CT475

Machine Learning & Data Mining

Assignment 3

Name: Taidgh Murray

Student I.D: 15315901

Discipline: College of Science

Course: 4BS2 Undenominated Science (Computing)

In this assignment, we were asked to design, implement, and evaluate a machine learning algorithm, from scratch. The algorithm was to be chosen at our discretion, ensuring to ignore k-Nearest Neighbours, Naïve Bayes, or trivial algorithms such as ZeroR or 1R. It was possible to also create an algorithm from scratch. We were encouraged to work in pairs, but having an introverted, asocial personality, I decided to forego a partner for this assignment, and try to tackle the assignment alone.

As with the previous assignment, I decided to use Python as my language of choice. My experience with it trumps all other programming languages, and the wide variety of libraries available made much of the extraneous programming tasks (mainly reading the ‘owls.csv’ file (or any .csv file the user might want to evaluate the algorithm on), and the random seed generation). Python is simple to understand, and quite portable, which allowed me to easily work on my assignment wherever I could get access to the internet. It also means that reading the code for the algorithm I’ve chosen is much easier than a language like Java, R or MATLAB.

I decided to implement the Classification and Regression Tree algorithm, or CART algorithm. The CART Algorithm can also be called a decision tree algorithm. It in