

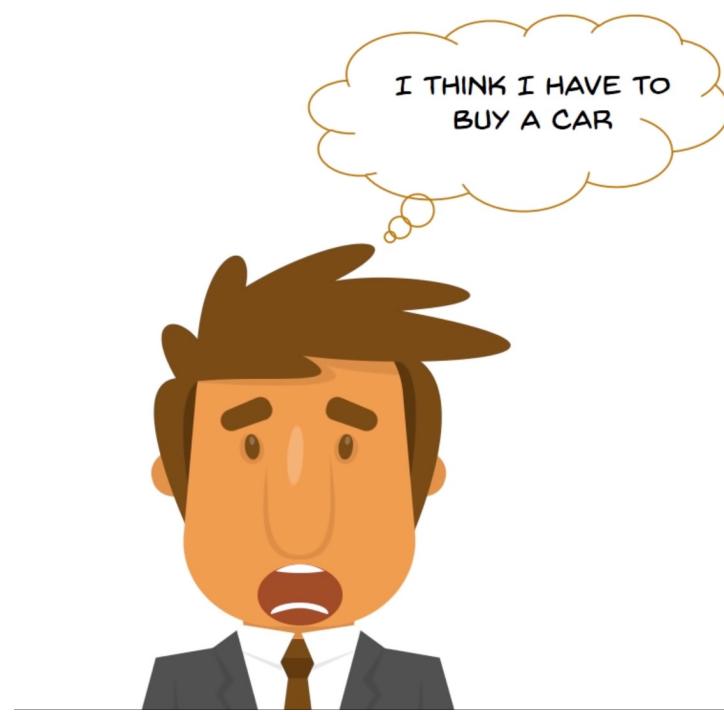


Machine Learning

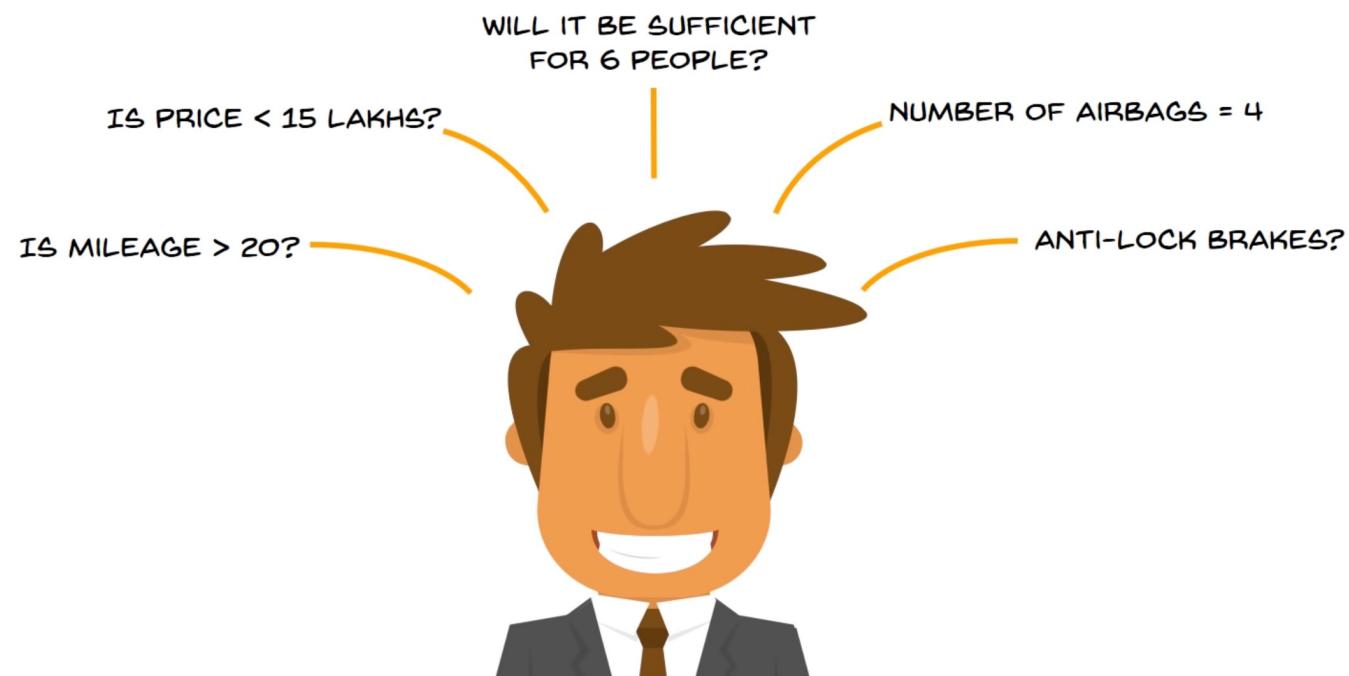
Decision Trees

Phd. César Astudillo | Facultad de Ingeniería

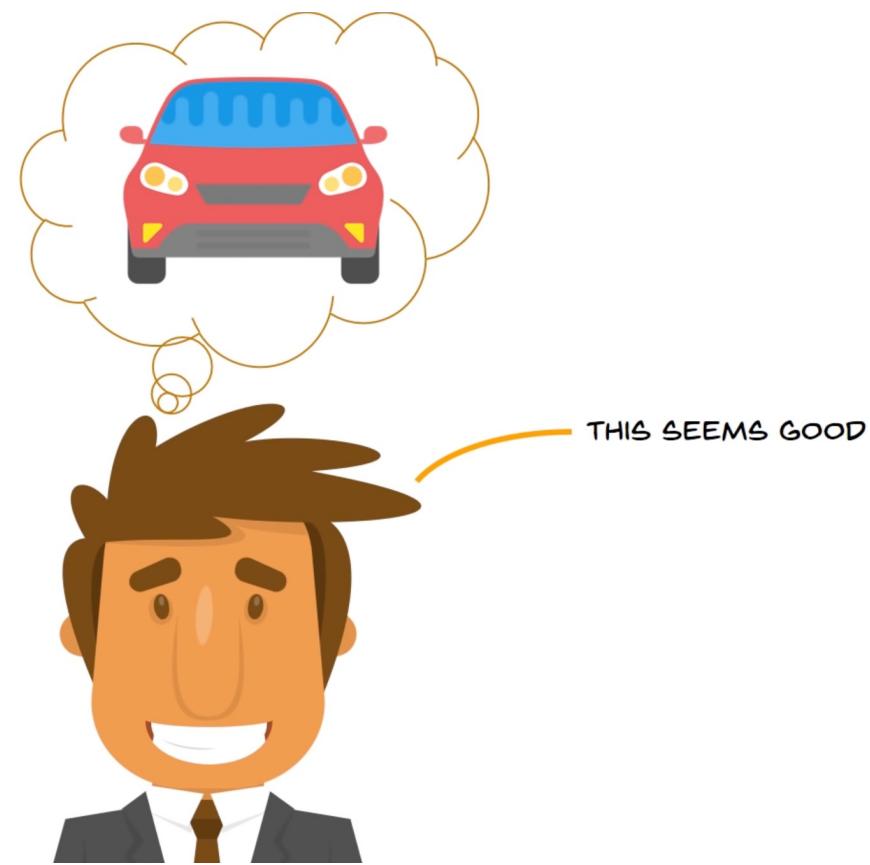
Decision Trees



Decision Trees



Decision Trees



THIS SEEMS GOOD



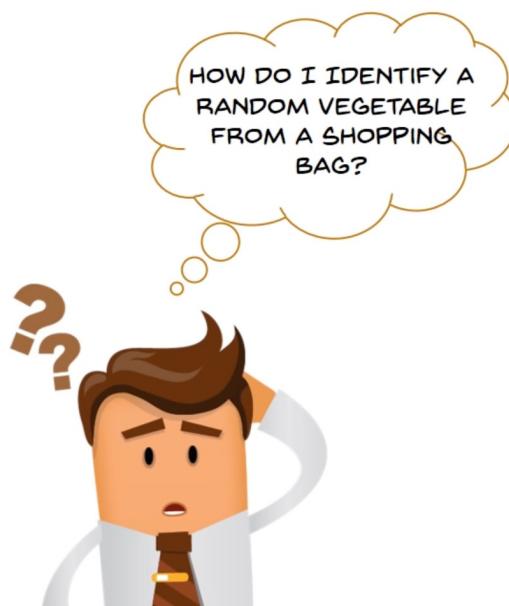
What is a decision tree?

