Information gain is defined as the difference between the original information requirement and the new requirement. It tells us how much would be gained by branching on a attribute.
Biased toward tests with many outcomes, prefers to select attributes having a large number of values. What if you have a attribute that is a unique identifier, result in a large number of partitions

Steps:
First calculate the expected information needed to classify a tuple in D:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i)$$

$$p_i = |C_{i,D}| / |D|$$

p_i is the probability of that a tuple D belongs to class C_j

for each attribute A:
{

Calculate the expected information required to classify a tuple from D based on the partitioning by A

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Info(D_j)$$

Calculate the information gain for the attribute A:

$$Gain(A) = Info(D) - Info_A(D)$$

}
Then choose the attribute with the highest information gain to split on.

Gain ratio attempts to over come the bias with information gain on attributes having a large number of values. It does this by applying a form of normalization to info gain using a split information value:

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right)$$

SplitInfo considers the number of tuples having the outcome with respect to the total number of tuples in D.

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

Gini index measures the impurity of a data partition or a set of training tuples. Gini index considers the binary split of each attribute. To determine the best binary split examine all subsets of A excluding the power and empty sets.

Steps for selection using Gini index:

First calculate the impurity of D:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2$$

For each attribute A

{

for all possible binary splits of A, A partions D into D_1 and D_2

{

Calculate the gini index of that split:

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)$$

}

Choose the binary split with the lowest gini index.

Calculate the change in impurity for the lowest gini index:

(The symbol for “change in” is delta(triangle))

$$\Delta Gini(A) = Gini(D) - Gini_A(D)$$

}

Select the attribute that maximizes the reduction in impurity and split on.

Information Gain:

$$Info(D) = -\frac{4}{24}\log_2(\frac{4}{24}) - \frac{5}{24}\log_2(\frac{5}{24}) - \frac{15}{24}\log_2(\frac{15}{24}) = 1.32\,bits$$

$$Info_{age}(D) = \frac{8}{24}\times(-\frac{2}{8}\log_2(\frac{2}{8}) - \frac{2}{8}\log_2(\frac{2}{8}) - \frac{4}{8}\log_2(\frac{4}{8})) = 0.5$$

$$+\frac{8}{24}\times(-\frac{1}{8}\log_2(\frac{1}{8}) - \frac{2}{8}\log_2(\frac{2}{8}) - \frac{5}{8}\log_2(\frac{5}{8})) = 0.43$$

$$+\frac{8}{24}\times(-\frac{1}{8}\log_2(\frac{1}{8}) - \frac{1}{8}\log_2(\frac{1}{8}) - \frac{6}{8}\log_2(\frac{6}{8})) = 0.35$$

$$Info_{age}(D) = 1.28\,bits$$

$$Gain(age) = 1.32 - 1.28 = 0.04\,bits$$

$$Info_{specRx}(D) = \frac{12}{24}\times(-\frac{3}{12}\log_2(\frac{3}{12}) - \frac{2}{12}\log_2(\frac{2}{12}) - \frac{7}{12}\log_2(\frac{7}{12})) = 0.69$$

$$+\frac{12}{24}\times(-\frac{1}{12}\log_2(\frac{1}{12}) - \frac{3}{12}\log_2(\frac{3}{12}) - \frac{8}{12}\log_2(\frac{8}{12})) = 0.59$$

$$Info_{age}(D) = 1.28\,bits$$

$$Gain(specRx) = 1.32 - 1.28 = 0.04\,bits$$

$$Info_{asig}(D) = \frac{12}{24}\times(-\frac{5}{12}\log_2(\frac{5}{12}) - \frac{7}{12}\log_2(\frac{7}{12})) = 0.48$$

$$+\frac{12}{24}\times(-\frac{4}{12}\log_2(\frac{4}{12}) - \frac{8}{12}\log_2(\frac{8}{12})) = 0.45$$

