COFFEE AND RUST: DETECTION AND PREVENTION FOR IMPROVING EXPORTATION QUALITY

Isabella Quintero
Universidad EAFIT
Colombia
iquinterov@eafit.edu.co

Sofia Vega
Universidad EAFIT
Colombia
svegal@eafit.edu.co

Mauricio Toro
Universidad Eafit
Colombia
mtorobe@eafit.edu.co

ABSTRACT

Coffee is a very important product around the world, it is consumed by most of the people we know, therefore, it exists a huge market around it, and the economy of many countries is based on it, such as Brazil, Vietnam and Colombia.

Along the years this countries, and many others, had confronted many problems related to coffee production, the most relevant of all is Rust, it is caused by the fungus Hemileia vastatrix, this plague has threatened crops across Latin-America and Africa, it infects the leaves on the coffee plant, which makes it not suitable for human consumption, and consequently it cannot be exported anymore, this causes very large loses of money and time. What makes this problem so big and uncontrollable is that when farmers can detect it, is too late to handle.

That is why, it becomes necessary to be able to monitor in real time the things that make rust appears in crops, and control the spread, so that the loses can be reduced and the incomes do not descend.

Keywords
Decission tree, physical-chemical factors, early detection, prevention.

ACM CLASSIFICATION Keywords

CCS → Applied computing → Computers in other domains → Agriculture

CCS → Applied computing → Life and medical sciences → Computational biology → Biological networks

1. INTRODUCTION

Nowadays, it is very common for almost everybody to drink a cup of coffee regularly, what we do not always notice, is all that happens in order to produce this simple cup. Coffee producers around the world have confronted different problem related to the coffee plantation, one of them more harmful than the others: Rust.

“The current coffee rust outbreak dates back to 2012. It is a continuing problem, disrupting coffee-growing activities throughout Central America, where more than 1.3 million people depend on the cultivation of this crop. The damage caused by this infestation is compounding the effects of the fluctuation in international coffee prices, particularly since 2013, and the drought conditions in 2015 affecting the quality and quantity of coffee production. One of the main effects of the low coffee yields has been a reduction in incomes and employment opportunities, particularly in the case of small coffee growers. This is the hardest hit population due to its lack of economic means, preventing them from engaging in good crop management practices such as applying fertilizers and fungicides, among others, and their lack of agricultural education for the proper management of coffee plantations.” [1]

For that reason, the purpose of this project is to create a solution by being able to analyze each and every variable involved in the infection and spread of rust in coffee plants, so the problem can be detected in real time, and, then, it can be controlled.

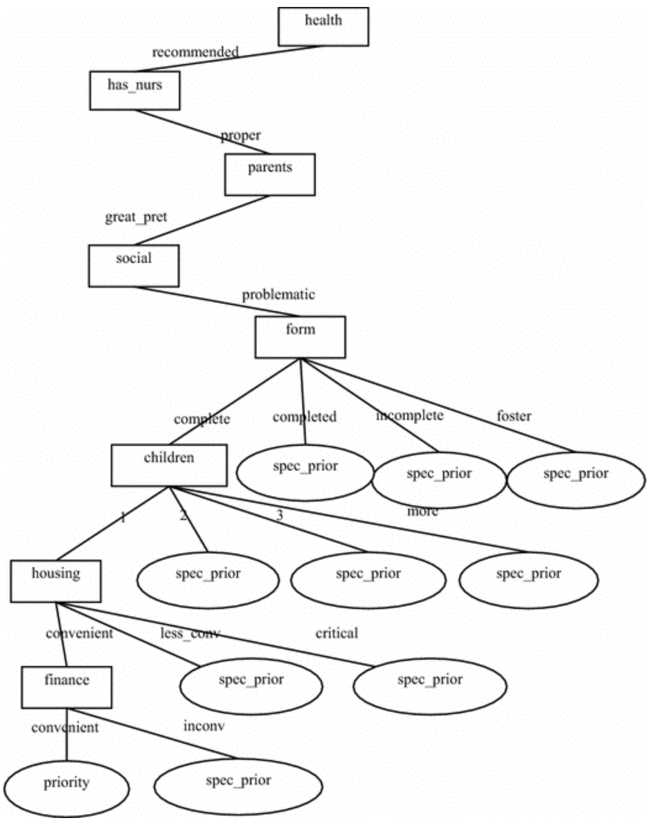
2. PROBLEM

In Colombia, Coffee is the main exportation product, in particular, Caturra variety (coffea arabica) is the most important part of this economy, but it has a very low resistance to rust plague, the prime problem that affects coffee around the world. Besides, the detection of this plague in crops takes a lot of time, therefore, it is very hard to control, which leads to large loses.

3. RELATED WORK

3.1 ID3 ALgorithm

This algorithm constructs a decision tree by a set of examples, is used for classification of future instances. Each example has various attributes that belong to a class, the tree leaf nodes contain the name of the class, while non-leaf nodes are the decision nodes where each one of them, corresponds to a possible attribute value. Every node of decision is a proof of the attribute with other tree that begins from it, the algorithm ID3 use the “information gain” to decide which aspect goes in each node of decision, with this we can define how well an attribute divide de examples of training in every class; it uses a concept named “Entropy” that corresponds to the quantity of an attribute´s uncertainty and it´s formula equals to $\sum -px \, log2 \; px$, where x is the set of classes in S and px is the proportion of S that belongs to the class x



Branch of decision tree based on ID3 algorithm [5]

3.2 C4.5 Algorithm

This algorithm is the successor of ID3; the difference is that C4.5 converts the trained trees into sets of if-then rules. Each one is evaluated for their accuracy looking for established the order in which they should be applied, when this thing happens is called “Pruning” and means removing a rule´s precondition if the accuracy of the rule improves without it.

