

Al Muhit Muhtadi  
2005013

## Accuracy Results Table:

---

	Average Accuracy Over 20 Runs	
Attribute selection strategy	Information Gain	Gini Impurity
Always select the best attribute	0.937638%	0.934166%
Select one randomly from the top three attributes	0.807993%	0.812608%

---

## Calculation:

Information gain is calculated by comparing the entropy of the original set of data to the entropy of the two child sets. Gini impurity is calculated by subtracting the sum of the squared probabilities of each class from one.

## Performance:

Both are metrics used in decision tree algorithms to measure informativeness, and the best one to use depends on the following factors

### *Target variable*

Information gain is better when the target variable has many possible values, while Gini impurity is better when the target variable has only a few possible values

### *Class balance*

Gini impurity is better when the classes are balanced, while information gain is better when the classes are imbalanced

### *Dataset size*

Gini impurity can be more efficient to calculate than information gain, especially when dealing with large datasets