Decision Tree Induction

1. class labeled training tuples

2. internal node checked - test on an attribute using attribute selection measures

3. tree pruning is needed to remove outliers

3. leaf node class label

4. tuple X association class label is unknown

5. decision tree -> classification rules

Advantages

1. appropriate for knowledge discovery

2. can handle high dimensional data

3. simple and fast, accurate

As we are well aware of, Decision Tree Induction is an apt method to identify class labels for each tuple of the training data set. A Decision Tree is a flowchart-like tree structure. The training data set, as a whole is first taken. The data is partitioned by testing the relevant attributes using appropriate attribute selection measures. Every tree has a root node which denotes a test on an attribute dividing into partitions and labeling tuples with their respective classes. Further on, every other internal (non-leaf) node represents a test on an attribute and every branch branching out indicates the result of the test. The leaf nodes or terminal nodes gives the class labels. For every tuple X, a path is traced from the root to the leaf nodes to predict the class label for that tuple.

It is important to use Decision Tree Induction for our structure here because it allows us to convert them into classification rules easily. This methodology proves to be vital for exploratory knowledge discovery. An significant purpose of using this technique is that it can handle high dimensional data. From the performance perspective, it is simple, fast and accurate.

In our design, the data set contains a large set of tuples with unknown class labels. Every tuple holds values from all the parameters of one source at a particular zone. With the standards, labels and domain knowledge at hand, it is possible to segregate these tuples in time according to the class prediction provided by the decision tree algorithm. Every tuple at the end of partition, now bears a respective label such as good, average or poor. Further tree pruning can be implied to identify outliers as well as to gain accurate results.