Exercise 2

Classes: mammals, reptile, fish, amphibian, bird
Total dataset: 15

- Mammals: 5
- Reptile: 3
- Fish: 3
- Amphibian: 2
- Bird: 2

Entropy(S) = -[(5/15)log2(5/15) + (3/15)log2(3/15) + (3/15)log2(3/15) + (2/15)log2(2/15) + (2/15)log2(2/15)]
= -[-0.5283 + -0.4644 + -0.4644 + -0.3876 + -0.3876]
= 2.2323
1.8447

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

$$Gain(A) = Entropy(D) - \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times Entropy(D_j)$$

Body Temperatures
D1: Warm-blooded: 7

- Mammals: 5
- Bird: 2

D2: Cold-blooded: 8

- Reptile: 3
- Fish: 3
- Amphibian: 2

Entropy(D1) = - [(5/7)log2(5/7) + (2/7)log2(2/7)]
= - [-0.3467 + -0.5164]
= 0.8631

Entropy(D2) = - [(3/8)log2(3/8) + (3/8)log2(3/8) + (2/8)log2(2/8)]
= - [-0.5306 + -0.5306 + -0.5]
= 1.5612

Gain(Body Temperature) = 2.2323 – [(7/15)*0.8631 + (8/15)*1.5612]
= 2.2323 – [0.40278 + 0.83264]
= 2.2323 – 1.23542
= 0.99688

SplitInfo(Body Temperature) = -[(7/15)log2(7/15) + (8/15)log2(8/15)]
= -[-0.5131 + -0.4837]
= 0.9968

GainRatio(Body Temperature) = 0.99688 / 0.9968 = ***1.0***

<u>Skin Cover</u>
D1: Hair: 3
- Mammals: 3

D2: Scales: 6
- Reptile: 3
- Fish: 3

D3: None: 2
- Amphibian: 2

D4: Feathers: 2
- Bird: 2

D5: Fur: 1
- Mammals: 1

D6: Quills: 1
- Mammals: 1

Entropy(D1) = - [(3/3)log2(3/3)] = 0
Entropy(D2) = - [(3/6)log2(3/6) + (3/6)log2(3/6)]
= - [-0.5 + -0.5]
= 1
Entropy(D3) = - [(2/2)log2(2/2)] = 0
Entropy(D4) = - [(2/2)log2(2/2)] = 0
Entropy(D5) = - [(1/1)log2(1/1)] = 0
Entropy(D6) = - [(1/1)log2(1/1)] = 0

Gain(Skin Cover) = 2.2323 – [(6/15)*1] = 1.8323

SplitInfo(Skin Cover) = -[(3/15)log2(3/15) + (6/15)log2(6/15) + (2/15)log2(2/15) + (2/15)log2(2/15) + (1/15)log2(1/15) + (1/15)log2(1/15)]
= -[-0.4644 + -0.5288 + -0.3876 + -0.3876 + -0.2604 + -0.2604]
= 2.2892

GainRatio(Skin Cover) = 1.8323 / 2.2892 = ***0.8004***

<u>Gives Birth</u>
D1: Yes: 6
- Mammals: 5
- Fish: 1

D2: No: 9
- Reptile: 3
- Fish: 2
- Amphibian: 2
- Bird: 2

Entropy(D1) = - [(5/6)log2(5/6) + (1/6)log2(1/6)]

= - [-0.2192 + -0.4308]
= 0.65

Entropy(D2) = - [(3/9)log2(3/9) + (2/9)log2(2/9) + (2/9)log2(2/9) + (2/9)log2(2/9)]
= -[-0.5283 + -0.4822 + -0.4822 + -0.4822]
= 1.9749

Gain(Gives Birth) = 2.2323 – [(6/15)*0.65 + (9/15)*1.9749]
= 2.2323 – [0.26 + 1.18494]
= 0.78736

SplitInfo(Gives Birth) = -[(6/15)log2(6/15) + (9/15)log2(9/15)]
= -[-0.5288 + -0.4422]
= 0.971

GainRaio(Gives Birth) = 0.78736 / 0.971 = ***0.8109***

<u>Aquatic Creature</u>
D1: Yes: 4

- Fish: 3
- Mammals: 1

D2: No: 7

- Mammals: 4
- Reptile: 2
- Bird: 1

D3: Semi: 4

- Amphibian: 2
- Reptile: 1
- Bird: 1

Entropy(D1) = - [(3/4)log2(3/4) + (1/4)log2(1/4)]
= - [-0.3113 + -0.5]
= 0.8113

Entropy(D2) = - [(4/7)log2(4/7) + (2/7)log2(2/7) + (1/7)log2(1/7)]
= - [-0.4613 + -0.5164 + -0.4011]
= 1.3788

