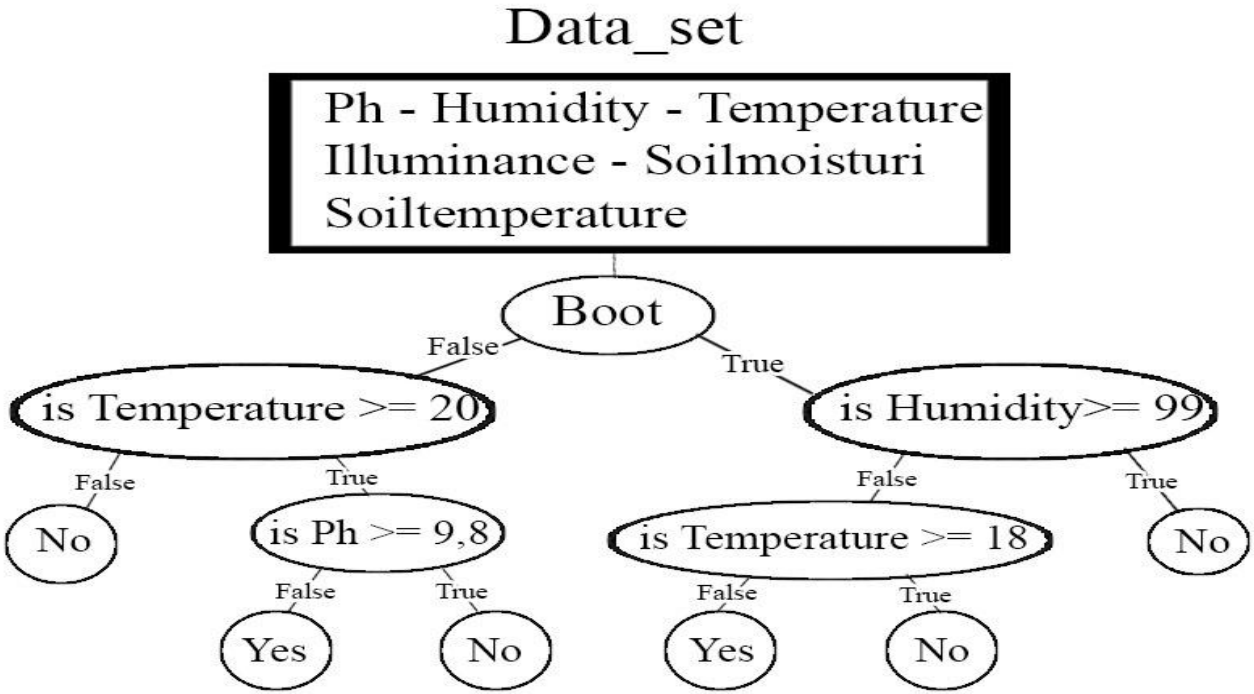

Figure 1: *Tree construction example*

Data Structure Operations

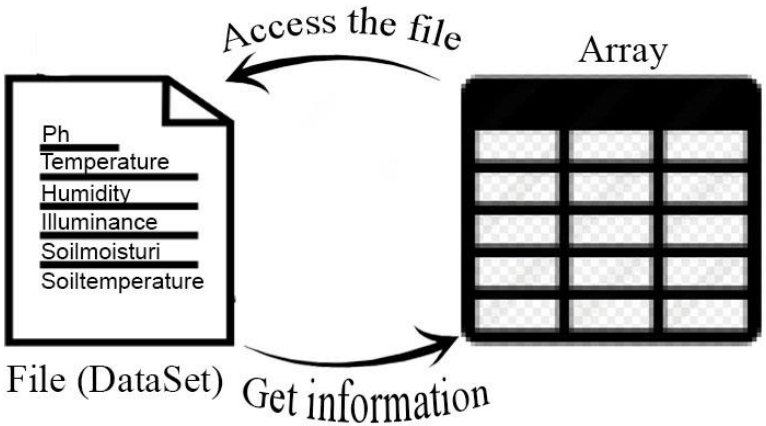

Figure 2: File reading.

Method	Complexity
Read data	O(n^2)
Unique values	O(n)
Is numeric?	O (1)
Find Best Split	O(n^2)
Gini	O(n)
Tree Building	O(n^2)

Table 1: Table to report complexity analysis

Design Criteria of the Data Structure

- The CART algorithm was chosen because it has a great predictive capacity with respect to the other algorithms such as ID3, C4.5 or the CHAID
- The most striking aspect of this algorithm is that CART selects the cut that leads to the greatest decrease in impurity.
- The criterion of division of this algorithm allows to generate a tree with an acceptable purity with respect to the other algorithms.
- CART can work with continuous variables, which are adjusted to variables of the Data Set given.

