1 Structure and Design

The decision tree classifier is implemented using the **iterative dichotomizer 3** algorithm, as outlined in the lecture slides. The algorithm from the lecture slides is improved with an insight on how to handle incomplete datasets drawn from the chapter on learning decision trees in Russel and Norvig's *Artificial Intelligence: A Modern Approach*.

The `ID3` class contains an inner class `Dataset`, which is an internal representation of the data passed to the classifier as an argument to the `train()` method. It was added to the design to abstract some of the illegible array manipulation logic that otherwise arose in the implementation of the core ID3 algorithm. As a result of this addition, the code for the training phase is much more legible and easier to modify as well as assess for correctness.

1.1 Training

The `train()` method wraps a call to the recursive `id3()` method, which constructs the decision tree using the pre-defined `TreeNode` data structure. The method recurses on increasingly small subsets of the original dataset and increasingly reduced sets of remaining attributes to split the dataset on. At each level of recursion, four cases can be encountered:

1. There are no more attributes to split the dataset by
2. All examples in the remaining dataset have the same class
3. There are no examples in this subdivision of the dataset
4. None of the above apply

1.2 Classification

Classification is carried out using the stored `TreeNode` data structure. Each node in this data structure has a `value` attribute; for inner node this represents the attribute based on which to make a decision, while for leaf nodes it represents the classification result. The `classify()` method recursively traverses the decision tree, at each node selecting the node's child that matches the example's value for the current node's attribute.

1.3 Testing

The implementation was tested on the provided simple tests. Further tests were developed by modifying the provided tests to include a number of edge cases.