**INTRODUCTION**

***DECISION TREES, POST PRUNING AND RANDOM FORESTS***

Group Members: -

BHARAT RAGHUNATHAN (2015AAPS263H)

V THEJAS (2015A7PS142H)

NITISH RAVISHANKAR (2015A7PS152H)

The Decision Tree is initially built using the **ID3** (Iterative Dichotomiser 3) algorithm which depends on a single statistical measure of **ENTROPY** or **INFORMATION GAIN** (of each attribute) given by

 where c is the no. of possible values.



Entropy for one attribute with respect to another attribute.

The attribute with Minimum Entropy maximizes Information Gain, and hence, the attribute with maximum entropy is placed at the root of the Decision Tree.

Subsequently, the attributes with higher information gain(s) and lower entropies are placed closer to the root of the tree.

ID3 does not guarantee an optimal solution (may get stuck in local optima) because it follows a **greedy** approach and does not **backtrack**. Due to this, ID3 has the tendency to **overfit** the training data.The attributes with higher information gain are **preferentially selected** and there is an Inductive Bias (not easily evident) that follows: -

Approximate Inductive Bias of Decision Trees: Shorter Trees are preferred over longer trees AND Trees that place high information gain attributes closer to the root are preferred over those that do not.

This directly follows from a principle:

Ockham’s Razor: Prefer Shorter Hypotheses over Longer ones.

ID3 exhibits what is called a *Preferential Bias,* wherein it searches all possible hypotheses but prefers one hypotheses over another with respect to a statistical measure.

In order to reduce overfitting tendencies, we have tried to do *reduced-error pruning*, which tries to prune/remove subtrees and replace them with leaf node(s) by backtracking till there is no reduction in the prediction accuracy (accuracy either improves or stays same).

After this, we have also tried to implement *Random Forests*, which builds an entire tree of possible decision trees (represented by a vector of decision trees) fitting a particular training example or an **ensemble** of decision trees and then takes a majority vote to classify a given test instance as positive or negative and we have also measured accuracy, precision, recall and F-Measure for each of these to evaluate the performance of each of these algorithms’ implementations.