# Decision Tree

# Agenda

**After this session, you will know:**

- The Algorithm used in Decision Tree
- Decision Tree - Classifier
- Entropy
- Information Gain
- Decision Tree – Regression
- Gini Impurity
- Random Forest

# The Algorithm behind Decision Tree

# Iterative Dichotomizer 3 (ID3)

Iterative Dichotomizer 3 (ID3) algorithm is the core algorithm that is used for Decision Tree.

ID3 employs a top-down, greedy search through the space of possible branches with no backtracking.

ID3 uses followings to construct a Decision Tree:

**01**

Entropy

**02**

Information Gain

# C4.5 Algorithm

1. C4.5 algorithm has made some significant improvements over ID3.
2. Handling both continuous and discrete attributes.
3. In order to handle continuous variables, C4.5 creates thresholds and splits the lists as per the threshold.
4. C4.5 allows missing values to be marked as missing.
5. The algorithm simply do not consider the missing values for the calculation of entropy and information gain.
6. C4.5 goes back through the tree once its been created.
7. The algorithm simply do not consider the missing values for the calculation of entropy and information gain.

# C5.0 Algorithm

## Advantages of C5.0 Algorithms

**Speed**

Significantly faster than C4.5

**Memory Usage**

More memory efficient

**Weighting**

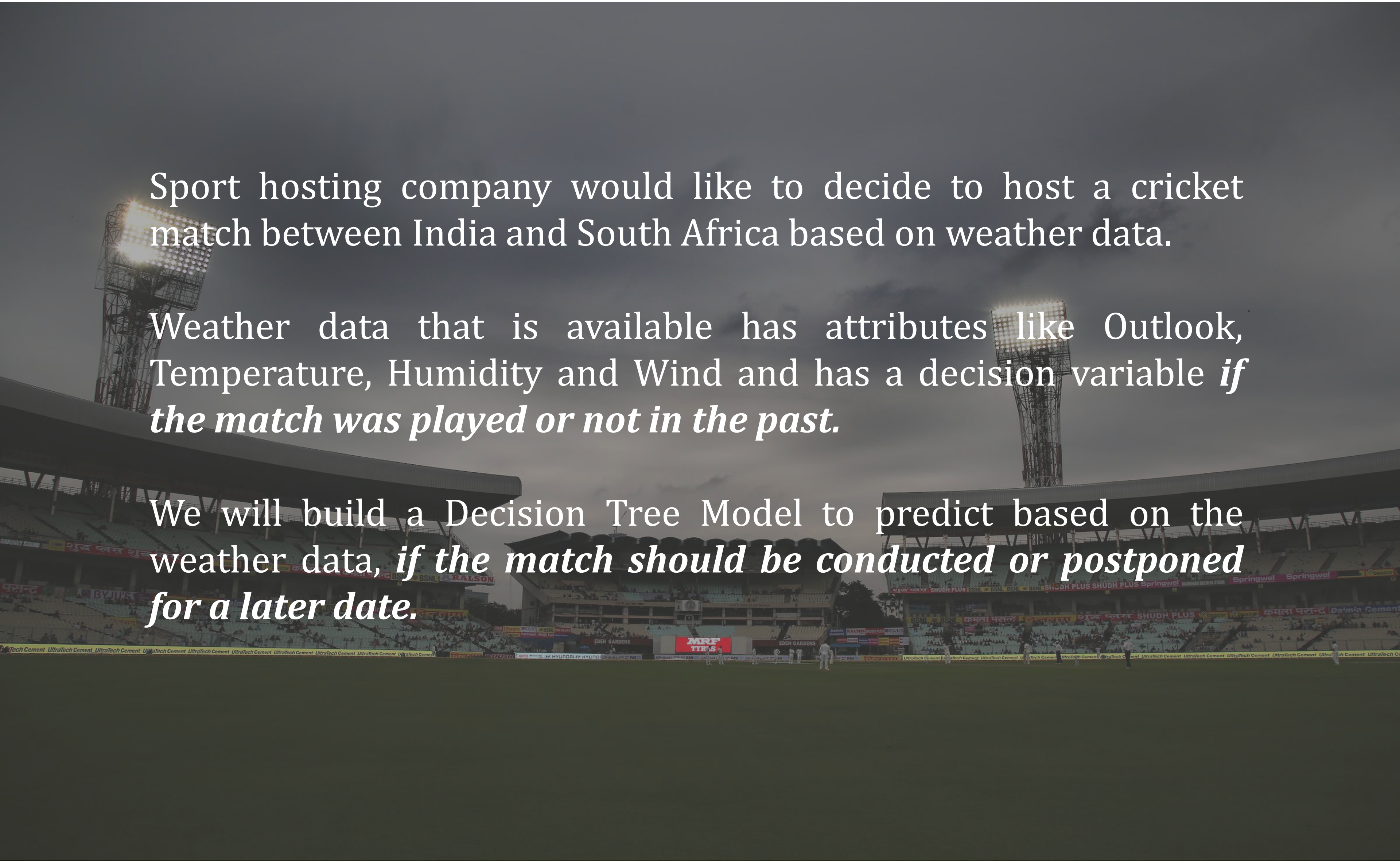
Allows to weight difference cases and misclassification types