What is a decision tree

Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction

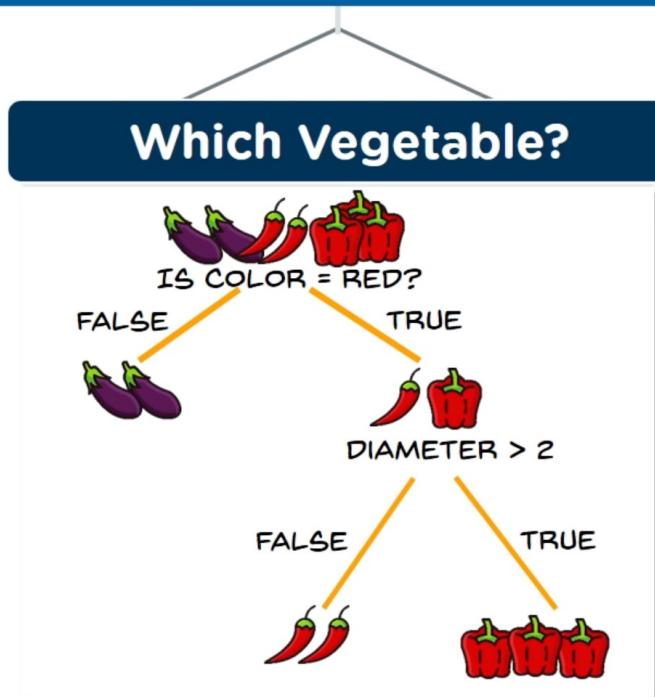
What is a decision tree

Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction



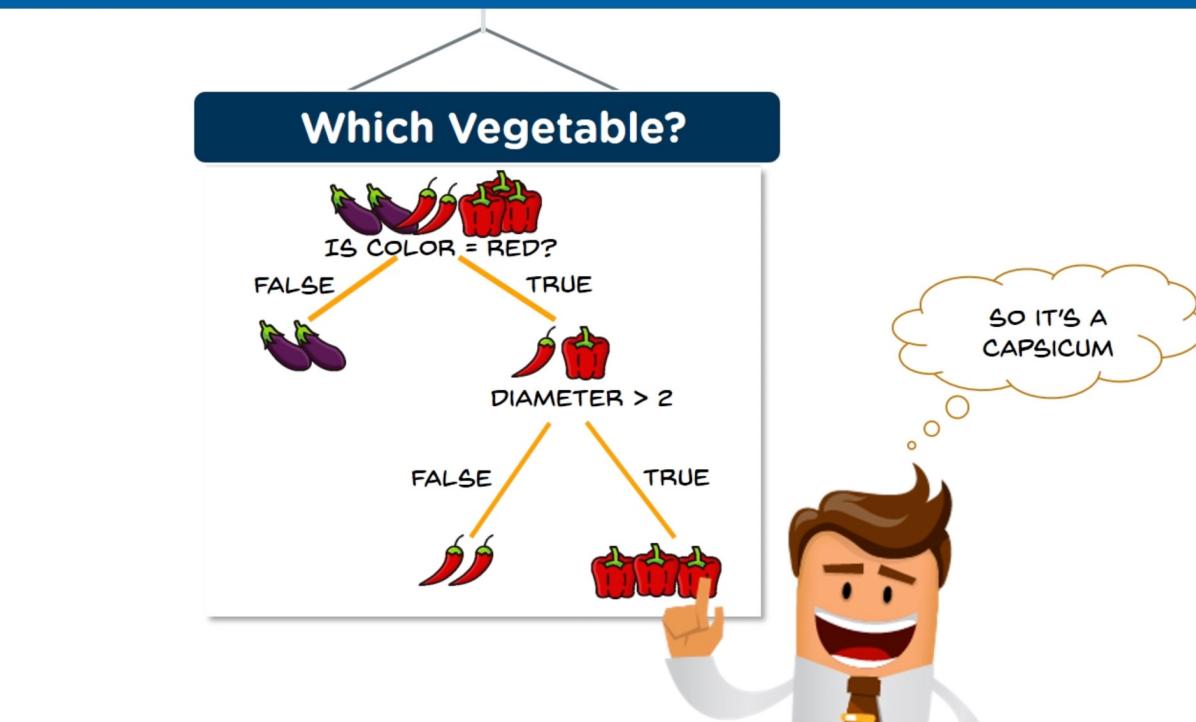
What is a decision tree

Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction



What is a decision tree

Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction



A close-up photograph of autumn maple leaves against a dark background. The leaves are a vibrant orange-yellow color, with some showing signs of aging or damage. They are arranged in several branches that curve across the frame. The lighting is soft, highlighting the texture of the leaves.

Problems that a decision tree can
solve

What is a decision tree

Classification

Regression



What is a decision tree

Classification



A classification tree will determine a set of logical if-then conditions to classify problems.

For example, discriminating between three types of flowers based on certain features

Regression



What is a decision tree

Classification

A classification tree will determine a set of logical if-then conditions to classify problems.

For example, discriminating between three types of flowers based on certain features



Regression

Regression tree is used when the target variable is numerical or continuous in nature. We fit a regression model to the target variable using each of the independent variables. Each split is made based on the sum of squared error.

A photograph of several tall palm trees with green fronds, standing against a clear, vibrant blue sky. The trees are positioned at different heights and angles, creating a sense of depth and tropical atmosphere.

