Step 1: Calculate Entropy for the Root Node

To use the ID3 algorithm, we first need to calculate the entropy of the target variable "Play" for the entire dataset. Entropy measures the uncertainty or disorder of the dataset. Lower entropy indicates a purer node (more of one class).

Formula for Entropy:

Entropy

(

𝑆

)

=

−

∑

𝑝

𝑖

log

⁡

2

(

𝑝

𝑖

)

Entropy(S)=−∑p

i

​

log

2

​

(p

i

​

)

Where

𝑝

𝑖

p

i

​

is the probability of class

𝑖

i in the dataset.

For the "Play" variable, we have:

5 "No" instances

7 "Yes" instances Total = 12 instances.

The probability of "No" (

𝑝

No

p

No

​

) =

5

12

12

5

​

The probability of "Yes" (

𝑝

Yes

p

Yes

​

) =

7

12

12

7

​

Entropy of the root node:

Entropy

(

𝑆

)

=

−

(

5

12

)

log

⁡

2

(

5

12

)

−

(

7

12

)

log

⁡

2

(

7

12

)

Entropy(S)=−(

12

5

​

)log

2

​

(

12

5

​

)−(

12

7

​

)log

2

​

(

12

7

​

)

Let's calculate this value.

Entropy

(

𝑆

)

=

−

(

0.4167

×

log

⁡

2

(

0.4167

)

)

−

(

0.5833

×

log

⁡

2

(

0.5833

)

)

Entropy(S)=−(0.4167×log

2

​

(0.4167))−(0.5833×log

2

​

(0.5833))

We can now calculate these logarithmic values.

Entropy

(

𝑆

)

=

−

(

0.4167

×

−

1.2637

)

−

(

0.5833

×

−

0.7859

)

Entropy(S)=−(0.4167×−1.2637)−(0.5833×−0.7859)

Entropy

(

𝑆

)

=

0.5264

+

0.4575

=

0.9839

Entropy(S)=0.5264+0.4575=0.9839

So, the entropy of the root node is 0.9839.

Step 2: Calculate Information Gain for Each Feature

Now that we know the entropy at the root node, we will calculate the information gain for each feature (Whether, Temperature, Wind) to determine which feature provides the best split. The feature with the highest information gain will be chosen for the first split.

Information Gain Formula:

Information Gain

(

𝑆

,

𝐴

)

=

Entropy

(

𝑆

)

−

∑

(

∣

𝑆

𝑣

∣

∣

𝑆

∣

×

Entropy

(

𝑆

𝑣

)

)

Information Gain(S,A)=Entropy(S)−∑(

∣S∣

∣S

v

​

∣

​

×Entropy(S

v

​

))

Where:

𝑆

𝑣

S

v

​

is the subset of the dataset for which the feature

𝐴

A takes a specific value.

∣

𝑆

𝑣

∣

∣

𝑆

∣

∣S∣

∣S

v

​

∣

​

is the proportion of instances in that subset.

a) Information Gain for the "Whether" Feature:

Let's split the dataset based on "Whether," which has three values: Sunny, Rain, Cloudy.

Subset 1: "Sunny"

Instances: 1, 4, 5, 9, 12 (5 instances)

Yes: 2, No: 3

Entropy of "Sunny":

Entropy

(

𝑆

Sunny

)

=

−

(

2

5

×

log

⁡

2

(

2

5

)

)

−

(

3

5

×

log

⁡

2

(

3

5

)

)

Entropy(S

Sunny

​

)=−(

5

2

​

×log

2

​

(

5

2

​

))−(

5

3

​

×log

2

​

(

5

3

​

))

Calculating it:

Entropy

(

𝑆

Sunny

)

=

0.971

Entropy(S

Sunny

​

)=0.971

Subset 2: "Rain"

Instances: 2, 7, 8, 11 (4 instances)

Yes: 3, No: 1

Entropy of "Rain":

Entropy

(

𝑆

Rain

)

=

−

(

3

4

×

log

⁡

2

(

3

4

)

)

−

(

1

4

×

log

⁡

2

(

1

4

)

)

Entropy(S

Rain

​

)=−(

4

3

​

×log

2

​

(

4

3

​

))−(

4

1

​

×log

2

​

(

4

1

​

))

Calculating it:

Entropy

(

𝑆

Rain

)

=

0.811

Entropy(S

Rain

​

)=0.811

Subset 3: "Cloudy"

Instances: 3, 6, 10 (3 instances)

Yes: 2, No: 1

Entropy of "Cloudy":

Entropy

(

𝑆

Cloudy

)

=

−

(

2

3

×

log

⁡

2

(

2

3

)