**Winnowing**

Automatically winnows the attributes to remove that may not be helpful

# Decision Tree - Classification

# Business Problem | To decide to Play or not Play



Sport hosting company would like to decide to host a cricket match between India and South Africa based on weather data.

Weather data that is available has attributes like Outlook, Temperature, Humidity and Wind and has a decision variable ***if the match was played or not in the past.***

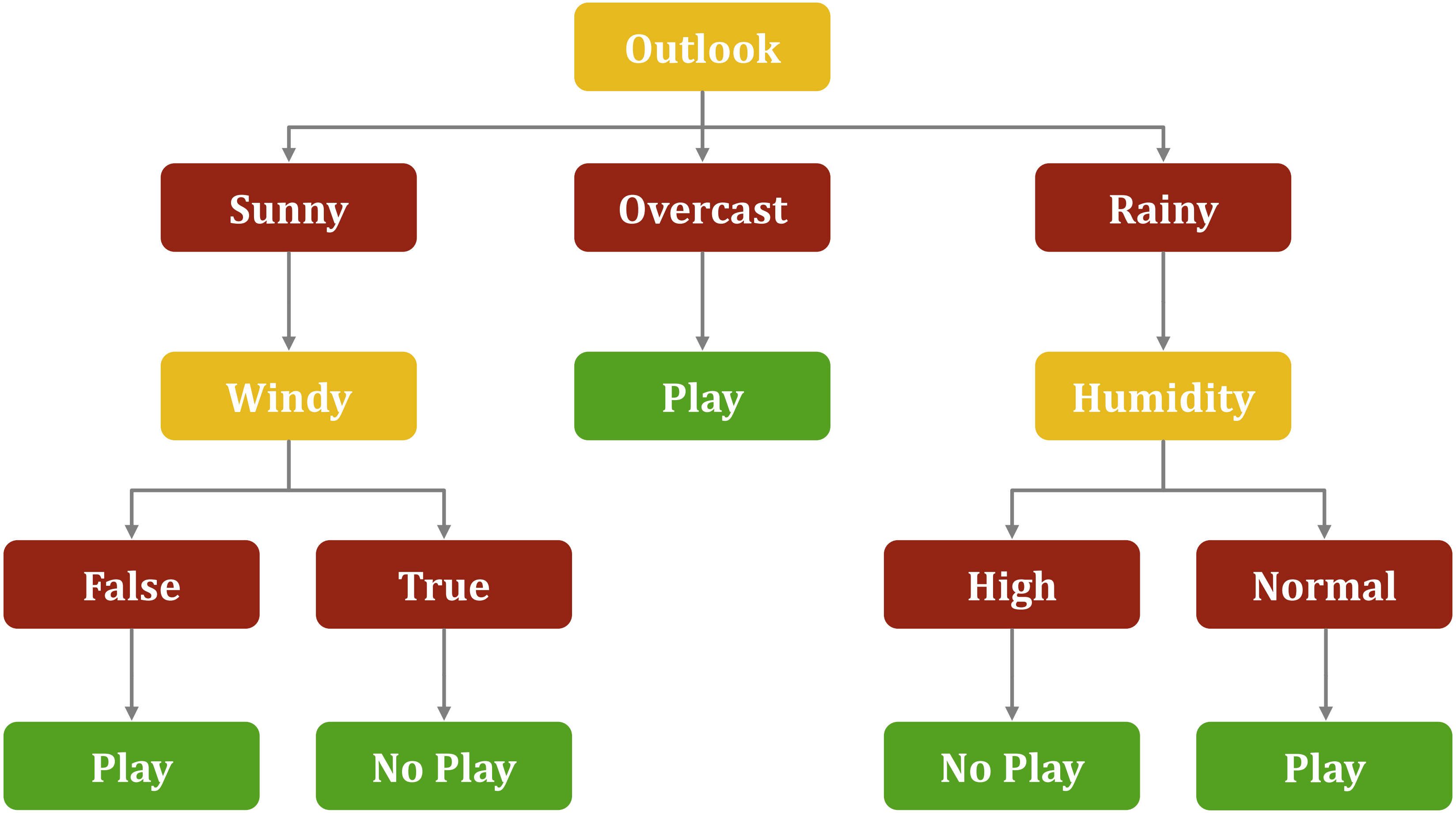
We will build a Decision Tree Model to predict based on the weather data, ***if the match should be conducted or postponed for a later date.***