Time and Memory Consumption

	Data Set 1	Data set 2	Data Set 3	Data Set 4
File Reading	0.0169 sg	0.0100 sg	0.0100 sg	0.0117 sg
Tree Building	3.6307 sg	1.3421 sg	2.0775 sg	3.3724 sg
Tree printing	0.0155 sg	0.0010 sg	0.0069 sg	0.0156 sg

Table 2: Execution time of the operations of the data structure for each data set

Implementation

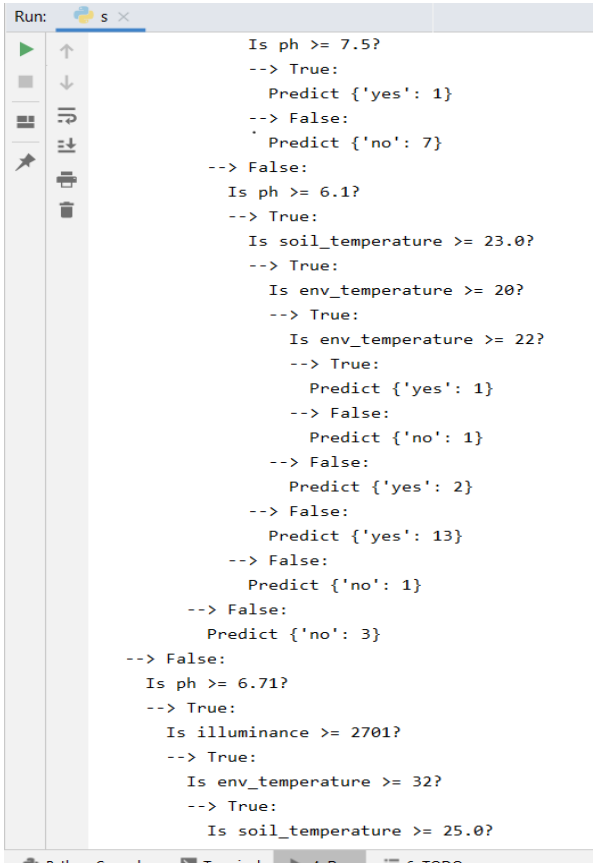

Figure 3: Part of a printed tree

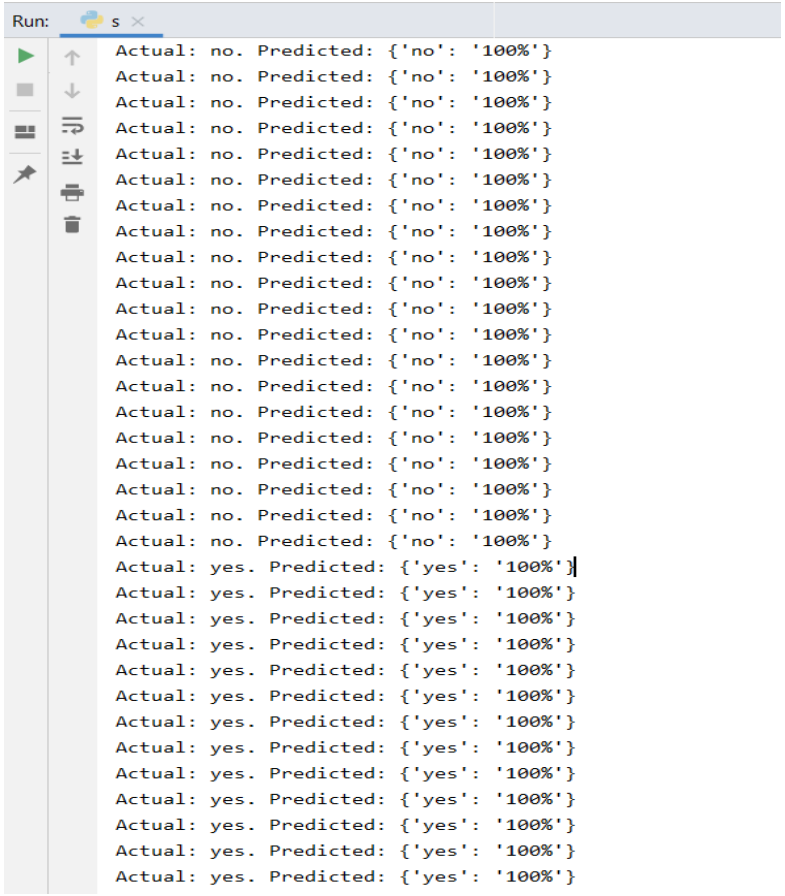

Figure 4: Printing of some evaluated data

Report in arXiv