

Data structures for the detection of infected crops.

Mauricio Jaramillo Uparela

Juan Pablo Ossa Zapata

Medellín, October 29

Designed data structure

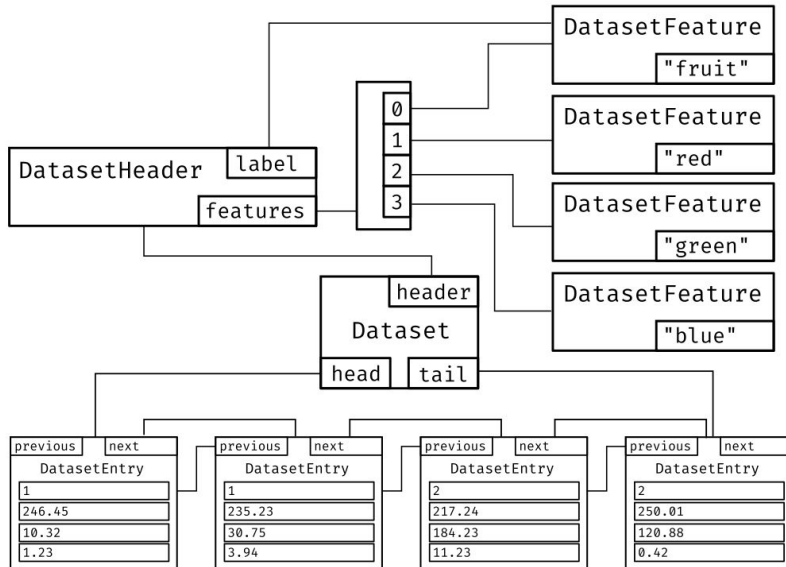


Figure 1: Dataset: header, features, entries.

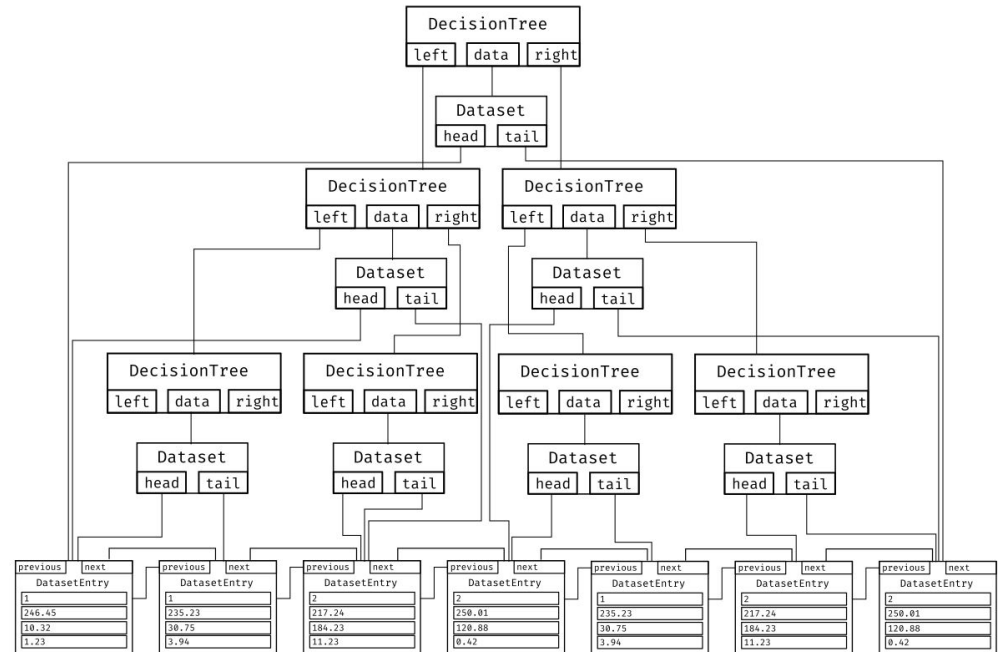


Figure 2: Example of the structure of a decision tree.