$$Info_{age}(D) = 0.93\,bits$$

$$Gain(asig) = 1.32 - 0.93 = 0.39\,bits$$

$$Info_{tears}(D) = \frac{12}{24}\times(-1\log_2(1)) = 0$$

$$+\frac{12}{24}\times(-\frac{4}{12}\log_2(\frac{4}{12}) - \frac{5}{12}\log_2(\frac{5}{12}) - \frac{3}{12}\log_2(\frac{3}{12})) = 0.77$$

$$Info_{tears}(D) = 0.77\,bits$$

$$Gain(tears) = 1.32 - 0.77 = 0.55\,bits$$

attribute ranks: tears > asig > specRx = age

Gain Ratio

$$SplitInfo_{age}(D)=\frac{8}{24}\times\log_2(\frac{8}{24})+\frac{8}{24}\times\log_2(\frac{8}{24})+\frac{8}{24}\times\log_2(\frac{8}{24})=1.58$$

$$SplitInfo_{specRx}(D)=\frac{12}{24}\times\log_2(\frac{12}{24})+\frac{12}{24}\times\log_2(\frac{12}{24})=1.5$$

$$GainRatio(age)=\frac{0.04}{1.58}=0.025$$

$$GainRatio(specRx)=\frac{0.04}{1.5}=0.0026$$

$$SplitInfo_{astig}(D)=\frac{12}{24}\times\log_2(\frac{12}{24})+\frac{12}{24}\times\log_2(\frac{12}{24})=1.5$$

$$SplitInfo_{tears}(D)=\frac{12}{24}\times\log_2(\frac{12}{24})+\frac{12}{24}\times\log_2(\frac{12}{24})=1.5$$

$$GainRatio(astig)=\frac{0.39}{1.5}=0.26$$

$$GainRatio(tears)=\frac{0.55}{1.5}=0.36$$

Attribute ranks: tears > astig > age > specRx

Gini index

$$Gini(D) = 1 - (\frac{4}{24})^2 - (\frac{5}{24})^2 - (\frac{15}{24})^2 = 0.538$$

Subsets for age:

{1,2} ,{2,3} ,{1,3} ,{1} ,{2} ,{3}

$$Gini_{age} \in \{1,2\}^{(D)}$$

$$= \frac{16}{24} Gini(\{1,2\}) + \frac{8}{24} Gini(\{3\})$$

$$= \frac{16}{24}(1 - (\frac{3}{16})^2 - (\frac{4}{16})^2 - (\frac{9}{16})^2) + \frac{8}{16}(1 - (\frac{1}{8})^2 - (\frac{1}{8})^2 - (\frac{6}{8})^2)$$

$$= \frac{16}{24} 0.585 + \frac{8}{16} 0.406$$

$$= 1.296$$

$$= Gini_{age} \in \{3\}^{(D)}$$

$$Gini_{age} \in \{2,3\}^{(D)}$$

$$= \frac{16}{24} Gini(\{2,3\}) + \frac{8}{24} Gini(\{1\})$$

$$= \frac{16}{24}(1 - (\frac{2}{16})^2 - (\frac{3}{16})^2 - (\frac{11}{16})^2) + \frac{8}{16}(1 - (\frac{2}{8})^2 - (\frac{2}{8})^2 - (\frac{4}{8})^2)$$

$$= \frac{16}{24} 0.476 + \frac{8}{16} 0.625$$

$$= 1.442$$

$$= Gini_{age} \in \{1\}^{(D)}$$

$$Gini_{age} \in \{1,3\}^{(D)}$$

$$= \frac{16}{24} Gini(\{1,3\}) + \frac{8}{24} Gini(\{2\})$$

$$= \frac{16}{24}(1 - (\frac{3}{16})^2 - (\frac{3}{16})^2 - (\frac{10}{16})^2) + \frac{8}{16}(1 - (\frac{1}{8})^2 - (\frac{2}{8})^2 - (\frac{5}{8})^2)$$

$$= \frac{16}{24} 0.539 + \frac{8}{16} 0.531$$

$$= 1.385$$

$$= Gini_{age} \in \{2\}^{(D)}$$

Best binary split is {1,2} as it has the lowest Gini index.

$$\Delta Gini(age) = 0.538 - 1.296 = -0.758$$

Repeat the process for all attributes and then select the highest delta gini index!