Advantages of decision trees

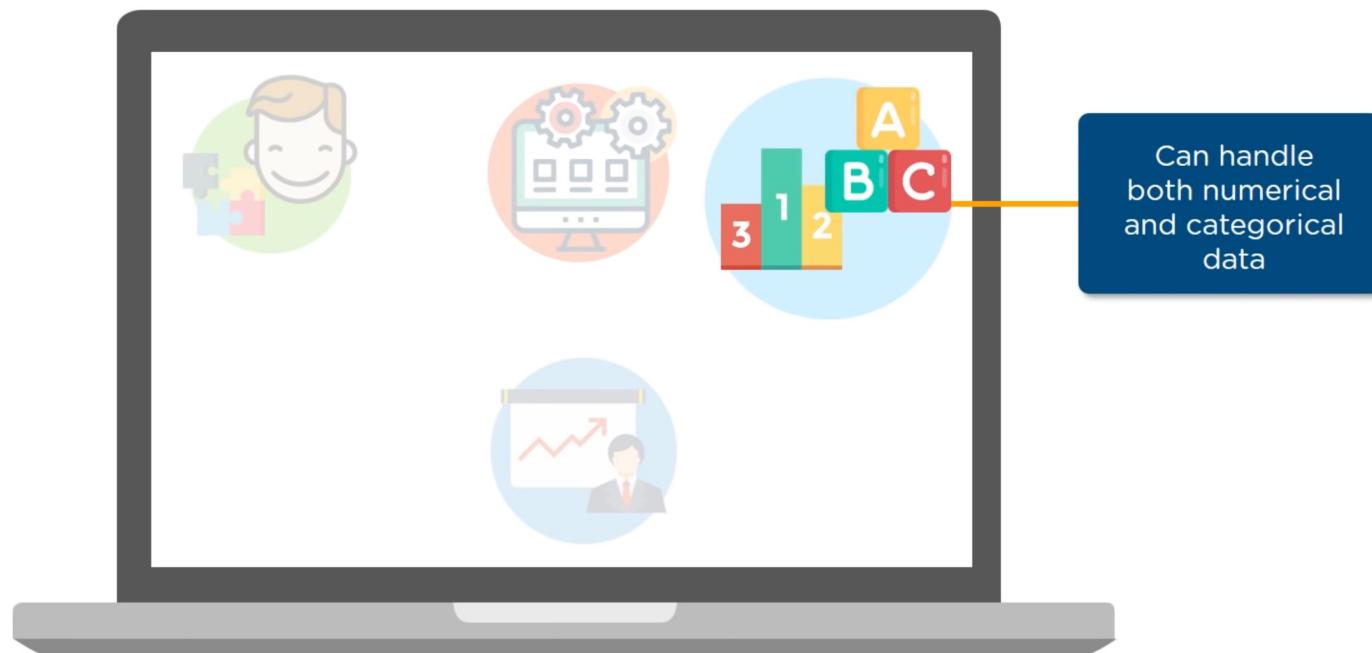
Advantages of decision trees



Advantages of decision trees

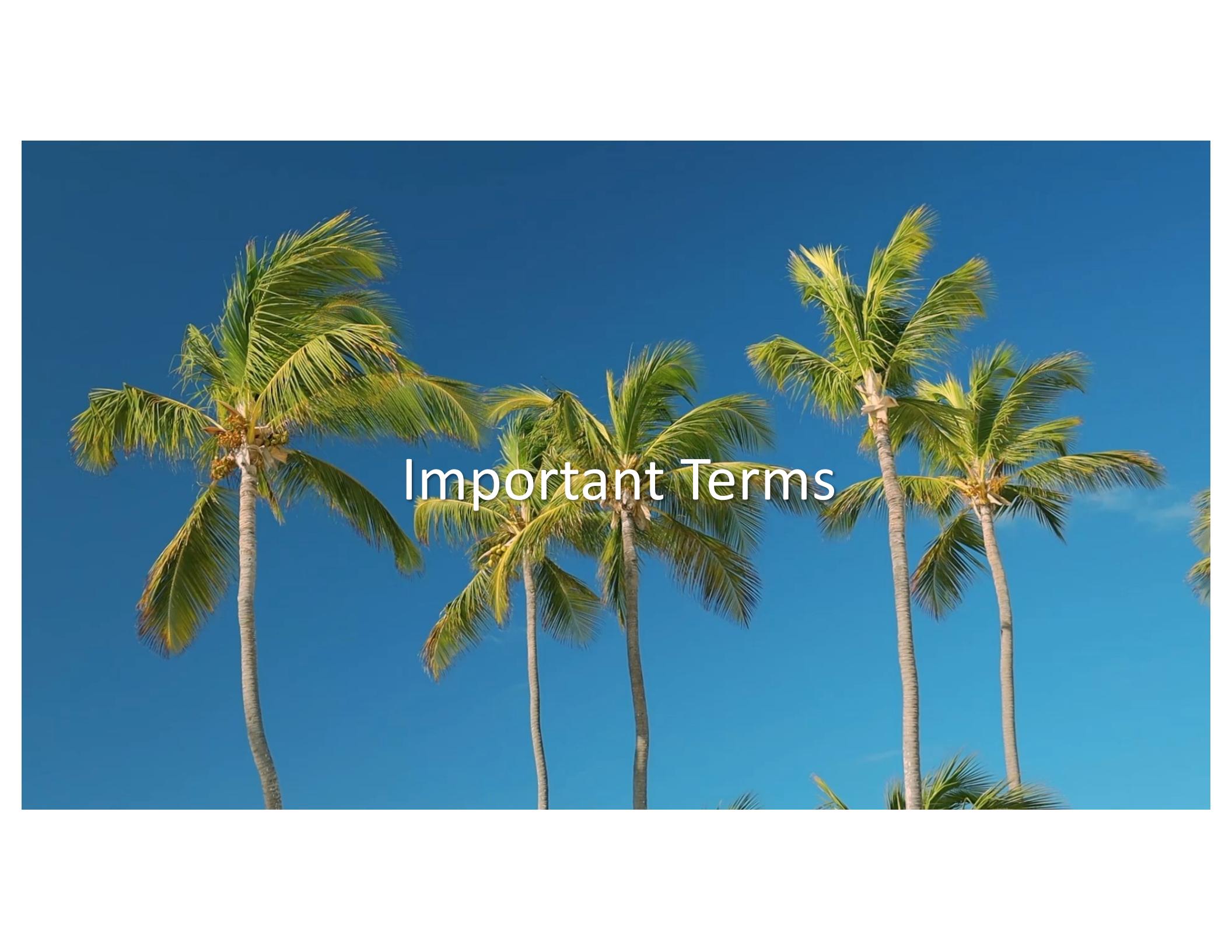


Advantages of decision trees



Advantages of decision trees



A photograph of several tall palm trees with lush green fronds, standing against a clear, vibrant blue sky. The trees are positioned at different heights and angles, creating a sense of depth and tropical atmosphere.