Entropy(D3) = - [(2/4)log2(2/4) + (1/4)log2(1/4) + (1/4)log2(1/4)]
= - [-0.5 + -0.5 + -0.5]
= 1.5

Gain(Aquatic Creature) = 2.2323 – [(4/15)*0.8113 + (7/15)*1.3788 + (4/15)*1.5]
= 2.2323 – [0.2163 + 0.64344 + 0.4]
= 2.2323 – 1.25974
= 0.97256

SplitInfo(Aquatic Creature) = -[(4/15)log2(4/15) + (7/15)log2(7/15) + (4/15)log2(4/15)]
= -[-0.5085 + -0.5131 + -0.5085]
= 1.5301

GainRatio(Aquatic Creature) = 0.97256 / 1.5301 = ***0.6356***

<u>Aerial Creature</u>
D1: Yes: 2

- Mammal: 1
- Bird: 1

D2: No: 13

- Mammal: 4
- Reptile: 3
- Fish: 3
- Amphibian: 2
- Bird: 1

Entropy(D1) = - [(1/2)log2(1/2) + (1/2)log2(1/2)]
= - [-0.5 + -0.5]
= 1.0

Entropy(D2) = - [(4/13)log2(4/13) + (3/13)log2(3/13) + (3/13)log2(3/13) + (2/13)log2(2/13) + (1/13)log2(1/13)]
= - [-0.5232 + -0.4881 + -0.4881 + -0.4154 + -0.285]
= 2.1998

Gain(Aerial Creature) = 2.2323 – [(2/15)*1.0 + (13/15)*2.1998]
= 2.2323 – [0.1333 + 1.9065]
= 0.1925

SplitInfo(Aerial Creature) = - [(2/15)log2(2/15) + (13/15)log2(13/15)]
= - [-0.3876 + -0.1789]
= 0.5665

GainRatio(Aerial Creature) = 0.1925/0.5665 = ***0.3398***

<u>Has Legs</u>
D1: Yes: 10

- Mammal: 4
- Amphibian: 2
- Reptile: 2
- Bird: 2

D2: No: 5

- Reptile: 1
- Fish: 3

- Mammal: 1

Entropy(D1) = - [(4/10)log2(4/10) + (2/10)log2(2/10) + (2/10)log2(2/10) + (2/10)log2(2/10)]
= - [-0.52877 + -0.4644 + -0.4644 + -0.4644]
= 1.92197

Entropy(D2) = - [(1/5)log2(1/5) + (3/5)log2(3/5) + (1/5)log2(1/5)]
= - [-0.4644 + -0.4422 + -0.4644]
= 1.371

Gain(Has Legs) = 2.2323 - [(10/15)*1.92197 + (5/15)*1.371]
= 2.2323 – [1.2813 + 0.457]
= 2.2323 – 1.7383
= 0.494

SplitInfo(Has Legs) = - [(10/15)log2(10/15) + (5/15)log2(5/15)]
= - [-0.39 + -0.5283]
= 0.9183

GainRatio(Has Legs) = 0.494 / 0.9183 = ***0.538***

<u>Hibernate</u>
D1: Yes: 5
- Reptile: 1
- Amphibian: 2
- Mammal: 2

D2: No: 10
- Mammal: 3
- Fish: 3
- Reptile: 2
- Bird: 2

Entropy(D1) = - [(1/5)log2(1/5) + (2/5)log2(2/5) + (2/5)log2(2/5)]
= - [-0.4644 + -0.5288 + -0.5288]
= 1.522

Entropy(D2) = - [(3/10)log2(3/10) + (3/10)log2(3/10) + (2/10)log2(2/10) + (2/10)log2(2/10)]
= - [-0.5211 + -0.5211 + -0.4644 + -0.4644]
= 1.971

Gain(Hibernate) = 2.2323 – [(5/15)*1.522 + (10/15)*1.971]
= 2.2323 – [0.5073 + 1.314]
= 0.411

SplitInfo(Hibernate) = - [(5/15)log2(5/15) + (10/15)log2(10/15)]

= - [-0.5283 + -0.390]
= 0.9183

GainRatio(Hibernate) = 0.411 / 0.9183 = ***0.4476***

GainRatio(Body Temperature) = ***1.0***
GainRatio(Skin Cover) = ***0.8004***
GainRaio(Gives Birth) = ***0.8109***
GainRatio(Aquatic Creature) = ***0.6356***
GainRatio(Aerial Creature) = ***0.3398***
GainRatio(Has Legs) = ***0.538***
GainRatio(Hibernate) = ***0.4476***

Therefore, we split at **Body Temperature**