# Weather Data

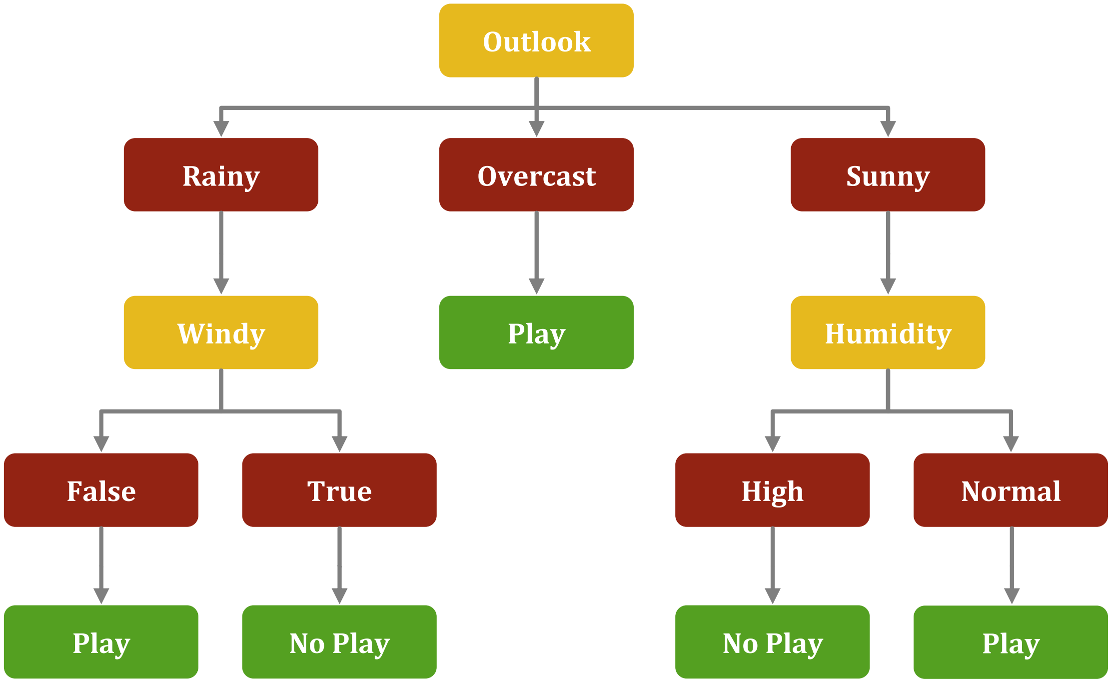
| Outlook | Temperature | Humidity | Windy | Play |
|---|---|---|---|---|
| sunny | hot | high | false | no |
| sunny | hot | high | true | no |
| overcast | hot | high | false | yes |
| rainy | mild | high | false | yes |
| rainy | cool | normal | false | yes |
| rainy | cool | normal | true | no |
| overcast | cool | normal | true | yes |
| sunny | mild | high | false | no |
| sunny | cool | normal | false | yes |
| rainy | mild | normal | false | yes |
| sunny | mild | normal | true | yes |
| overcast | mild | high | true | yes |
| overcast | hot | normal | false | yes |
| rainy | mild | high | true | no |