Important Terms

Decision Tree – Important Terms

ENTROPY

ENTROPY IS THE MEASURE OF RANDOMNESS OR UNPREDICTABILITY IN THE DATASET

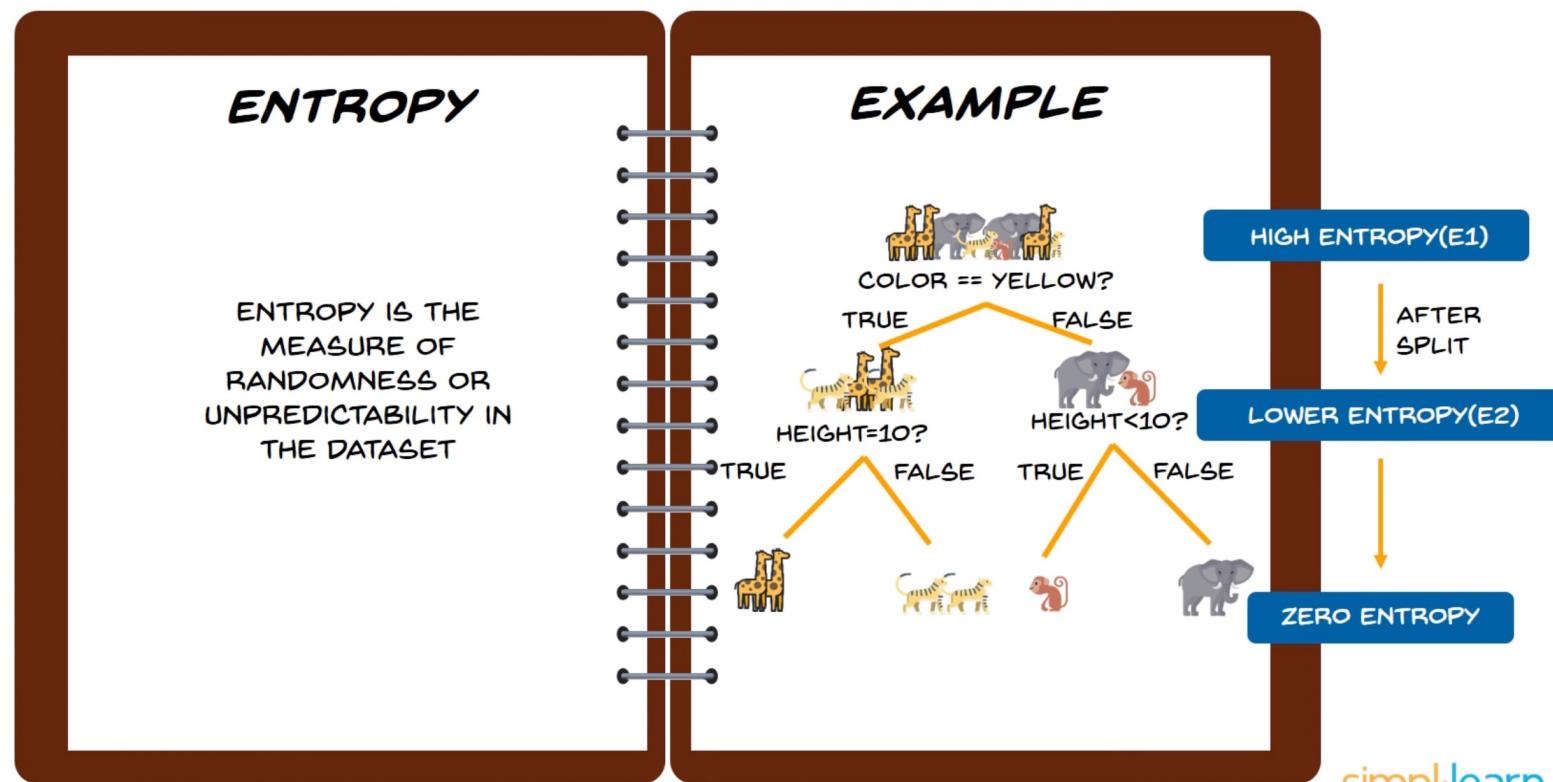
EXAMPLE



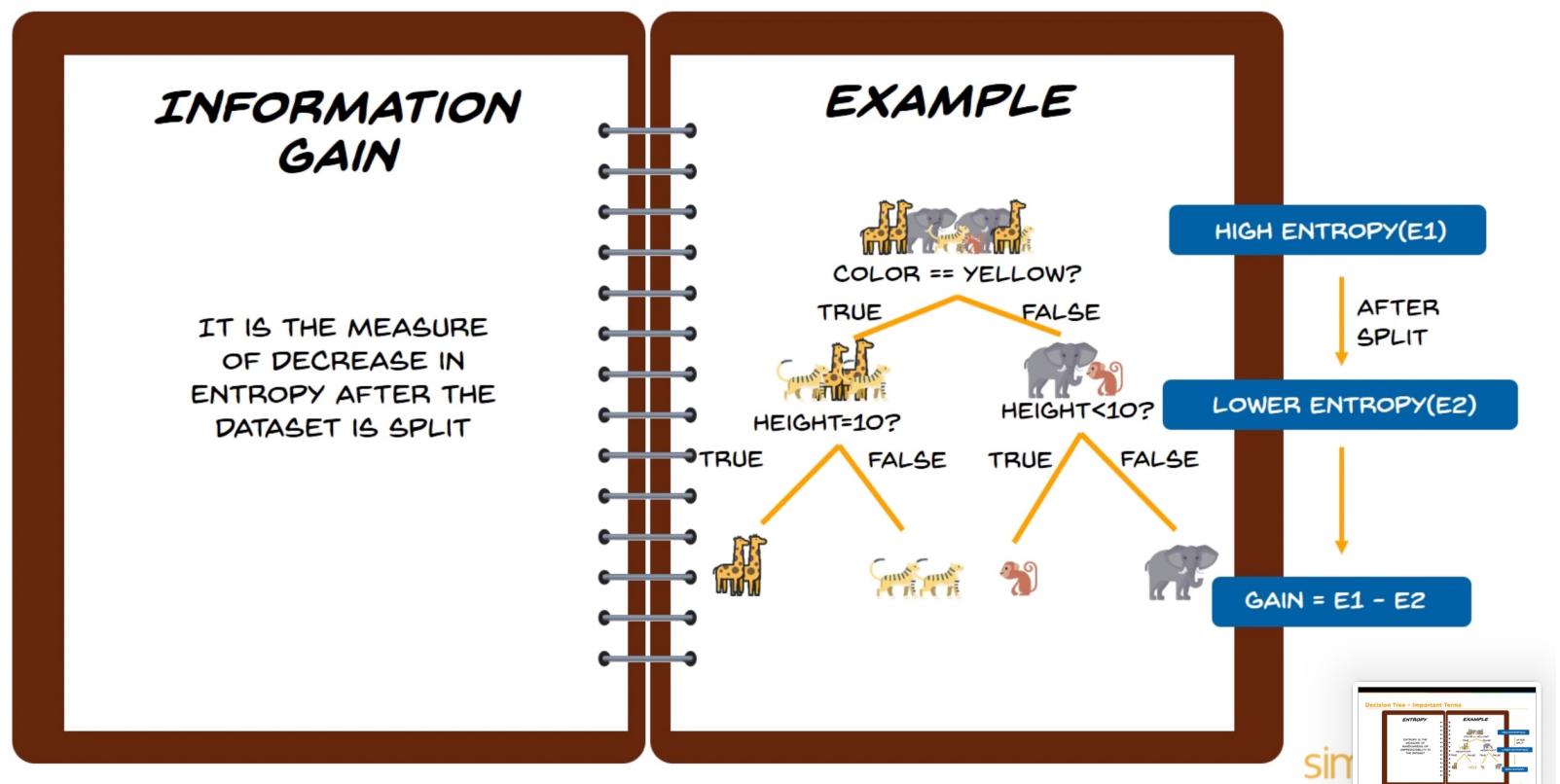
HIG

THIS DATASET HAS A VERY HIGH ENTROPY

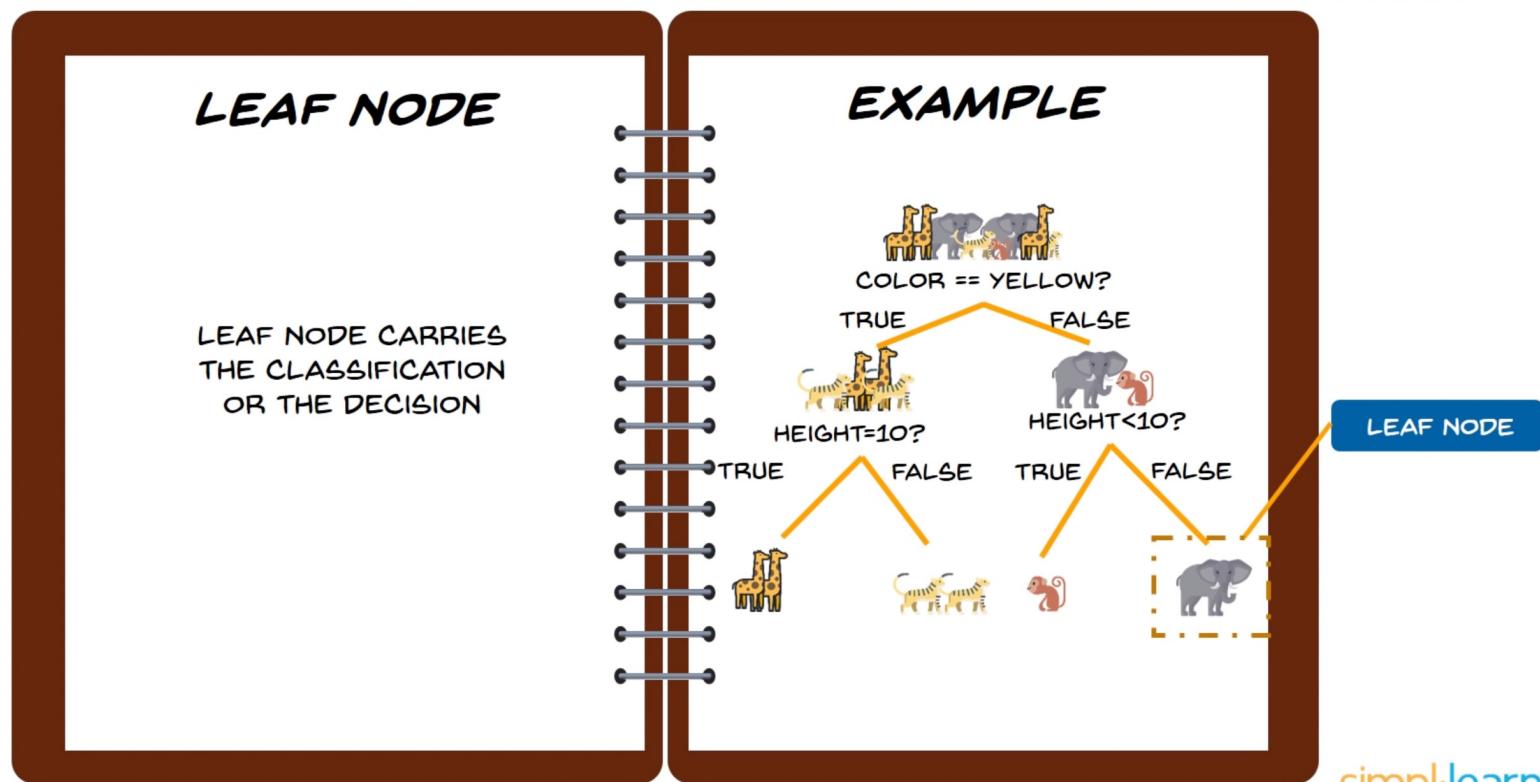
Decision Tree – Important Terms



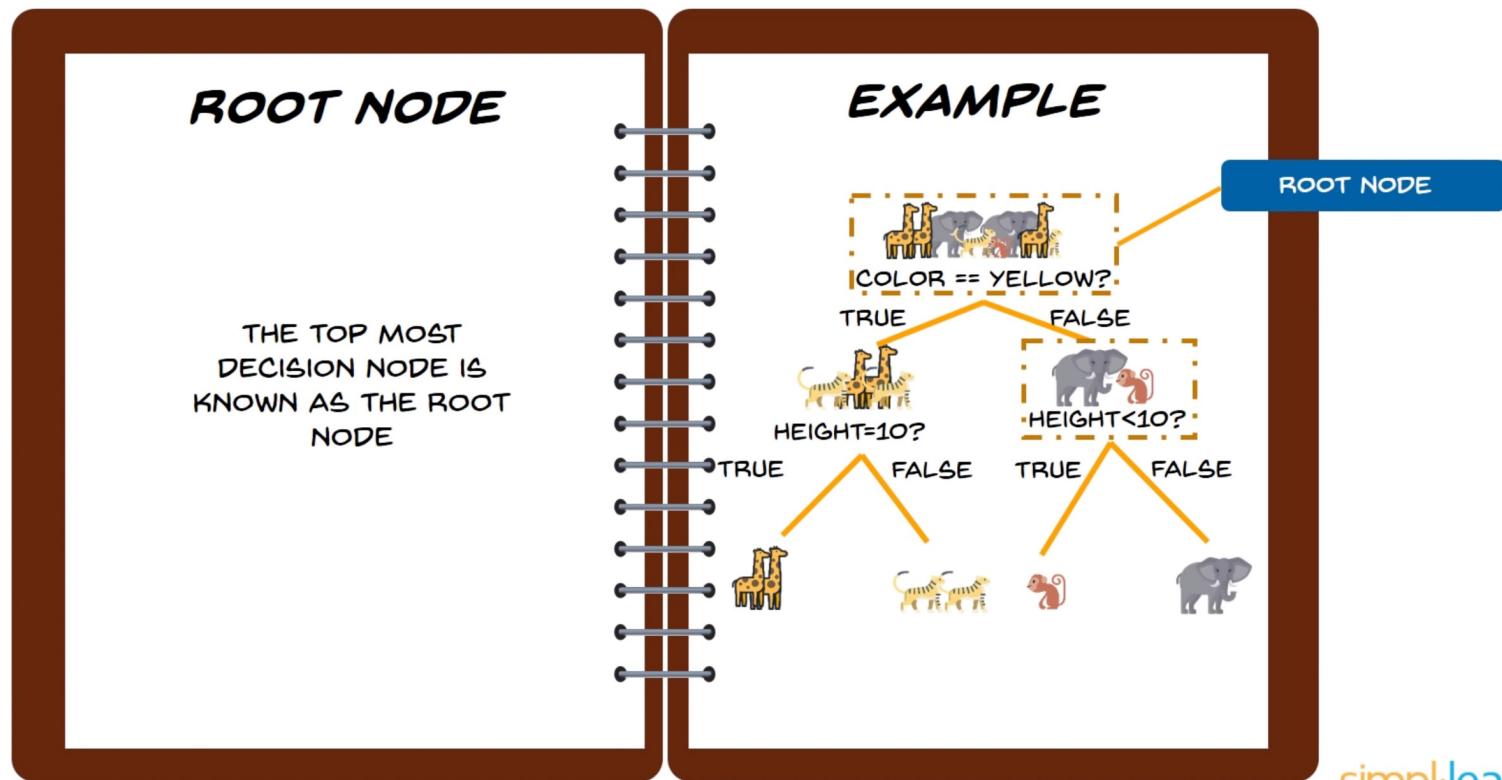
