

**The Entropy for all overall collection**

**info(D) = -**  **(P = no. of** Yes **or** No **/ all of** Yes **and** No**)**

**Info(D) = - ((8/14) \* + (6/14) \* ) = 0.984**

**Calculate the Gain of Outlook**

**info Outlook(D) = -**

**Outlook**

**Rain**

**Sunny**

**Overcast**

**Yes**

**2**

**No**

**3**

**Yes**

**2**

**No**

**3**

**Yes**

**4**

**No**

**0**

**Yes => 8**

**No => 6**

**(TP) = Total of probability of attribute (Rain, Sunny, Overcast) / Total of probability of Outlook**

***Info* Outlook (*D*) = ((5/14) \* (-(2/5) \* log2(2/5)) - ((3/5) \* log2(3/5))) + ((5/14) \* ((-2/5) \* log2(2/5)) - ((3/5) \* log2(3/5))) + ((4/14) \* ((-4/4) \* log2(4/4)) - ((0/4) \* log2(0/4))) = 0.694 bits**

**Gain (Outlook) = Info(D) – Info outlook(D) = 0.984 – 0.694 = 0.29**

**- ((5/14) \* log2(5/14)) - ((5/14) \* log2 (5/14)) - ((4/14) \* log2 (4/14)) =1.5774**

**Sunny [5]**

**Rain [5]**

**Overcast [4]**

**Saving**

**Calculate the Gain of** **Temperature**

**info Temperature (D) = -**

**Cool**

**Hot**

**Mild**

**Yes**

**2**

**No**

**2**

**Yes**

**2**

**No**

**2**

**Yes**

**4**

**No**

**2**

**Yes => 8**

**No => 6**

**(TP) = Total of probability of attribute (Cool, Hot, Mild) / Total of probability of Temperature**

***Info* Temperature (*D*) = ((4/14) \* (-(2/4) \* log2(2/4)) - ((2/4) \* log2(2/4))) + ((4/14) \* ((-2/4) \* log2(2/4)) - ((2/4) \* log2(2/4))) + ((6/14) \* ((-4/6) \* log2(4/6)) - ((2/6) \* log2(2/6))) = 0.965 bits**

**Gain (Temperature) = Info(D) – Info outlook(D) = 0.984 – 0.965 = 0.019**

**-((4/14) \* log2(4/14))-((4/14)\*log2 (4/14))-((6/14)\*log2 (6/14) = 1.5567**

**Cool [5]**

**Hot [5]**

**Mild [4]**

**Temperature**

**Calculate the Gain of** **Humidity**

**Yes => 8**

**No => 6**

**Normal**

**High**

**Yes**

**5**

**No**

**2**

**Yes**

**3**

**No**

**4**

**info Humidity (D) = -**

**(TP) = Total of probability of attribute (normal, High) / Total of probability of Humidity**

***Info* Humidity (*D*) = ((7/14) \* (-(5/7) \* log2(5/7)) - ((2/7) \* log2(2/7))) + ((7/14) \* ((-3/7) \* log2(3/7)) - ((4/7) \* log2(4/7))) = 0.924 bits**

**Gain (Humidity) = Info(D) – Info outlook(D) = 0.984 – 0.924 = 0.06**

**- ((7/14) \* log2(7/14)) - ((7/14) \* log2 (7/14)) = 1**

**Normal [7]**

**High [7]**

**Humidity**

**Calculate the Gain of Wind**

**Yes => 8**

**No => 6**

**Strong**

**Weak**

**Yes**

**3**

**No**

**2**

**Yes**

**6**

**No**

**3**

**info Wind (D) = -**

**(TP) = Total of probability of attribute (Strong, Weak) / Total of probability of Wind**

***Info* Wind (*D*) = ((5/14) \* (-(3/5) \* log2(3/5)) - ((2/5) \* log2(2/5))) + ((9/14) \* ((-3/9) \* log2(3/9)) - ((6/9) \* log2(6/9))) = 0.937 bits**

**Gain (Wind) = Info(D) – Info outlook(D) = 0.984 – 0.937 = 0.047**

**- ((6/14) \* log2(6/14)) - ((8/14) \* log2 (8/14)) = 0.985**

**Strong [6]**

**Weak [8]**

**Wind**

**Gain Ratio(A) = Gain(A) / Split Info(A)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Wind** | **Humidity** | **Temperature** | **Outlook** |  |
| **0.047** | **0.06** | **0.019** | **0.29** | **Gain** |
| **0.985** | **1** | **1.5567** | **1.5774** | **Split Info** |
| **0.0477** | **0.06** | **0.012** | **0.184** | **Gain Ratio** |

**Outlook attribute with the maximum gain ratio is selected as the splitting attribute**

…..

??