# Decision Tree Output (Classification)

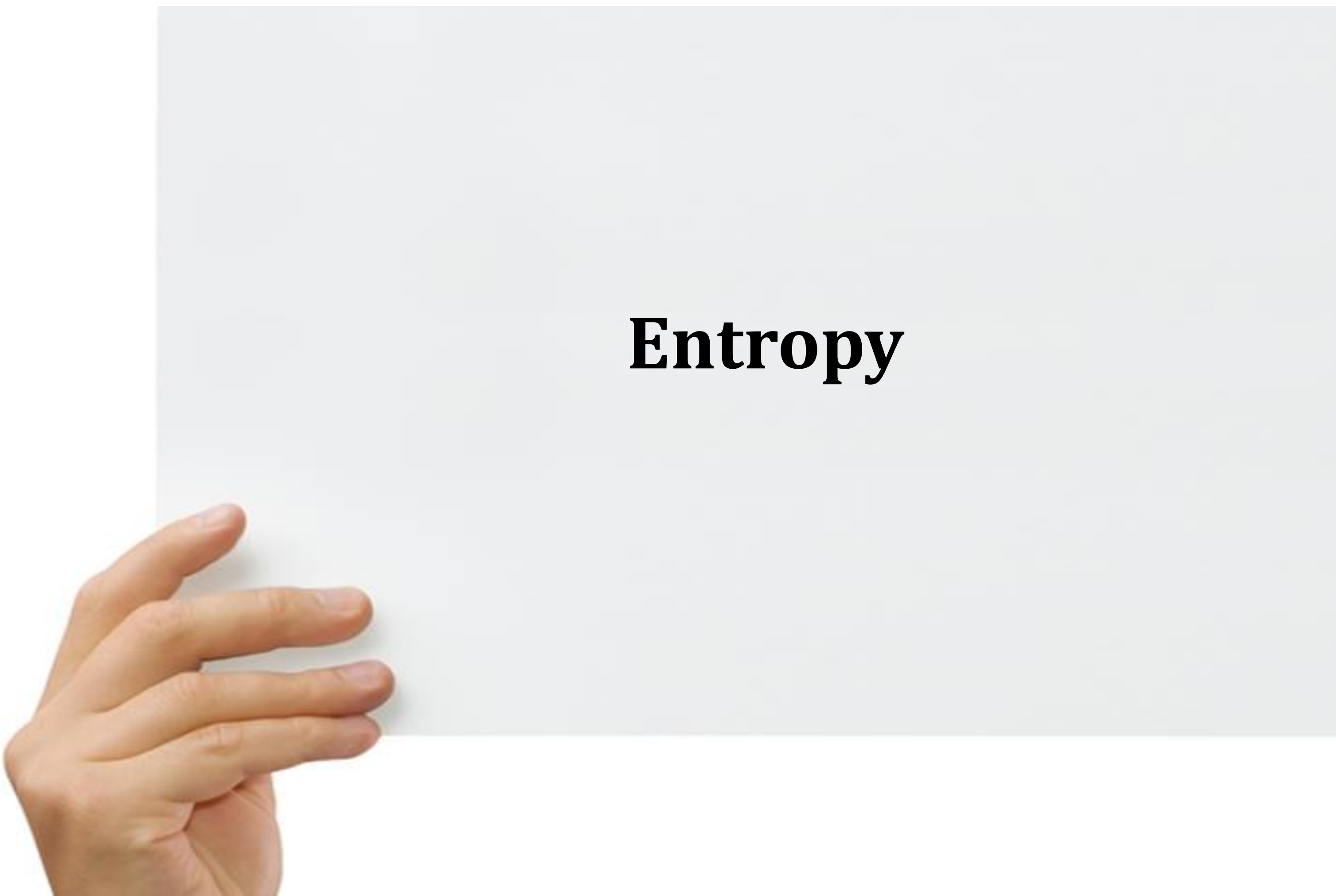
**Decision Tree Model to predict the weather data**



# Characteristics




- The starting node is called root node
- Every non-leaf node denotes a representation of the attribute value
- Every branch denotes the rest of the value representation
- Every leaf or the terminal node represents the value of the target attribute