)

−

(

1

3

×

log

⁡

2

(

1

3

)

)

Entropy(S

Cloudy

​

)=−(

3

2

​

×log

2

​

(

3

2

​

))−(

3

1

​

×log

2

​

(

3

1

​

))

Calculating it:

Entropy

(

𝑆

Cloudy

)

=

0.918

Entropy(S

Cloudy

​

)=0.918

Total Weighted Entropy for "Whether":

Entropy

(

𝑆

Whether

)

=

(

5

12

×

0.971

)

+

(

4

12

×

0.811

)

+

(

3

12

×

0.918

)

Entropy(S

Whether

​

)=(

12

5

​

×0.971)+(

12

4

​

×0.811)+(

12

3

​

×0.918)

Entropy

(

𝑆

Whether

)

=

0.4046

+

0.2703

+

0.2295

=

0.9044

Entropy(S

Whether

​

)=0.4046+0.2703+0.2295=0.9044

Information Gain for "Whether":

Information Gain

(

𝑆

,

Whether

)

=

0.9839

−

0.9044

=

0.0795

Information Gain(S,Whether)=0.9839−0.9044=0.0795

b) Information Gain for the "Temperature" Feature:

Temperature has three values: High, Normal, Low. We repeat the same process as above.

Subset 1: "High"

Instances: 1, 3, 4, 11 (4 instances)

Yes: 2, No: 2

Entropy of "High" = 1 (since it's an equal mix of Yes and No).

Subset 2: "Normal"

Instances: 2, 7, 9, 12 (4 instances)

Yes: 2, No: 2

Entropy of "Normal" = 1.

Subset 3: "Low"

Instances: 5, 6, 8, 10 (4 instances)

Yes: 3, No: 1

Entropy of "Low" = 0.811.

Total Weighted Entropy for "Temperature":

Entropy

(

𝑆

Temperature

)

=

(

4

12

×

1

)

+

(

4

12

×

1

)

+

(

4

12

×

0.811

)

Entropy(S

Temperature

​

)=(

12

4

​

×1)+(

12

4

​

×1)+(

12

4

​

×0.811)

Entropy

(

𝑆

Temperature

)

=

0.333

+

0.333

+

0.2703

=

0.9363

Entropy(S

Temperature

​

)=0.333+0.333+0.2703=0.9363

Information Gain for "Temperature":

Information Gain

(

𝑆

,

Temperature

)

=

0.9839

−

0.9363

=

0.0476

Information Gain(S,Temperature)=0.9839−0.9363=0.0476

c) Information Gain for the "Wind" Feature:

Wind has two values: High and Normal.

Subset 1: "High"

Instances: 1, 2, 4, 7, 9, 11 (6 instances)

Yes: 3, No: 3

Entropy of "High" = 1.

Subset 2: "Normal"

Instances: 3, 5, 6, 8, 10, 12 (6 instances)

Yes: 4, No: 2

Entropy of "Normal":

Entropy

(

𝑆

Normal

)

=

−

(

4

6

×

log

⁡

2

(

4

6

)

)

−

(

2

6

×

log

⁡

2

(

2

6

)

)

Entropy(S

Normal

​

)=−(

6

4

​

×log

2

​

(

6

4

​

))−(

6

2

​

×log

2

​

(

6

2

​

))

Calculating:

Entropy

(

𝑆

Normal

)

=

0.918

Entropy(S

Normal

​

)=0.918

Total Weighted Entropy for "Wind":

Entropy

(

𝑆

Wind

)

=

(

6

12

×

1

)

+

(

6

12

×

0.918

)

Entropy(S

Wind

​

)=(

12

6

​

×1)+(

12

6

​

×0.918)

Entropy

(

𝑆

Wind

)

=

0.5

+

0.459

=

0.959

Entropy(S

Wind

​

)=0.5+0.459=0.959

Information Gain for "Wind":

Information Gain

(

𝑆

,

Wind

)

=

0.9839

−

0.959

=

0.0249

Information Gain(S,Wind)=0.9839−0.959=0.0249

Step 3: Choose the Feature with the Highest Information Gain

Information Gain for "Whether" = 0.0795

Information Gain for "Temperature" = 0.0476

Information Gain for "Wind" = 0.0249

Since "Whether" has the highest information gain of 0.0795, it will be chosen as the first split.

Next Steps:

Split the dataset based on the "Whether" feature.

Recursively apply the same process to the subsets created by this split, repeating the entropy and information gain calculation for each subset until all leaves are pure or no further splits can be made.

This is how the ID3 algorithm works using information gain as the criterion, focusing on reducing entropy to build the decision tree.