Data Structure Operations

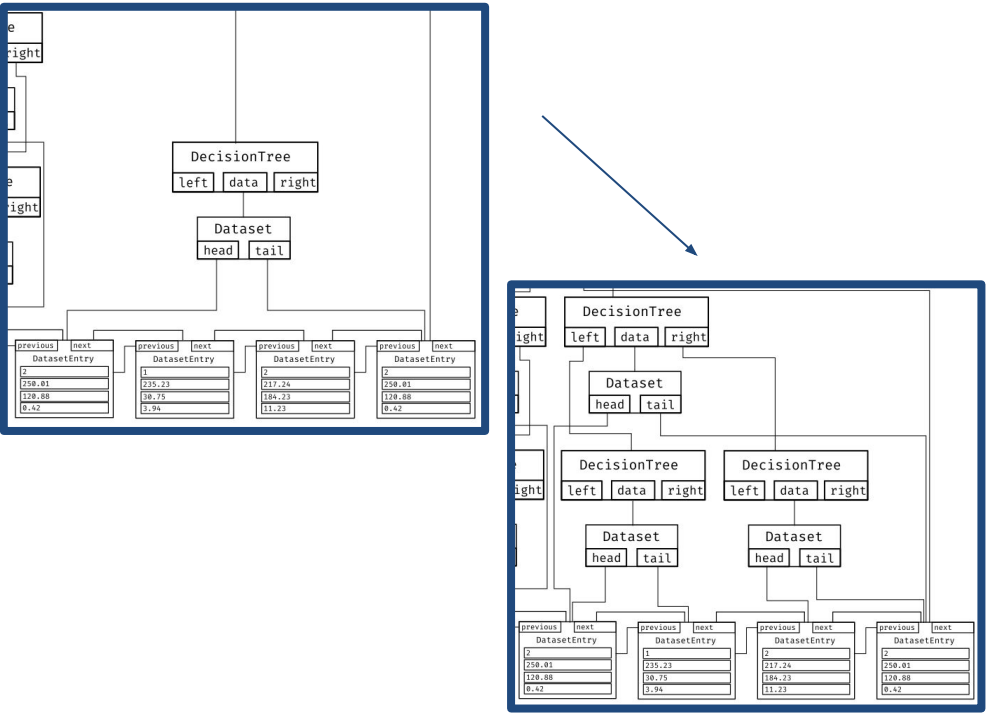


Figure 3: Training a decision tree.

Task	Complexity
Loading from disk	$O(n)$
Find best information gain split	$O(mno)$
Train decision tree	$O(n^3mo)$
Build full random forest	$O(2^n n^3 mo)$

Table 1: complexity of operations.
 m: Number of entries.
 n: Number of features.
 o: Precision used for split.

Design Criteria of the Data Structure

- Reduce memory redundancy of shared data: Two or more datasets can point to different segments of the same linked list.
- Efficient loading from disk: Linked lists allow $O(1)$ insertion at any position with a reference.
- Random access not required: All operations work on all the data at once.