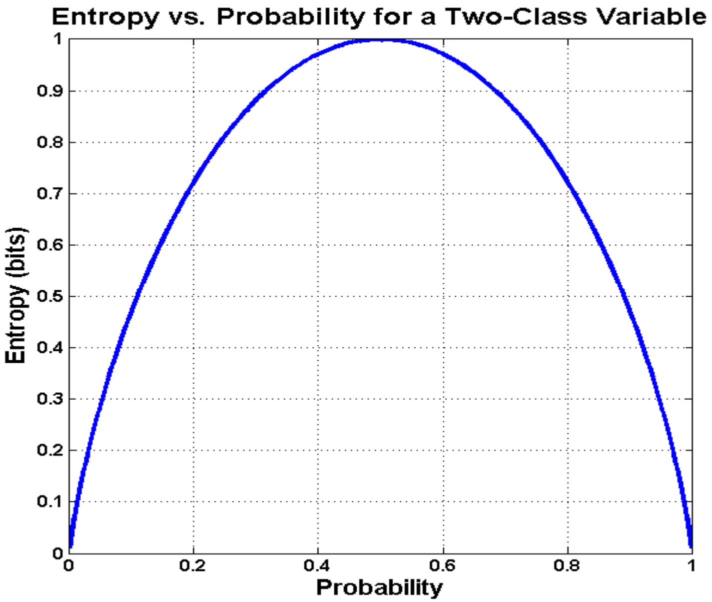

Entropy

# Entropy

Iterative dichotomizer (ID3) algorithm uses entropy to calculate the homogeneity of a sample.

If the sample is completely homogeneous, the Entropy = 0

If the sample is equally divided, the Entropy = 1




Entropy $= -p \log_2 p - q \log_2 q$

Entropy $= -0.5 \log_2 0.5 - 0.5 \log_2 0.5 = 1$

$$H = -\sum_i p_i(\log_2 p_i)$$

# Types of Entropy

There are two types of Entropy

**Entropy of the Target Variable (Categorical)**

**01**

**Entropy of 2 Variables**

**02**

Each of the independent variable & the target variable

# Entropy of Categorical Target Variable

**Formula** $$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

| Play Match? | |
|---|---|
| Yes | No |
| 9 | 5 |

Entropy (Play Match)

= Entropy (5, 9)

= Entropy (0.36, 0.64)

= – (0.36 $\log_2$ 0.36) – (0.64 $\log_2$ 0.64)

= 0.94

# Entropy of 2 Variables

$$E(T,X)=\sum_{c\in X}P(c)E(c)$$

| | | Play Match | | |
|---|---|---|---|---|
| | | Yes | No | |
| Outlook | Rainy | 3 | 2 | 5 |
| | Overcast | 4 | 0 | 4 |
| | Sunny | 2 | 3 | 5 |
| | | | | 14 |

Entropy (Play Match, Outlook)

= P(Rainy)*E(3,2) + P(Overcast)*E(4,0) + P(Sunny)*E(2,3)

= (5/14)*(0.971) + (4/14)(0.0) + (5/14)(0.971)

= 0.693

# Information Gain

# Information Gain

Information Gain is based on the decrease in entropy after a dataset is split on an attribute.

$$Gain\ (T, X) = Entropy\ before\ split - Entropy\ after\ split$$

$$Gain\ (T, X) = Entropy\ (T) - Entropy\ (T, X)$$

Constructing Decision Tree is all about finding attribute that returns the highest Information Gain.

# Entropy of the Target Variable – Step 1

| Play |
| --- |
| No |
| No |
| Yes |
| Yes |
| Yes |
| No |
| Yes |
| No |
| Yes |
| Yes |
| Yes |
| Yes |
| Yes |
| No |

$$\text{Entropy (T)} = \text{Entropy (5,9)}$$

$$= \text{Entropy}(0.36, 0.64)$$

$$= -(0.36\log_2 0.36) - (0.64\log_2 0.64)$$

$$= 0.94$$

# Calculate Information Gain – Step 2

- The entropy for each branch is calculated
- Then it is added proportionally, to get total entropy for the split
- The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain, or decrease in entropy

$$Gain(T,X) = Entropy(T) - Entropy(T,X)$$

Gain(Play,Outlook) = E(Play) – E(Play,Outlook)
= 0.940 – 0.693
= 0.247

# How Does It Select the Root Node? – Step 2(contd)

| | | Play Match | | |
|---|---|---|---|---|
| | | Yes | No | |
| Outlook | Rainy | 3 | 2 | 5 |
| | Overcast | 4 | 0 | 4 |
| | Sunny | 2 | 3 | 5 |
| | | | | 14 |