Decision Tree – Important Terms



Decision Tree – Important Terms



Decision Tree – Important Terms



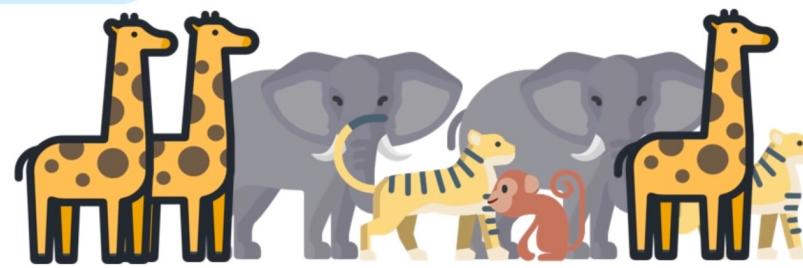
How does a Decision Tree work ?



How does a Decision Tree work ?

PROBLEM STATEMENT

TO CLASSIFY THE DIFFERENT TYPES OF ANIMALS BASED ON THEIR FEATURES USING DECISION TREE

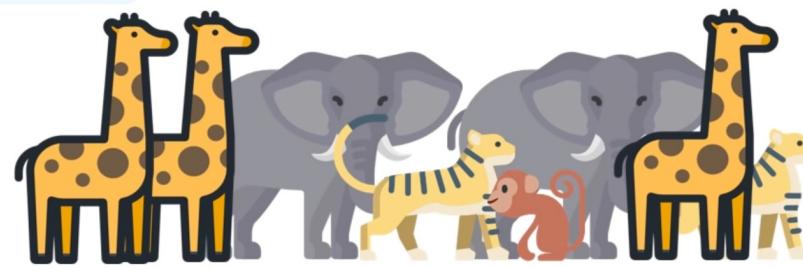


How does a Decision Tree work ?

