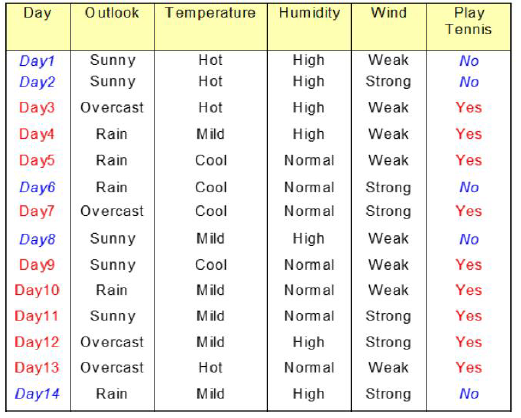
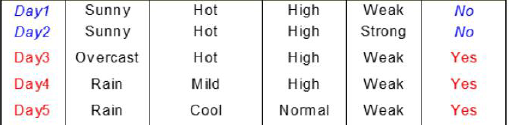
Table 1: Original Dataset



1. Bootstrapped dataset creation

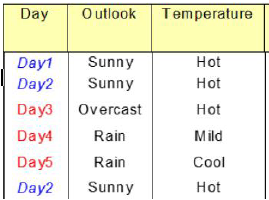
Table 1: Bootstrapped Dataset 1





**2. Create decision trees using random subset of variables or columns [ Here, we considered only 2 columns randomly]**

Table 2: Random subset of variables or columns from Bootstrapped dataset (1)



Outlook

Sunny [Yes: 0, No: 3]

Overcast [Yes: 1, No: 0]

Rain [Yes: 2, No: 0]

GINI(Outlook=sunny) = 1 - (0/3)^2-(3/3)^2 = 1 - 0 - 1 = 0

GINI(Outlook= Overcast) = 1 - (1/1)^2-(0/1)^2 = 1 - 1 - 0 = 0

GINI(Outlook= Rain) = 1 - (2/2)^2-(0/2)^2 = 1 - 1 - 0 = 0

Now,

Gini impurity of parent node = weighted average of Gini impurities of leaf nodes

**GINI(Outlook)** = (3/6)\*0 + (1/6)\*0 + (2/6)\*0 = 0

Temperature

Hot [Yes: 1, No: 3]

Mild [Yes: 1, No: 0]

cool [Yes: 1, No: 0]

GINI(Temperature=Hot)=1 -(1/4)^2-(3/4)^2= 1 - 0.0625 - 0.5625 = 0.375

GINI(Temperature=Mild) = 1 - (1/1)^2-(0/1)^2 = 1 - 1 - 0 = 0

GINI(Temperature=Cool) = 1 - (1/1)^2-(0/1)^2 = 1 - 1 - 0 = 0

**GINI(Temperature)** = (4/6)\* 0.375 + (1/6)\*0 + (1/6)\*0 = 0.25

The lowest impurity means, the feature with lowest impurity separates the classes well.

As GINI(Outlook) < GINI(Temperature), so Outlook will be in the root of our decision tree.

Sunny

Overcast

Rain

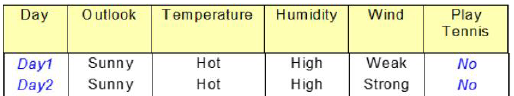
Now, we should consider for next level nodes for better separation.

Sunny

Overcast

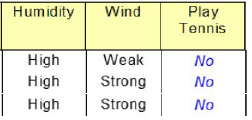
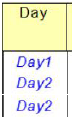
Rain

Select this from bootstrapped dataset, where Outlook=Sunny



Consider the bootstrapped dataset again and take random subset of variables or columns excluding the root node variable. Also only consider the instances, which have specific root/parent variable (e.g., Outlook) value (e.g., Sunny).

Table 3: Random subset of variables or columns from Bootstrapped dataset (2)



Humidity

High [Yes: 0, No: 3]

GINI(Humidity= High) = 1 - (0/3)^2-(3/3)^2 = 1 - 0 - 1 = 0

Now,

Gini impurity of parent node = weighted average of Gini impurities of leaf nodes

**GINI(**Humidity**)** = (3/3)\*0 = 0

Wind

Strong [Yes: 0, No: 2]

Weak [Yes: 0, No: 1]

GINI(Wind = Strong)=1 -(0/2)^2-(2/2)^2= 1 - 0 - 1 = 0

GINI(Wind = Weak) = 1 - (0/1)^2-(1/1)^2 = 1 - 0 - 1 = 0

**GINI(**Wind**)** = (2/3)\* 0 + (1/3)\*0 = 0

As GINI(Wind) = GINI(Humidity), so Wind or Humidity will be the level 2 factor of our decision tree.