**Gain = 0.247**

| | | Play Match | | |
|---|---|---|---|---|
| | | Yes | No | |
| Temp | Hot | 2 | 2 | 4 |
| | Mild | 4 | 2 | 6 |
| | Cold | 3 | 1 | 6 |
| | | | | 14 |

**Gain = 0.029**

| | | Play Match | | |
|---|---|---|---|---|
| | | Yes | No | |
| Windy | False | 6 | 2 | 8 |
| | True | 3 | 3 | 6 |
| | | | | 14 |

**Gain = 0.048**

| | | Play Match | | |
|---|---|---|---|---|
| | | Yes | No | |
| Humidity | High | 3 | 4 | 7 |
| | Normal | 6 | 1 | 7 |
| | | | | 14 |

**Gain = 0.152**

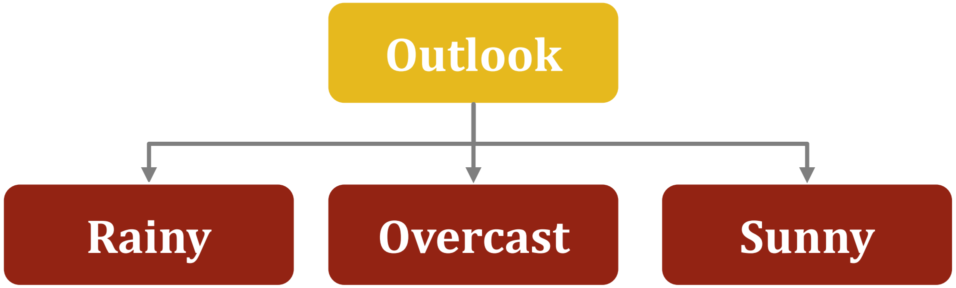
Highest Information Gain. So we choose Outlook as the Root Node

# Compare Information Gain – Step 3

Choose attribute with the largest information gain as the decision node, divide the dataset by its branches and repeat the same process on every branch

| | | Play Match | |
|---|---|---|---|
| | | Yes | No |
| Outlook | Rainy | 3 | 2 |
| | Overcast | 4 | 0 |
| | Sunny | 2 | 3 |
| | | | |

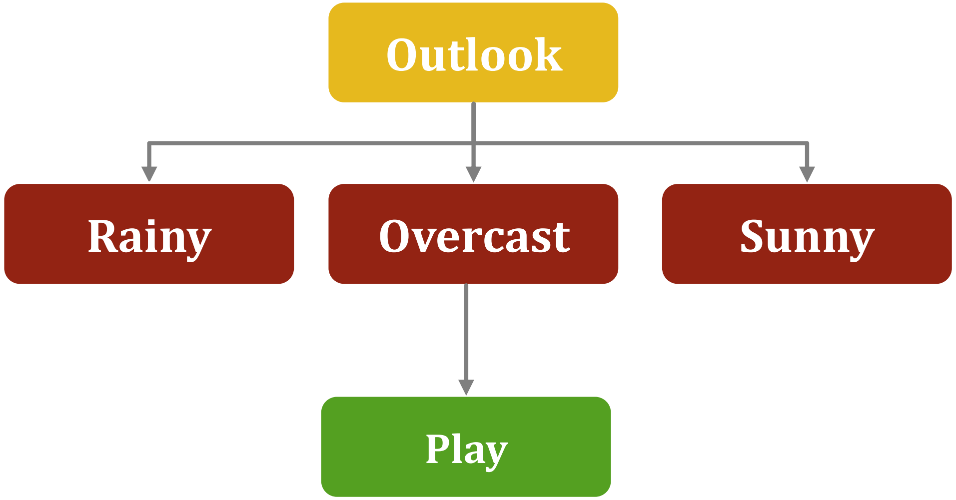
**Gain = 0.247**



# How does it Select a Leaf Node? – Step 3a

- A branch with entropy of 0 is a **leaf node**

| Temp | Humidity | Windy | Play |
|---|---|---|---|
| Hot | High | FALSE | Yes |
| Cool | Normal | TRUE | Yes |
| Mild | High | TRUE | Yes |
| Hot | Normal | FALSE | Yes |