Time and Memory Consumption

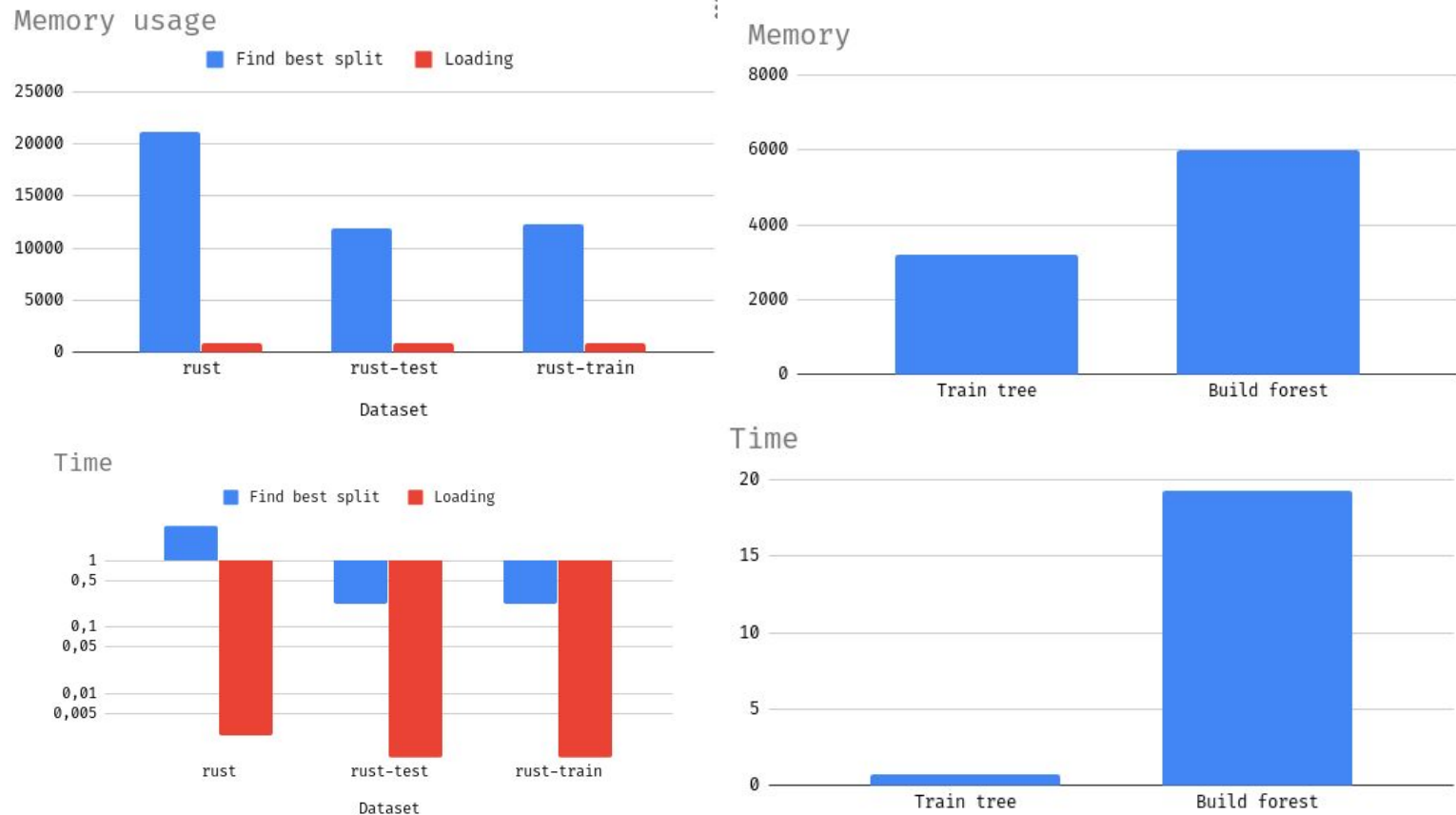


Figure 4: Memory (Kb) and time (Sec) usage of various tasks

Implementation

```

Training tree 101011
Training tree 011011
Training tree 111011
Training tree 000111
Training tree 100111
Training tree 010111
Training tree 110111
Training tree 001111
Training tree 101111
Training tree 011111
Testing forest:
10111010011011011111111111111100101011111111
1100111100011101111111111111011000011110100
1010111001100001000001101110010111001110011
0111110100100010111110110001110000001111
010001101011101110001110101110111011010010
011111001110100001011111010011001110101011
01101010101101011101110001011110101101100
00010111110101100111
Passed 179 out of 300: 0.596667 accuracy.
Testing tree:
10010111011100101110111001110100101110
0011111001111011101011111001100111100100
1000111011110001100101011110111011100110001
0111110110111000111110011000111101100101
0010011001111010100111101011101110111000
011011011110100111011111001110011101001
1110011011101111111111001011111111100100
11010110110100110101
Passed 188 out of 300: 0.626667 accuracy.

```

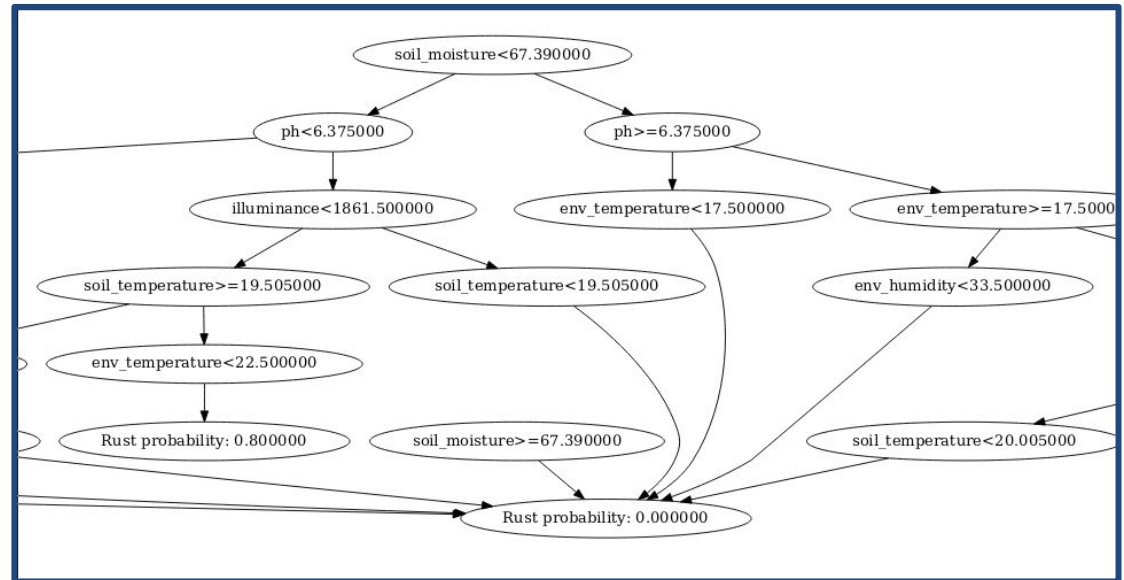


Figure 5: Example output of the program.

Figure 6: Part of the output tree visualization