Sunny

Overcast

Rain

Strong

Weak

No

No



Sunny

Overcast

Rain

Strong

Weak

No

No

Yes



Table 4: Random subset of variables or columns from Bootstrapped dataset (3)

Sunny

Overcast

Rain

Strong

Weak

No

No

Yes

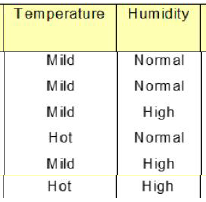
Yes

Table 1: Bootstrapped Dataset 2





**2. Create decision trees using random subset of variables or columns [ Here, we considered only 2 columns randomly]**



Temperature

Mild [Yes: 3, No: 1]

Hot [Yes: 1, No: 1]

GINI(Temperature=Mild)=1-(3/4)^2-(1/4)^2=1 - 0.5625 - 0.0625 = 0.375

GINI(Temperature = Hot) = 1 - (1/2)^2-(1/2)^2 = 0.5

Now,

Gini impurity of parent node = weighted average of Gini impurities of leaf nodes

**GINI(**Temperature**)** = (4/6)\*0.375 + (2/6)\*0.5 = 0.417

Humidity

High [Yes: 1, No: 2]

Normal [Yes: 3, No: 0]

GINI(Humidity = High) = 1 -(1/3)^2-(2/3)^2= 1 - 0.1111 - 0.4444 = 0.444

GINI(Humidity = Normal) = 1 - (3/3)^2-(0/3)^2 = 1 - 1 - 0 = 0

**GINI(**Humidity**)** = (3/6)\* 0.444 + (3/6)\*0 = 0.22223

Normal

High

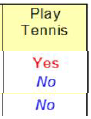
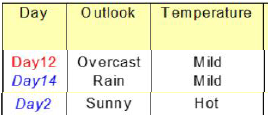
Now, we should consider for next level nodes for better separation.

High

Normal







Temperature

Mild [Yes: 1, No: 1]

Hot [Yes: 0, No: 1]

GINI(Temperature=Mild)=1-(1/2)^2-(1/2)^2= 0.5

GINI(Temperature = Hot) = 1 - (0/1)^2-(1/1)^2 = 1-0-1=0

Now,

Gini impurity of parent node = weighted average of Gini impurities of leaf nodes

**GINI(**Temperature**)** = (2/3)\*0.5 + (1/3)\*0 = 0.333

Outlook

Sunny [Yes: 0, No: 1]

Overcast [Yes: 1, No: 0]

Rain [Yes: 0, No: 1]

GINI(Outlook=sunny) = 0

GINI(Outlook= Overcast) = 0

GINI(Outlook= Rain) = 0

Now,

Gini impurity of parent node = weighted average of Gini impurities of leaf nodes

**GINI(Outlook)** = (1/3)\*0 + (1/3)\*0 + (1/3)\*0 = 0

High

Normal

Sunny

Overcast

Rain

No

No

Yes





High

Normal

Sunny

Overcast

Rain

No

No

Yes

Yes

NOW, A Query:



Sunny

Overcast

Rain

Strong

Weak

No

No

Yes

Yes

Tree-1

Bagging = Yes: 1

High

Normal

Sunny

Overcast

Rain

No

No

Yes

Yes

Tree-2

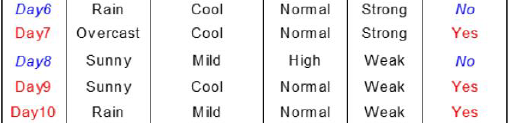
Bagging = Yes: 2

If Tree 3 result is NO.

Then Bagging: Yes: 2, No: 1

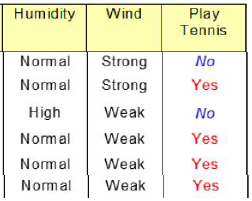
So, Final result of the query is YES

Table 1: Bootstrapped Dataset 3





**2. Create decision trees using random subset of variables or columns [ Here, we considered only 2 columns randomly]**



Humidity

High [Yes: 1, No: 2]

Normal [Yes: 3, No: 0]

GINI(Humidity = High) = 1 -(1/3)^2-(2/3)^2= 1 - 0.1111 - 0.4444 = 0.444

GINI(Humidity = Normal) = 1 - (3/3)^2-(0/3)^2 = 1 - 1 - 0 = 0

**GINI(**Humidity**)** = (3/6)\* 0.444 + (3/6)\*0 = 0.22223

Wind

Strong [Yes: 0, No: 2]

Weak [Yes: 0, No: 1]

GINI(Wind = Strong)=1 -(0/2)^2-(2/2)^2= 1 - 0 - 1 = 0

GINI(Wind = Weak) = 1 - (0/1)^2-(1/1)^2 = 1 - 0 - 1 = 0

**GINI(**Wind**)** = (2/3)\* 0 + (1/3)\*0 = 0

As GINI(Wind) = GINI(Humidity), so Wind or Humidity will be the level 2 factor of our decision tree.