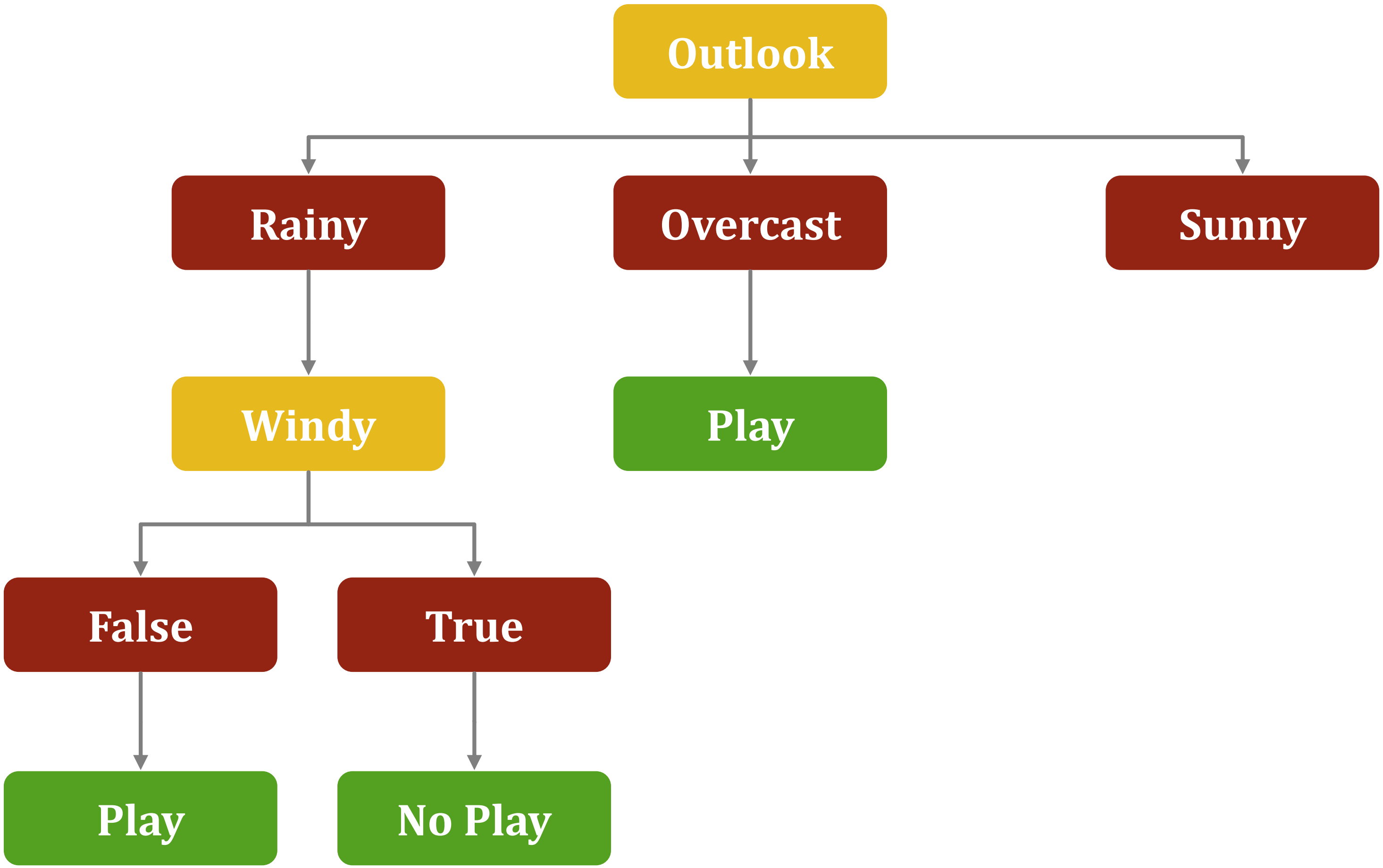


# How does it Select the next Node? – Step 3b

- A branch with entropy more than 0 needs further splitting

| Temp | Humidity | Windy | Play |
|---|---|---|---|
| Mild | High | FALSE | Yes |
| Cool | Normal | FALSE | Yes |
| Mild | High | FALSE | Yes |
| Cool | Normal | TRUE | No |
| Mild | High | TRUE | No |

# Outlook – Rainy – Step 3b (contd.)



# Decision Tree Output (Classification)

**Decision Tree Model to predict the weather data**