Yes

**Outlook**

**Rain**

**Sunny**

**Overcast**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Wind** | **Humidity** | **Temperature** | **Outlook** |
| No | Strong | Normal | Cool | Rain |
| No | Weak | Normal | Cool | Rain |
| No | Strong | High | Mild | Rain |
| Yes | Weak | High | Mild | Rain |
| Yes | Weak | Normal | Mild | Rain |

**The Entropy for all overall collection**

**info(D) = -**  **(P = no. of** Yes **or** No **/ all of** Yes **and** No**)**

**Info(D) = - ((2/5) \* + (3/5) \* ) = 0.971**

**Calculate the Gain of Temperature**

**info Temperature (D) = -**

**(TP) = Total of probability of attribute (Cool, Mild) / Total of probability of Temperature**

***Info* Temperature (*D*) = ((2/5) \* (-(2/2) \* log2(2/2)) - ((0/2) \* log2(0/2))) + ((3/5) \* ((-2/3) \* log2(2/3)) - ((1/3) \* log2(1/3))) = 0.551 bits**

**Yes => 2**

**No => 3**

**Cool**

**Mild**

**Yes**

**0**

**No**

**2**

**Yes**

**2**

**No**

**1**

**Gain (Temperature) = Info(D) – Info outlook(D) = 0.971 – 0.551 = 0.42**

**Temperature**

**-((2/5) \* log2(2/5)) - ((3/5) \* log2 (3/5) = 0.971**

**Mild [3]**

**Cool [2]**

**Calculate the Gain of Humidity**

**Yes => 2**

**No => 3**

**Normal**

**High**

**Yes**

**1**

**No**

**2**

**Yes**

**1**

**No**

**1**

**info Humidity (D) = -**

**(TP) = Total of probability of attribute (normal, High) / Total of probability of Humidity**

***Info* Humidity (*D*) = ((3/5) \* (-(1/3) \* log2(1/3)) - ((2/3) \* log2(2/3))) + ((2/5) \* ((-1/2) \* log2(1/2)) - ((1/2) \* log2(1/2))) = 0.9508 bits**

**Gain (Humidity) = Info(D) – Info outlook(D) = 0.971 – 0.9508 = 0.0202**

**- ((3/5) \* log2(3/5)) - ((2/5) \* log2 (2/5)) = 0.971**

**Normal [3]**

**High [2]**

**Humidity**

**Calculate the Gain of Wind**

**Yes => 2**

**No => 3**

**Strong**

**Weak**

**Yes**

**0**

**No**

**2**

**Yes**

**2**

**No**

**1**

**info Wind (D) = -**

**(TP) = Total of probability of attribute (Strong, Weak) / Total of probability of Wind**

***Info* Wind (*D*) = ((2/5) \* (-(0/2) \* log2(0/2)) - ((2/2) \* log2(2/2))) + ((3/5) \* ((-2/3) \* log2(2/3)) - ((1/3) \* log2(1/3))) = 0.551 bits**

**Gain (Wind) = Info(D) – Info outlook(D) = 0.984 – 0.551 = 0.433**

**- ((2/5) \* log2(2/5)) - ((3/5) \* log2 (3/5)) = 0.971**

**Strong [2]**

**Weak [3]**

**Wind**

**Gain Ratio(A) = Gain(A) / Split Info(A)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Wind** | **Humidity** | **Temperature** |  |
| **0.433** | **0.0202** | **0.42** | **Gain** |
| **0.971** | **0.971** | **0.971** | **Split Info** |
| **0.446** | **0.021** | **0.433** | **Gain Ratio** |

**Wind attribute with the maximum gain ratio is selected as the splitting attribute**

Yes

??

No

Strong

Weak

**Outlook**

**Rain**

**Sunny**

**Overcast**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Wind** | **Humidity** | **Temperature** | **Outlook** |
| No | Weak | Normal | Cool | Rain |
| Yes | Weak | High | Mild | Rain |
| Yes | Weak | Normal | Mild | Rain |

**The Entropy for all overall collection**

**info(D) = -**  **(P = no. of** Yes **or** No **/ all of** Yes **and** No**)**

**Info(D) = - ((1/3) \* + (2/3) \* ) = 0.918**

**Calculate the Gain of Temperature**

**info Temperature (D) = -**

**(TP) = Total of probability of attribute (Cool, Mild) / Total of probability of Temperature**

**Info Temperature (D) = ((1/3) \* (-(1/1) \* log2(1/1)) - ((0/1) \* log2(0/1))) + ((2/3) \* ((-2/2) \* log2(2/2)) - ((0/2) \* log2(0/2))) = 0 bits**

**Yes => 2**

**No => 1**

**Cool**

**Mild**

**Yes**

**0**

**No**

**1**

**Yes**

**2**

**No**

**0**

**Gain (Temperature) = Info(D) – Info outlook(D) = 0.971 – 0 = 0.971**

**Mild [2]**

**-((2/3) \* log2(2/3)) - ((1/3) \* log2 (1/3) = 0.918**

**Temperature**

**Cool [1]**

**Calculate the Gain of Humidity**

**Yes => 1**

**No => 2**

**Normal**

**High**

**Yes**

**1**

**No**

**1**

**Yes**

**0**

**No**

**1**

**info Humidity (D) = -**

**(TP) = Total of probability of attribute (normal, High) / Total of probability of Humidity**

***Info* Humidity (*D*) = ((2/3) \* (-(1/2) \* log2(1/2)) - ((1/2) \* log2(1/2))) + ((1/3) \* ((-0/1) \* log2(0/1)) - ((1/1) \* log2(1/1))) = 0.667 bits**