3.3 C5.0 Algorithm

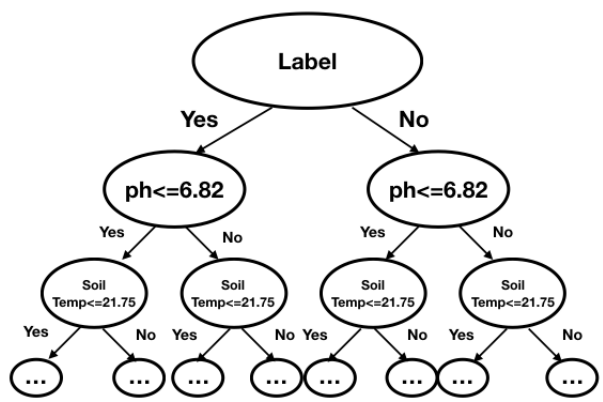
This algorithm constructs the trees based on a set of data of the optimized training under the criteria of information gain and corresponds to an evolution of its last version, the algorithm C4.5; the biggest advantages of this version are related with the efficiency in the tree´s construction time, the memory usage and the obtaining of shorter trees than in C4-5 algorithm, with the same predictive capacity.

Additionally, has the option of consider some attributes in order to focus the construction of the tree and being able to use a penalized learning can be used in which it is possible to assign a value to the possible outcomes.

3.4 CART Algorithm

This algorithm is a binary decision tree that is constructed by splitting a node into two little node recursively, beginning with the principal node, the root, that contains the whole learning sample; each split depends on the value of only one predictor variable, if X is a nominal categorical variable of I categories, there are (2I-1)-1 possible splits for this predictor; on the other hand, if X is an ordinal categorical or continuous variable with K different values, there are K-1 splits on X.

4. Tree Set



4.1 Operations of the data structure

4.1.2. Add data

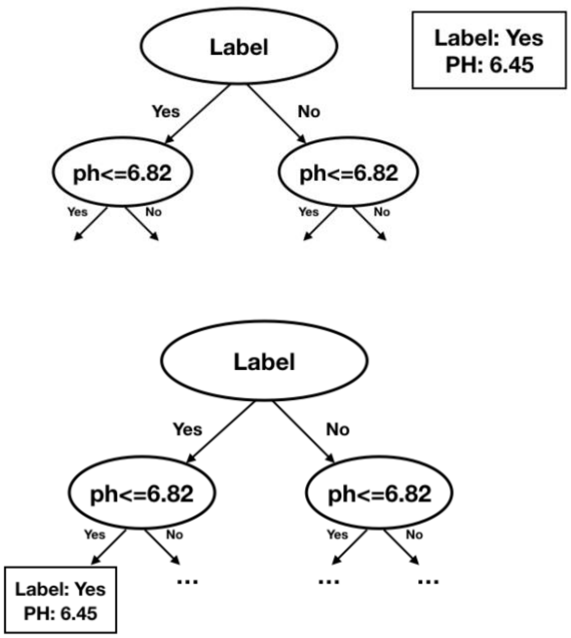


Figure 2: Add data to tree set.

4.2 Design criteria of the data structure

To chose this data structure we based mainly in th amount of variables we were taking into consideration, therefore, the tree set begins wirh label, this variable practily breaks in half all of our data, then, we use all the others variables, so we can obtain an improved classification, and consequently analyze more carefully in which cases rust can be detected in coffe plantations, leading us to determine the ideal enviroment conditions where coffee will be free from rust, and

4.3 Complexity analysis

Method	Complexity
Info()	O(n)
Infox	O(n)
Gain()	O(1)

REFERENCES

1. The impact of the coffee rust outbreack on the coff e sector in central America, 2016. Retrieved August 09, 2019 from Fews net: https://fews.net/sites/default/f iles/documents/reports/CENTRAL%20AMERICA %20-%2 0Special%20Report%20- %20Coffee%20Sector%20-%202 016.pdf

2. Dupouy, C. Aplicación de árboles de decisión para la estimación del escenario económico y la estimación de movimiento la tasa interés en Chile, Universidad de Chile, 2014, retrieved July 14. 20- 23.

3. CART Algorithms, 2019, 2013. Retrieved August 08, from IBM knowledge centerhttps://www.ibm.co m/support/knowledgecenter/en/SSLVMB_22.0.0/c om.ibm.spss.statistics.algorithms/alg_tree- cart.htm

4. Tree algorithms: ID3,C4.5, C5.0 and Cart, 2019. Retrieved August 08, 2019, from Medium Corporation US:https://mediu m.com/datadriveninvestor/tree-algorithms-id3-c4- 5-c5-0-and-cart-413387342164

5. https://www.researchgate.net/figure/Branch-of- decision-tree-based-on-ID3- algorithm_fig3_323401116