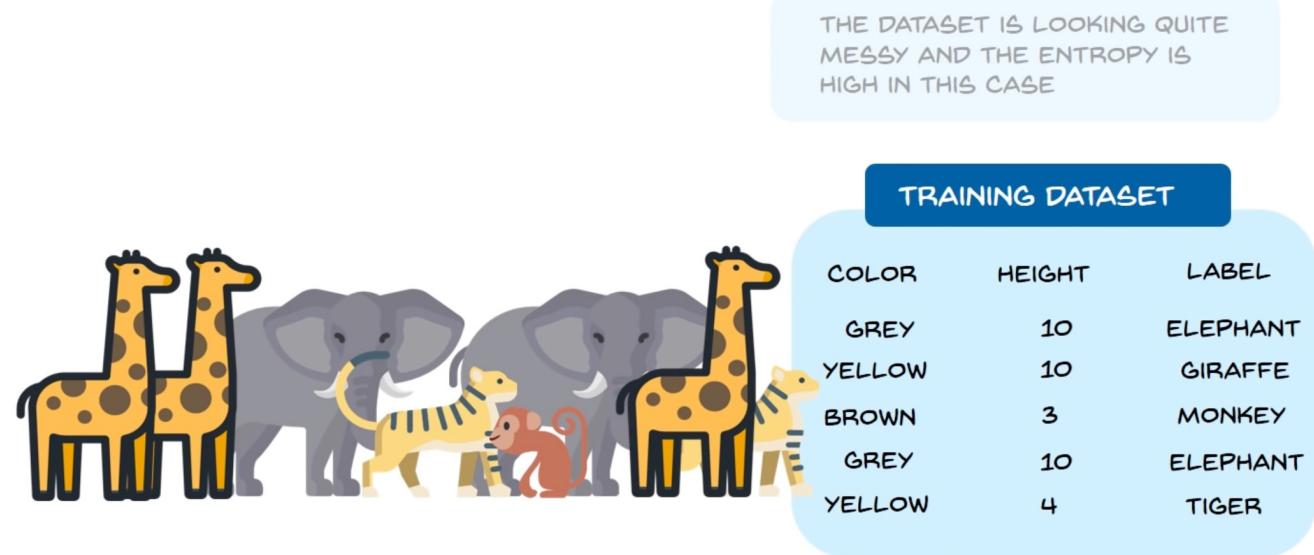
PROBLEM STATEMENT

TO CLASSIFY THE DIFFERENT TYPES OF ANIMALS BASED ON THEIR FEATURES USING DECISION TREE

THE DATASET IS LOOKING QUITE MESSY AND THE ENTROPY IS HIGH IN THIS CASE



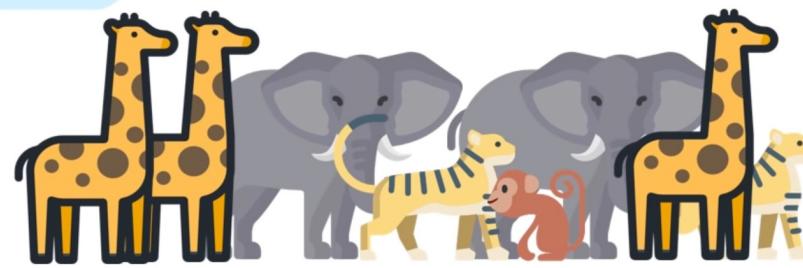
How does a Decision Tree work ?



How does a Decision Tree work ?

HOW TO SPLIT THE DATA

WE HAVE TO FRAME THE CONDITIONS THAT SPLIT THE DATA IN SUCH A WAY THAT THE INFORMATION GAIN IS THE HIGHEST



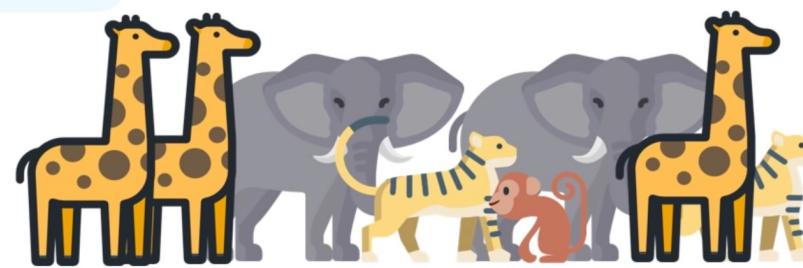
How does a Decision Tree work ?

HOW TO SPLIT THE DATA

WE HAVE TO FRAME THE CONDITIONS THAT SPLIT THE DATA IN SUCH A WAY THAT THE INFORMATION GAIN IS THE HIGHEST

NOTE

GAIN IS THE MEASURE OF DECREASE IN ENTROPY AFTER SPLITTING

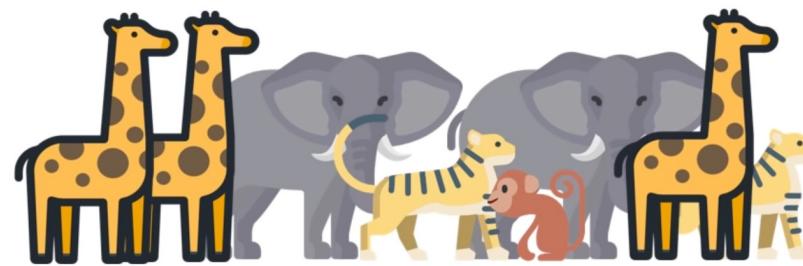


How does a Decision Tree work ?

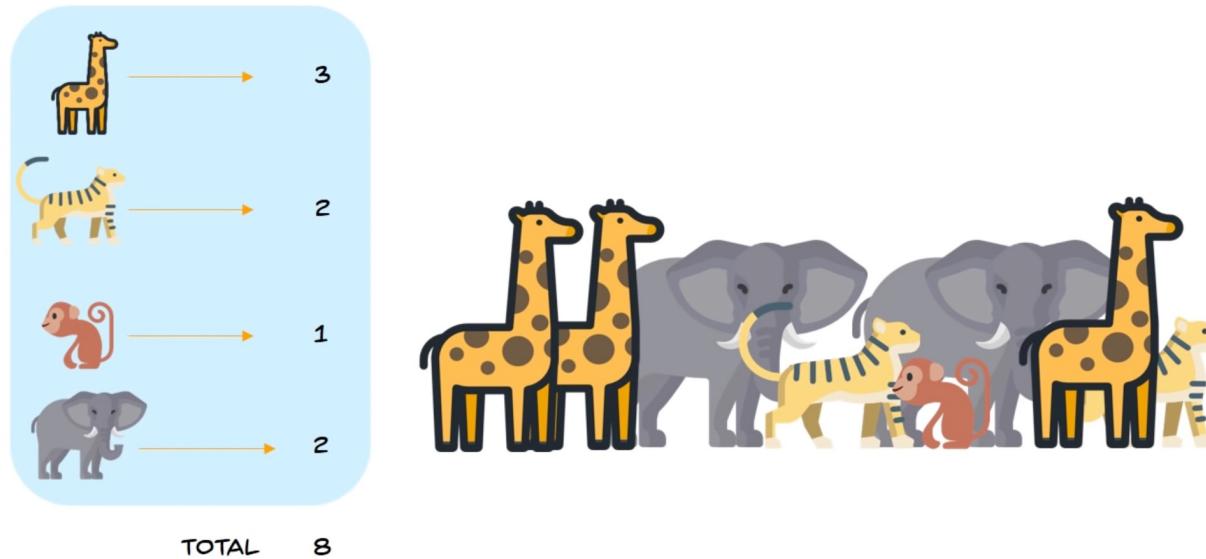
FORMULA FOR ENTROPY

$$\sum_{i=1}^k P(\text{value}_i) \cdot \log_2(P(\text{value}_i))$$

LET'S TRY TO CALCULATE
THE ENTROPY FOR THE
CURRENT DATASET



How does a Decision Tree work ?



How does a Decision Tree work ?

LET'S USE THE
FORMULA

$$\sum_{i=1}^k P(\text{value}_i) \cdot \log_2(P(\text{value}_i))$$

$$\text{ENTROPY} = \left(\frac{3}{8}\right) \log_2\left(\frac{3}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right) + \left(\frac{1}{8}\right) \log_2\left(\frac{1}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right)$$

ENTROPY=0.571



How does a Decision Tree work ?