**Gain (Humidity) = Info(D) – Info outlook(D) = 0.971 – 0.667 = 0.303**

**- ((2/3) \* log2(2/3)) - ((1/3) \* log2 (1/3)) = 0.918**

**Normal [2]**

**High [1]**

**Humidity**

**Gain Ratio(A) = Gain(A) / Split Info(A)**

|  |  |  |
| --- | --- | --- |
| **Humidity** | **Temperature** |  |
| **0.303** | **0.971** | **Gain** |
| **0.918** | **0.918** | **Split Info** |
| **0.33** | **1.058** | **Gain Ratio** |

**Temperature attribute with the maximum gain ratio is selected as the splitting attribute**

Yes

No

Strong

Weak

**Outlook**

**Rain**

**Sunny**

**Overcast**

??

Yes

No

Mild

Cool

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class** | **Wind** | **Humidity** | **Temperature** | **Outlook** |
| Yes | Weak | Normal | Cool | Sunny |
| No | Strong | High | Hot | Sunny |
| No | Weak | High | Hot | Sunny |
| Yes | Strong | Normal | Mild | Sunny |
| No | Weak | High | Mild | Sunny |

**The Entropy for all overall collection**

**info(D) = -**  **(P = no. of** Yes **or** No **/ all of** Yes **and** No**)**

**Info(D) = - ((2/5) \* + (3/5) \* ) = 0.971**

**Calculate the Gain of Temperature**

**info Temperature (D) = -**

**Cool**

**Hot**

**Mild**

**Yes**

**1**

**No**

**0**

**Yes**

**0**

**No**

**2**

**Yes**

**1**

**No**

**1**

**Yes => 2**

**No => 3**

**(TP) = Total of probability of attribute (Cool, Hot, Mild) / Total of probability of Temperature**

***Info* Temperature (*D*) = ((1/5) \* (-(1/1) \* log2(1/1)) - ((0/1) \* log2(0/1))) + ((2/5) \* ((-2/2) \* log2(2/2)) - ((0/2) \* log2(0/2))) + ((2/5) \* ((-1/2) \* log2(1/2)) - ((1/2) \* log2(1/2))) = 0.4 bits**

**Gain (Temperature) = Info(D) – Info outlook(D) = 0.984 – 0.4 = 0.584**

**-((2/5) \* log2(2/5)) - ((2/5) \* log2 (2/5)) - ((1/5) \* log2 (1/5) = 1.522**

**Cool [1]**

**Hot [2]**

**Mild [2]**

**Temperature**

**Calculate the Gain of Humidity**

**Yes => 2**

**No => 3**

**Normal**

**High**

**Yes**

**2**

**No**

**0**

**Yes**

**0**

**No**

**3**

**info Humidity (D) = -**

**(TP) = Total of probability of attribute (normal, High) / Total of probability of Humidity**

***Info* Humidity (*D*) = ((2/5) \* (-(2/2) \* log2(2/2)) - ((0/2) \* log2(0/2))) + ((3/5) \* ((-0/3) \* log2(0/3)) - ((3/3) \* log2(3/3))) = 0 bits**

**Gain (Humidity) = Info(D) – Info outlook(D) = 0.971 – 0 = 0.971**

**- ((2/5) \* log2(2/5)) - ((3/5) \* log2 (3/5)) = 0.971**

**Normal [2]**

**High [3]**

**Humidity**

**Calculate the Gain of Wind**

**Yes => 2**

**No => 3**

**Strong**

**Weak**

**Yes**

**1**

**No**

**1**

**Yes**

**1**

**No**

**2**

**info Wind (D) = -**

**(TP) = Total of probability of attribute (Strong, Weak) / Total of probability of Wind**

***Info* Wind (*D*) = ((2/5) \* (-(1/2) \* log2(1/2)) - ((1/2) \* log2(1/2))) + ((3/5) \* ((-2/3) \* log2(2/3)) - ((1/3) \* log2(1/3))) = 0.9508 bits**

**Gain (Wind) = Info(D) – Info outlook(D) = 0.984 – 0.9508 = 0.0202**

**- ((2/5) \* log2(2/5)) - ((3/5) \* log2 (3/5)) = 0.971**

**Strong [2]**

**Weak [3]**

**Wind**

**Gain Ratio(A) = Gain(A) / Split Info(A)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Wind** | **Humidity** | **Temperature** |  |
| **0.0202** | **0.971** | **0.584** | **Gain** |
| **0.971** | **0.971** | **1.522** | **Split Info** |
| **0.208** | **1** | **0.384** | **Gain Ratio** |

**Humidity attribute with the maximum gain ratio is selected as the splitting attribute**

**Outlook**

**Rain**

**Sunny**

**Overcast**

Yes

No

High

Normal

Yes

Yes

Mild

No

Cool

No

Weak

Strong