Gustavo Gutiérrez Gómez
Franco Valencia Adan

# Lab 5 Report

## Implementing Decision Trees

This lab revolved around the implementation of Decision Trees, which are defined as a decision support tool that uses a tree-like graph of decisions and the possible consequences. This tool is usually used in decision analysis and in machine learning.

For this Lab, we implemented the algorithm ID3 to construct a decision tree given a set of attributes, and a training dataset of instances that have different values of the attributes presented.

### Implementation vs WEKA

The advantages of writing a program versus a suite like WEKA, is that the program has our own implementation is that we can utilize that program as we want to; from being just a tool to practice and learn more about decisiones trees, to using it as a mean to serve a bigger program that takes advantage of the implementation and produces useful things or applications.

The disadvantages of writing a program versus WEKA is the complexity and possible struggle of implementing the algorithm, basically redoing something that has been done already by another person (or persons). The fact that a suite like WEKA exists, makes way easier to teachers to teach the subject and raise the curiosity and passion in students for this kind of tools, but in the other hand, it has a learning curve that we cannot ignore on learning how to use the software.

### Criteria of selected datasets

We decided to use the dataset named ‘1984 United States Congressional Voting Records Database’, because we were curious on the tendencies and preferences of citizens of the US back in 1984. The reason behind our curiosity is the political scenario that the world is living right now, with Donald Trump as the President of the US.

Finally, the other reason we selected this dataset was because it had enough instances (data rows), being useful to test if our implementation was good enough not to crash while processing that data.

### Comparison of decision trees

The following sections will present the generated tree in the same format as the one used in our own implementation. After each tree is presented, an analysis on the output will be made by us.

**Implemented tree (algorithm ID3)**

```
'physician-fee-freeze': 'n'
  'education-spending': 'n'
```

Gustavo Gutiérrez Gómez
Franco Valencia Adan

```
    'adoption-of-the-budget-resolution': 'n'
      'synfuels-corporation-cutback': 'n'
        'religious-groups-in-schools': 'n'
          'crime': 'n'
            ANSWER: 'democrat'
          'crime': 'y'
            ANSWER: 'republican'
        'religious-groups-in-schools': 'y'
          ANSWER: 'democrat'
      'synfuels-corporation-cutback': 'y'
        ANSWER: 'democrat'
    'adoption-of-the-budget-resolution': 'y'
      ANSWER: 'democrat'
  'education-spending': 'y'
    ANSWER: 'democrat'
'physician-fee-freeze': 'y'
  'export-administration-act-south-africa': 'n'
    'synfuels-corporation-cutback': 'n'
      ANSWER: 'republican'
    'synfuels-corporation-cutback': 'y'
      'adoption-of-the-budget-resolution': 'n'
        'superfund-right-to-sue': 'n'
          ANSWER: 'democrat'
        'superfund-right-to-sue': 'y'
          'water-project-cost-sharing': 'n'
            'handicapped-infants': 'n'
              ANSWER: 'democrat'
            'handicapped-infants': 'y'
              ANSWER: 'republican'
          'water-project-cost-sharing': 'y'
            ANSWER: 'republican'
      'adoption-of-the-budget-resolution': 'y'
        ANSWER: 'democrat'
  'export-administration-act-south-africa': 'y'
    'synfuels-corporation-cutback': 'n'
      'superfund-right-to-sue': 'n'
        ANSWER: 'republican'
      'superfund-right-to-sue': 'y'
        ANSWER: 'republican'
    'synfuels-corporation-cutback': 'y'
      'mx-missile': 'n'
        'adoption-of-the-budget-resolution': 'n'
          ANSWER: 'republican'
        'adoption-of-the-budget-resolution': 'y'
          'water-project-cost-sharing': 'n'
            ANSWER: 'republican'
          'water-project-cost-sharing': 'y'
            ANSWER: 'democrat'
      'mx-missile': 'y'
        'handicapped-infants': 'n'
          ANSWER: 'democrat'
        'handicapped-infants': 'y'
          'adoption-of-the-budget-resolution': 'n'
            ANSWER: 'democrat'
          'adoption-of-the-budget-resolution': 'y'
            ANSWER: 'republican'
```

Gustavo Gutiérrez Gómez
Franco Valencia Adan

**WEKA tree (algorithm J48)**

```
'physician-fee-freeze': 'n'
  ANSWER: 'democrat' (253.41/3.75)
'physician-fee-freeze': 'y'
  'synfuels-corporation-cutback': 'n'
    ANSWER: 'republican' (145.71/4.0)
  'synfuels-corporation-cutback': 'y'
    'mx-missile': 'n'
      'adoption-of-the-budget-resolution': 'n'
        ANSWER: 'republican' (22.61/3.32)
      'adoption-of-the-budget-resolution':y
        'anti-satellite-test-ban': 'n'
          ANSWER: 'democrat' (5.04/0.02)
        'anti-satellite-test-ban': 'y'
          ANSWER: 'republican' (2.21)
    'mx-missile': 'y'
      ANSWER: 'democrat' (6.03/1.03)
```

**Conclusions**

As you can see, the tree produced by our implementation was way bigger and more complicated. This can be attributed to the algorithms that our implementation used (ID3) versus the one used by WEKA (J48). As we analyse further the generated trees, we can observe that the WEKA's tree has 6 leaves and our implementation has 20; while the size of the WEKA's tree is of 11 and ours is of 39.

We can observe that one of the reasons of this big difference in leaves and size is the fact that WEKA'S tree separated at the very beginning the attribute `'physician-fee-freeze'` and giving an answer of the class right away, while out implementation made various branches from that initial node.

The final thought of this analysis is that ID3 is way simpler on the implementation, which can be useful at first, but J48 is an enhanced version based on ID3, so it is expected to produce a better and simpler decision tree.

Include the graphics of the trees or part of the trees you generated in WEKA and your own program. Are they different, and if so, why?

## Where are decision trees useful?

Based in what you have learned so far where would you use decision trees?

Decision trees are quite a good tool to use because they have several advantages, such as:

- Simple enough to understand and interpret
- Valuable even when there is little data
- New scenarios or attributes can be added easily
- Can be combined with other decision techniques

With all of these advantages, decisiones trees can be useful in areas like astronomy, control systems, financial analysis, manufacturing and production, object recognition, object

classification, pharmacology and many areas more. This areas can be beneficiated by implementations of decision trees that help them with the prediction of possible consequences, outcomes, resource costs and utility.