LET'S USE THE
FORMULA

$$\sum_{i=1}^k P(\text{value}_i) \cdot \log_2(P(\text{value}_i))$$

$$\text{ENTROPY} = \left(\frac{3}{8}\right) \log_2\left(\frac{3}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right) + \left(\frac{1}{8}\right) \log_2\left(\frac{1}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right)$$

ENTROPY=0.571



WE WILL CALCULATE
THE ENTROPY OF THE
DATASET SIMILARLY
AFTER EVERY SPLIT TO
CALCULATE THE GAIN

How does a Decision Tree work ?



GAIN CAN BE
CALCULATED BY
FINDING THE
DIFFERENCE OF THE
SUBSEQUENT ENTROPY
VALUES AFTER SPLIT

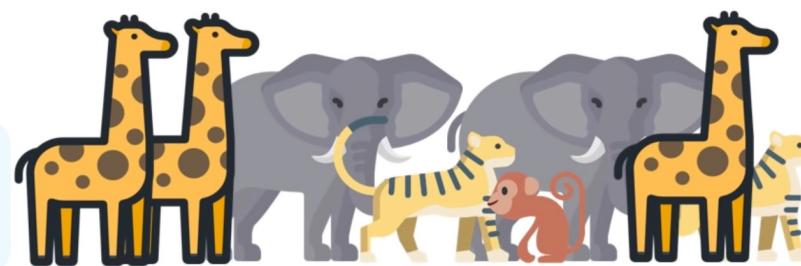
How does a Decision Tree work ?

NOW WE WILL TRY TO
CHOOSE A CONDITION
THAT GIVES US THE
HIGHEST GAIN



How does a Decision Tree work ?

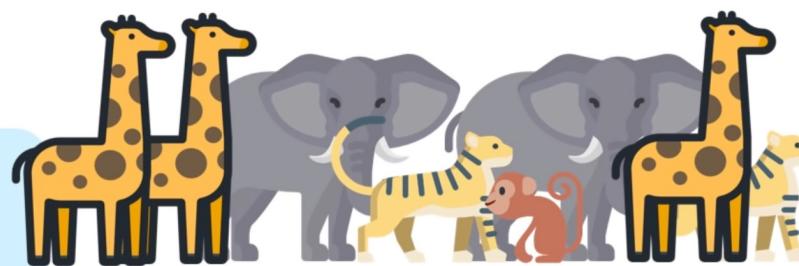
NOW WE WILL TRY TO
CHOOSE A CONDITION
THAT GIVES US THE
HIGHEST GAIN



WE WILL DO THAT BY
SPLITTING THE DATA
USING EACH CONDITION
AND CHECKING THE
GAIN THAT WE GET
OUT THEM.

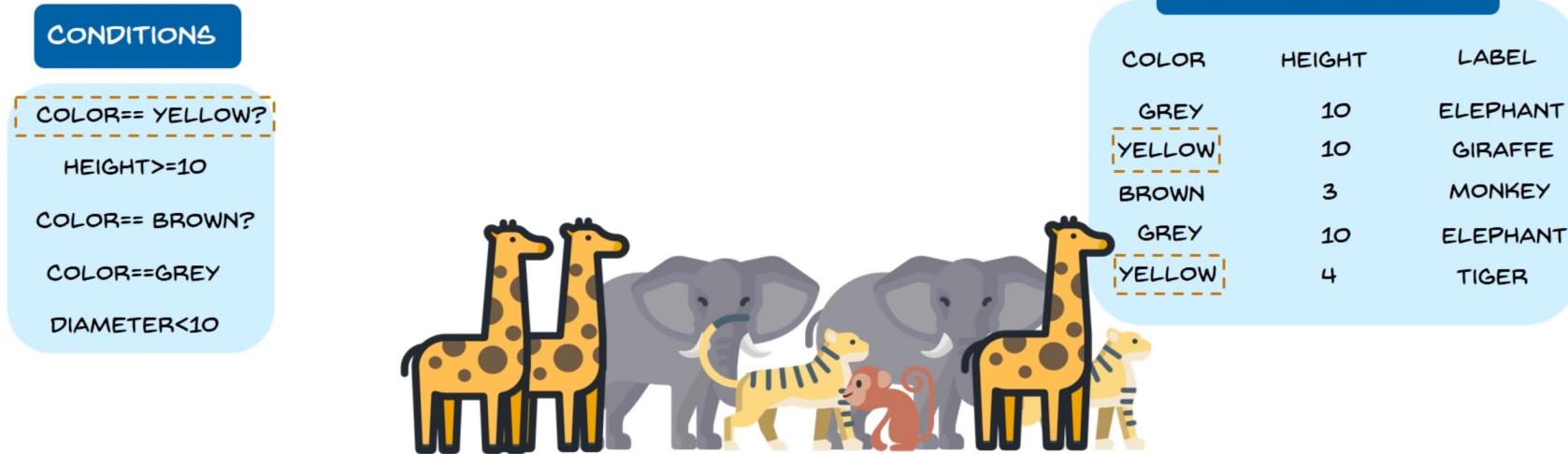
How does a Decision Tree work ?

THE CONDITION THAT GIVES US THE HIGHEST GAIN WILL BE USED TO MAKE THE FIRST SPLIT



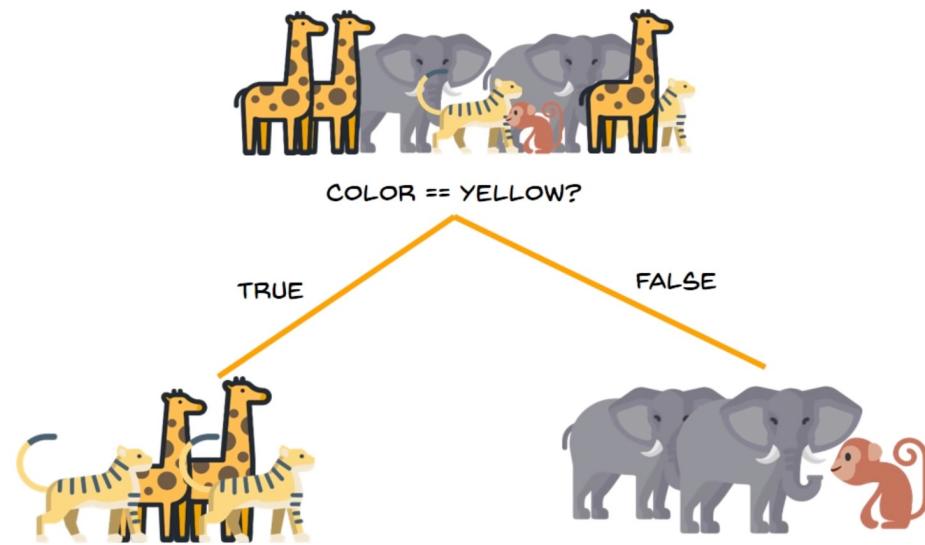
WE WILL DO THAT BY SPLITTING THE DATA USING EACH CONDITION AND CHECKING THE GAIN THAT WE GET OUT THEM.

How does a Decision Tree work ?

