**Assignment 1: Decision Tree Induction**

**Part I: Written Problem**

***1. Decision Tree Induction***

Necessary formulas:

* Entropy calculation: and
* Information gain if split by feature :

The number of instances according to worth taking

|  |  |
| --- | --- |
| Worth taking + | 5 |
| Worth taking - | 9 |

The Probability of the worth taking

|  |  |
| --- | --- |
|  |  |
|  |  |

Entropy of the worth taking

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |

The Probability of worth taking after divide to each type of features

|  |  |  |  |
| --- | --- | --- | --- |
| Personality | Hilarious |  |  |
|  |
| Boring |  |  |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Difficulty | Low |  |  |
|  |
| Medium |  |  |
|  |
| High |  |  |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| RMP Review | Awesome |  |  |
|  |
| Average |  |  |
|  |
| Awful |  |  |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| Easy A | Yes |  |  |
|  |
| No |  |  |
|  |

Of all the we can see the is highest so we choose the Difficulty as the Decision Tree root node

Medium

**Difficulty**

Low

High

Of all the instance with Difficulty = Low we have the following data table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Personality (Hilarious/Boring)** | **Difficulty (Low/Medium/ High)** | **RMP Reviews (Awesome/Average/Awful)** | **Easy A** | **Worth taking** |
| 1 | Hilarious | Low | Awesome | No | - |
| 3 | Boring | Low | Awesome | No | + |
| 4 | Hilarious | Low | Awesome | Yes | - |
| 5 | Hilarious | Low | Average | Yes | - |
| 12 | Boring | Low | Average | Yes | + |

We can see that only choosing Personality as the next node will immediately polarized the Worth taking result so the next feature for Difficulty = Low branch will be Easy A

**Difficulty**

Medium

Low

High

Personality

Hilarious

Boring

Of all the instances with Difficulty = Medium we have the following data table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Personality (Hilarious/Boring)** | **Difficulty (Low/Medium/ High)** | **RMP Reviews (Awesome/Average/Awful)** | **Easy A** | **Worth taking** |
| 2 | Boring | Medium | Awful | No | + |
| 6 | Hilarious | Medium | Awful | No | + |
| 8 | Hilarious | Medium | Awesome | No | + |
| 9 | Hilarious | Medium | Average | Yes | - |
| 11 | Boring | Medium | Awesome | Yes | - |

We can see that only choosing Easy A as the next node will immediately polarized the Worth taking result so the next feature for Difficulty = Medium branch will be Easy A

**Difficulty**

Medium

Low

High

Personality

Easy A

Hilarious

Boring

No

Yes

***2. Representing Boolean Functions***

(a).

**B**

**C**

**A**

**True**

**A**

**True**

**False**

**True**

**False**

True

True

True

True

False

False

False

False

(b).

**C**

**D**

**False**

**True**

**False**

True

True

False

False

False

**A**

**B**

**True**

**C**

**False**

**True**

**D**

**False**

True

True

True

True

False

False

False