1. Read pruning, entropy, and information gain in decision tree. Compute entropy manually from slide, put your steps and result in word file.

Solution:

Entropy(Conversion)

= Entropy(5,5,6)

= Entropy(0.3125,0.3125,0.375)

= - 0.3125\*log2(0.3125) - 0.3125\*log2(0.3125) - 0.375\*log2(0.375)

= 1.5794

Entropy(Conversion,Demand)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | High | Medium | Low |  |
| Heavy | 3 | 2 | 2 | 7 |
| Moderate | 2 | 1 | 1 | 4 |
| Low | 0 | 2 | 3 | 5 |

= P(Heavy)\*Entropy(3,2,2) + P(Moderate)\*Entropy(2,1,1) + P(Low)\*Entropy(0,2,3)

= 7/16 \* Entropy(0.429,0.286,0.286) + 4/16 \* Entropy(0.5,0.25,0.25) + 5/16 \* Entropy(0,0.4,0.6)

= 7/16 \* [-0.429 \* log2(0.429) – 0.286 \* log2(0.286) – 0.286 \* log2(0.286)] +

4/16 \* [-0.5 \* log2(0.5) – 0.25 \* log2(0.25) – 0.25 \* log2(0.25)] +

5/16 \* [-0 \* log2(0) – 0.4 \* log2(0.4) – 0.6 \* log2(0.6)]

= 7/16 \* 1.5567 + 4/16 \* 1.5 + 5/16 \* 0.9709

= 1.3591

Information Gain: G(Conversion,Demand) = 1.57 – 1.35 = 0.22

Entropy(Conversion,Strategic)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | High | Medium | Low |  |
| yes | 3 | 3 | 3 | 9 |
| no | 2 | 2 | 3 | 7 |

= P(yes)\*Entropy(3,3,3) + P(no)\*Entropy(2,2,3)

= 9/16 \* Entropy(0.33,0.33,0.33) + 7/16 \* Entropy(0.2857,0.2857,0.4286)

= 9/16 \* [-0. 33 \* log2(0.33) – 0.33 \* log2(0.33) – 0.33 \* log2(0.33)] +

7/16 \* [-0.2857 \* log2(0.2857) – 0.2857 \* log2(0.2857) – 0.4286 \* log2(0.4286)]

= 0.8907 + 0.6810 = 1.5717

Information Gain: G(Conversion,Strategic) = 1.5794 – 1.5717 = 0.0077

Entropy(Conversion,Campaign)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | High | Medium | Low |  |
| aggressive | 4 | 4 | 1 | 9 |
| lowkey | 1 | 1 | 5 | 7 |

= P(aggressive)\*Entropy(4,4,1) + P(lowkey)\*Entropy(1,1,5)

= 9/16 \* Entropy(0.44,0.44,0.11) + 7/16 \* Entropy(0.1429,0.1429,0.7143)

= 9/16 \* [-0.44 \* log2(0.44) – 0.44 \* log2(0.44) – 0.11 \* log2(0.11)] +

7/16 \* [-0.1429 \* log2(0.1429) – 0.1429 \* log2(0.1429) – 0.7143 \* log2(0.7143)]

= 0.7833 + 0.5027 = 1.286

Information Gain: G(Conversion,Campaign) = 1.5794 – 1.286 = 0.293