Assignment 4

Rishabh Daga 2018101015

- Problem = (2018101015%10 + 1) = 6.
- Flip entries of (5 + 1) % 12 + 1 = 7 and 1 % 12 + 1 = 2.

Problem Statement :

You want to assess if you are living a healthy lifestyle or not but you can't make an objective decision about yourself on the topic. You observe habits of a few people around you and decide if that lifestyle is healthy or not. Based on your view of other people's lifestyle design a decision tree to see given your habits are you leading a healthy lifestyle or not.

Dataset:

Sleep Pattern	Junk Food Consumption	Exercise daily	Healthy
Irregular Sleep	High	Yes	No
Irregular Sleep	Normal	No	Yes
Irregular Sleep	Low	Yes	Yes
Irregular Sleep	Low	No	Yes
Good Sleep	High	Yes	Yes
Good Sleep	Normal	No	No
Good Sleep	Low	Yes	No
Good Sleep	High	No	No
Long Sleep	High	No	No
Long Sleep	Normal	Yes	Yes
Long Sleep	Low	Yes	Yes
Long Sleep	Normal	No	Yes

Steps

```
Attributes = [
    'Sleep Pattern (0)' : {'Irregular Sleep', 'Good Sleep', 'Long Sleep'},
    'Junk Food Consumption (1)' : {'High', 'Low', 'Normal'},
    'Exercise Daily (2)' : {'Yes', 'No'}
]
Label = 'Healthy' : {'Yes', 'No'}
I(p, n) = I(7, 5) = 0.979
```

Layer 1

```
E(0) = 4/12 * I(p1, n1) + 4/12 * I(p2, n2) + 4/12 * I(p3, n3)
     = 4/12 * I(3, 1) + 4/12 * I(1, 3) + 4/12 * I(3, 1)
     = 4/12 * 0.811 + 4/12 * 0.811 + 4/12 * 0.811
     = 0.811
=> G(0) = 0.979 - E(0) = 0.168

E(1) = 4/12 * I(p1, n1) + 4/12 * I(p2, n2) + 4/12 * I(p3, n3)
     = 4/12 * I(1, 3) + 4/12 * I(3, 1) + 4/12 * I(3, 1)
     = 4/12 * 0.811 + 4/12 * 0.811 + 4/12 * 0.811
     = 0.811
=> G(1) = 0.979 - E(1) = 0.168

E(2) = 6/12 * I(p1, n1) + 6/12 * I(p2, n2)
     = 6/12 * I(4, 2) + 6/12 * I(3, 3)
     = 6/12 * 0.918 + 6/12 * 1
     = 0.959
=> G(2) = 0.979 - E(2) = 0.020
```

As gain is equal for both 'Sleep Pattern' and 'Junk Food Consumption', we can choose any attribute as root of the tree. Here, 'Sleep Pattern' is chosen as the root of the tree.

Layer 2

> Irregular Sleep

```
I(p, n) = I(3, 1) = 0.811
E(1) = 1/4 * I(p1, n1) + 1/4 * I(p2, n2) + 2/4 * I(p3, n3)
     = 1/4 * I(0, 1) + 1/4 * I(1, 0) + 2/4 * I(2, 0)
     = 1/4 * 0 + 1/4 * 0 + 2/4 * 0
     = 0
=> G(1) = 0.811 - E(1) = 0.811

E(2) = 2/4 * I(p1, n1) + 2/4 * I(p2, n2)
     = 2/4 * I(1, 1) + 2/4 * I(2, 0)
     = 2/4 * 1 + 2/4 * 0
     = 0.5
=> G(2) = 0.811 - E(2) = 0.311
```

As gain is higher for ‘Junk Food Consumption', it is chosen as the next attribute.

> High

I(p, n) = I(0, 1) = 0. Hence, leaf already reached.

> Normal

I(p, n) = I(1, 0) = 0. Hence, leaf already reached.

> Low

I(p, n) = I(2, 0) = 0.Hence, leaf already reached.

> Good Sleep

```
 I(p, n) = I(1, 3) = 0.811
E(1) = 2/4 * I(p1, n1) + 1/4 * I(p2, n2) + 1/4 * I(p3, n3)
     = 2/4 * I(1, 1) + 1/4 * I(0, 1) + 1/4 * I(0, 1)
     = 2/4 * 1 + 1/4 * 0 + 1/4 * 0
     = 0.5
=> G(1) = 0.811 - E(1) = 0.311

E(2) = 2/4 * I(p1, n1) + 2/4 * I(p2, n2)
     = 2/4 * I(1, 1) + 2/4 * I(0, 2)
     = 2/4 * 1 + 2/4 * 0
     = 0.5
=> G(2) = 0.811 - E(2) = 0.311
```

As gain is equal for both 'Sleep Pattern' and 'Junk Food Consumption', we can choose any attribute as the next attribute. Here, 'Junk Food Consumption' is chosen as the next attribute.

> High

I(p, n) = I(1, 1) = 1. 'Exercise Daily' chosen (last option).

> Normal

I(p, n) = I(0, 1) = 0. Hence, leaf already reached.

> Low

I(p, n) = I(0, 1) = 0.Hence, leaf already reached.

> Long Sleep

```
 I(p, n) = I(3, 1) = 0.811
E(1) = 1/4 * I(p1, n1) + 2/4 * I(p2, n2) + 1/4 * I(p3, n3)
     = 1/4 * I(0, 1) + 2/4 * I(2, 0) + 1/4 * I(1, 0)
     = 1/4 * 0 + 2/4 * 0 + 1/4 * 0
     = 0
=> G(1) = 0.811 - E(1) = 0.811

E(2) = 2/4 * I(p1, n1) + 2/4 * I(p2, n2)
     = 2/4 * I(2, 0) + 2/4 * I(1, 1)
     = 2/4 * 0 + 2/4 * 1
     = 0.5
=> G(2) = 0.811 - E(2) = 0.311
```

As gain is higher for ‘Junk Food Consumption', it is chosen as the next attribute.

> High

I(p, n) = I(0, 1) = 0. Hence, leaf already reached.

> Normal

I(p, n) = I(2, 0) = 0. Hence, leaf already reached.

> Low

I(p, n) = I(1, 0) = 0. Hence, leaf already reached.