Handling Numerical Data

SNo	User_Rating	Downloded
1	3.5	Yes
2	4.6	Yes
3	2.2	No
4	1.6	Yes
5	4.1	No
6	3.9	No
7	3.2	No
8	2.9	Yes
9	4.8	Yes
10	3.5	No

(Before) DataSet_I

Q How to Split the DataSet_I?

Ans As, earlier we have Categorical Data. So It is easier to categorise the data. But, Now we have the Numerical Data. It is difficult to split as per the Numeric values...

Q To Split the data when there is Numerical Data, what should be the steps?

Ans Step 1:- Sort the DataSet as per the Numerical Column.

(After)

SNo	User_Rating	Downloded
1	1.6	Yes
2	2.2	No
3	2.9	Yes
4	3.2	No
5	3.3	No
6	3.5	Yes
7	3.9	No
8	4.1	No
9	4.6	Yes
10	4.8	Yes

DataSet_I

Spiral

Date.....

Step 2:- Split the entire data as per the individual user_rating.

For example:-

User_Rating >= 1.6

1. → User_Rating <= 1.6

Yes

SNo	User_Rating	Downloaded
1	1.6	Yes

Dataset_First

No

SNo	User_Rating	Downloaded
2	2.2	No
3	2.9	Yes
4	3.2	No
5	3.3	No
6	3.5	Yes
7	3.9	No
8	4.1	No
9	4.6	Yes
10	4.8	Yes

Dataset_Second

2. User_Rating <= 2.2

Yes

SNo	User_Rating	Downloaded
1	1.6	Yes
2	2.2	No

No → Rest Complete Table..

3. Splitting with all the values, till → User_Rating <= 4.8

Yes → Complete table

No

Date.....

Step 3:- For every Dataset we get after Step 2 From each User_Rating Value
For example:-

1. User_Rating <= 1.6
   - Yes → Dataset_First ⇒ E(D_First)
   - No → Dataset_Second ⇒ E(D_Second)

2. User_Rating <= 2.2
   - Yes → Dataset_First ⇒ E(D_First)
   - No → Dataset_Second ⇒ E(D_Second)

⋮

10. User_Rating <= 4.8
   - Yes → Dataset_First ⇒ E(D_First)
   - No → Dataset_Second ⇒ E(D_Second)

Step 4:- Calculate the Weighted Entropy of each of the Dataset.
For example:-

1. User_Rating <= 1.6
   - Yes → Dataset_First → E(D_First)
   - No → Dataset_Second → E(D_Second)
   - E(D_First), E(D_Second) → WE1

2. User_Rating <= 2.2
   - Yes → E(D_First)
   - No → E(D_Second)
   - E(D_First), E(D_Second) → WE2

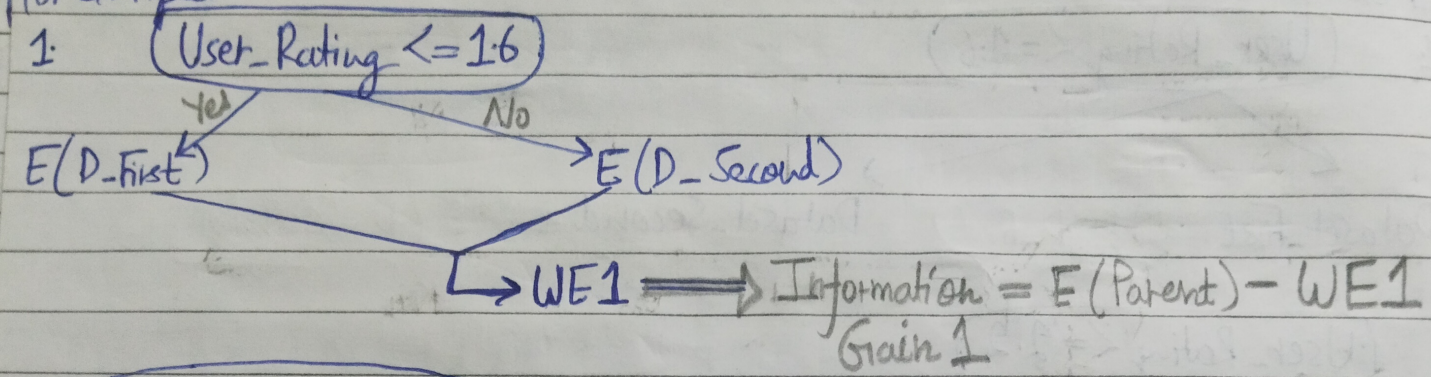
Spiral

Date.....

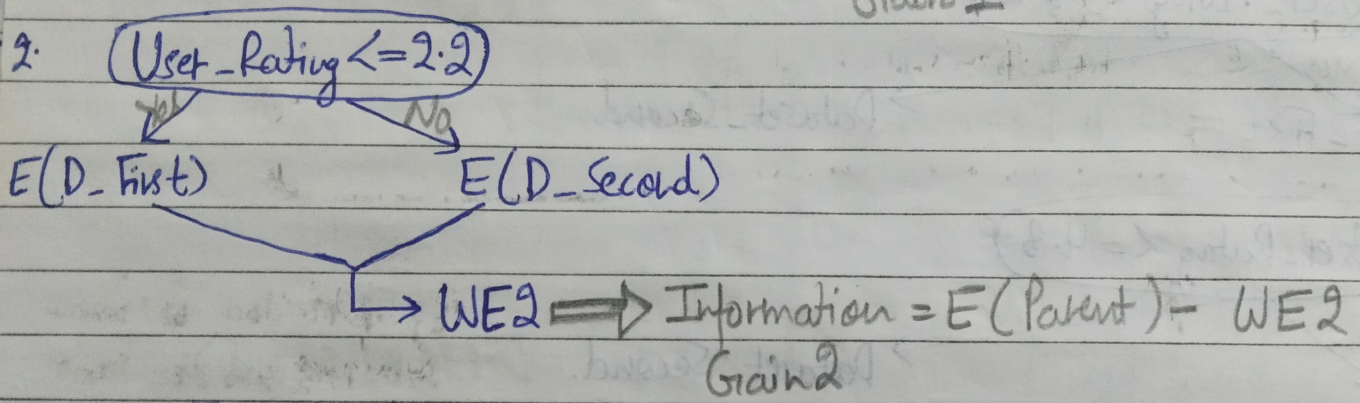
Step 5:- Calculate the Information Gain of every dataset.

For example:-

1. User_Rating <= 1.6



2. User_Rating <= 2.2

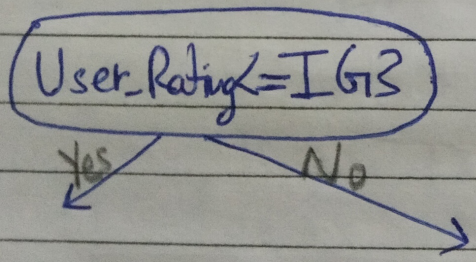


Step 6:- Find the Maximum Information Gain Value..

For example:-

$MAX\{IG1, IG2, IG3 \ldots IGn\}$

Assuming that "IG3" is the Maximum Value..



"Till you get all the Leaf Nodes."