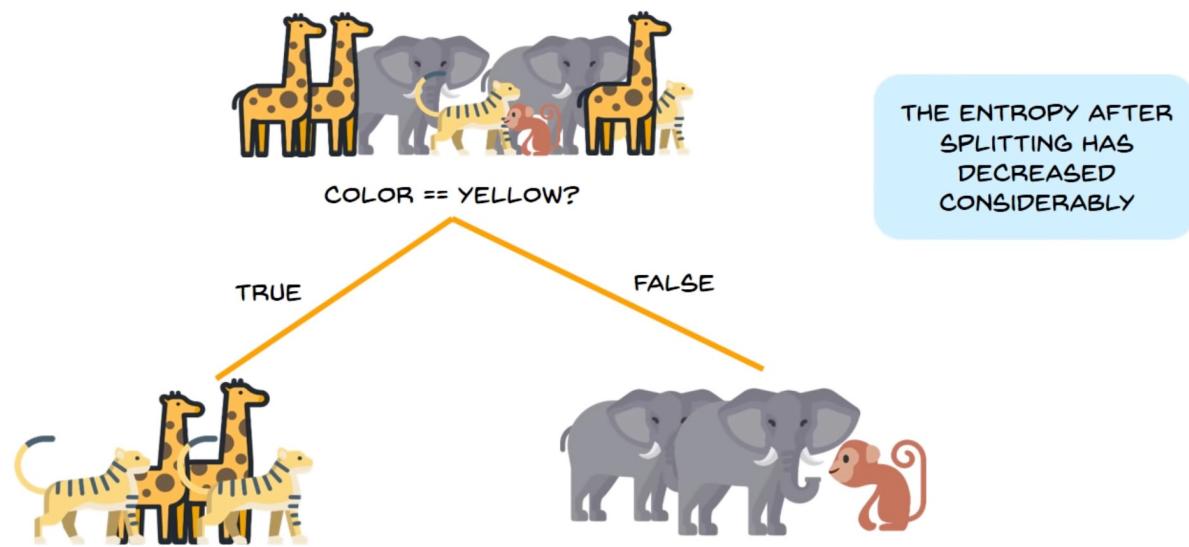


How does a Decision Tree work ?

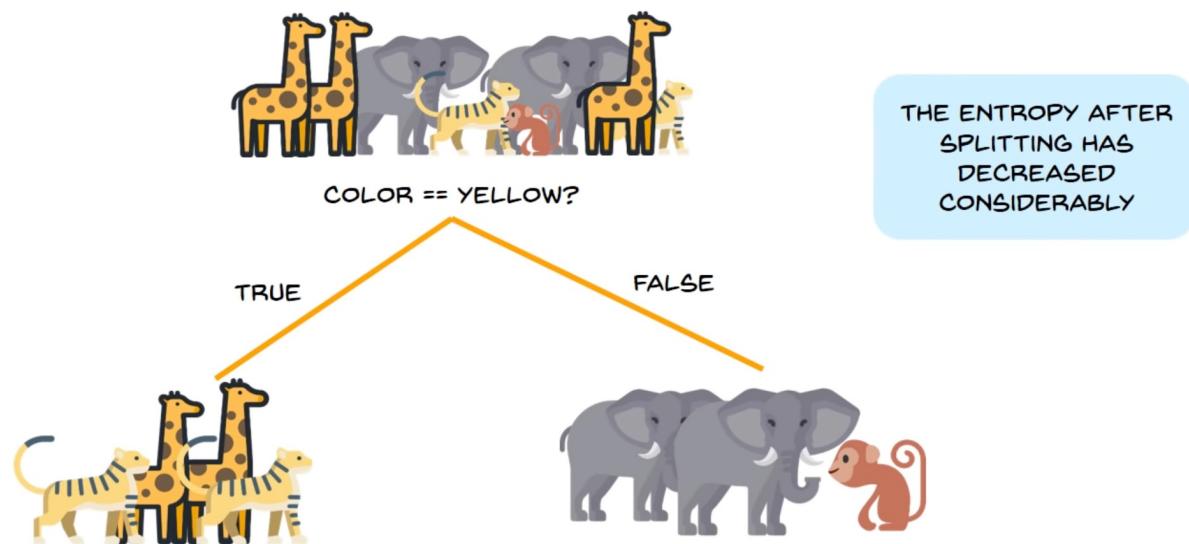
WE SPLIT THE DATA



How does a Decision Tree work ?

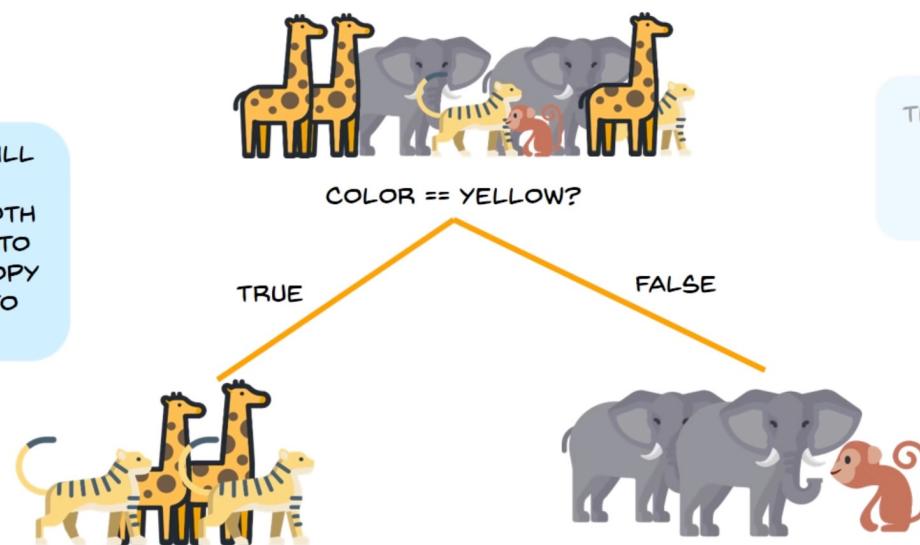


How does a Decision Tree work ?

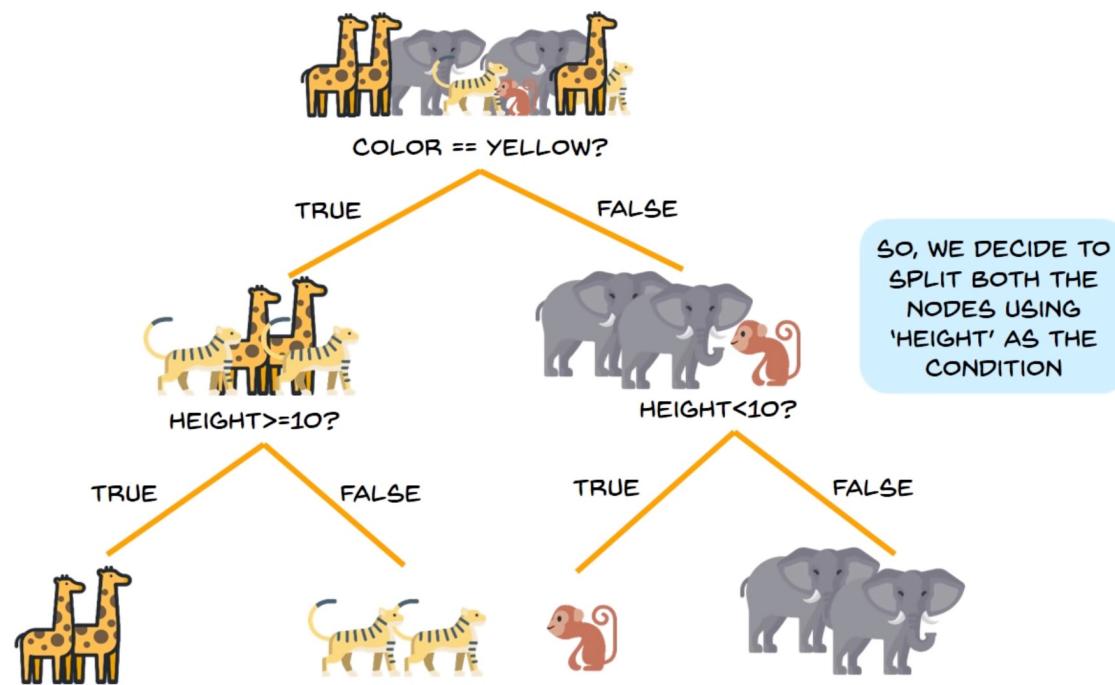


How does a Decision Tree work ?

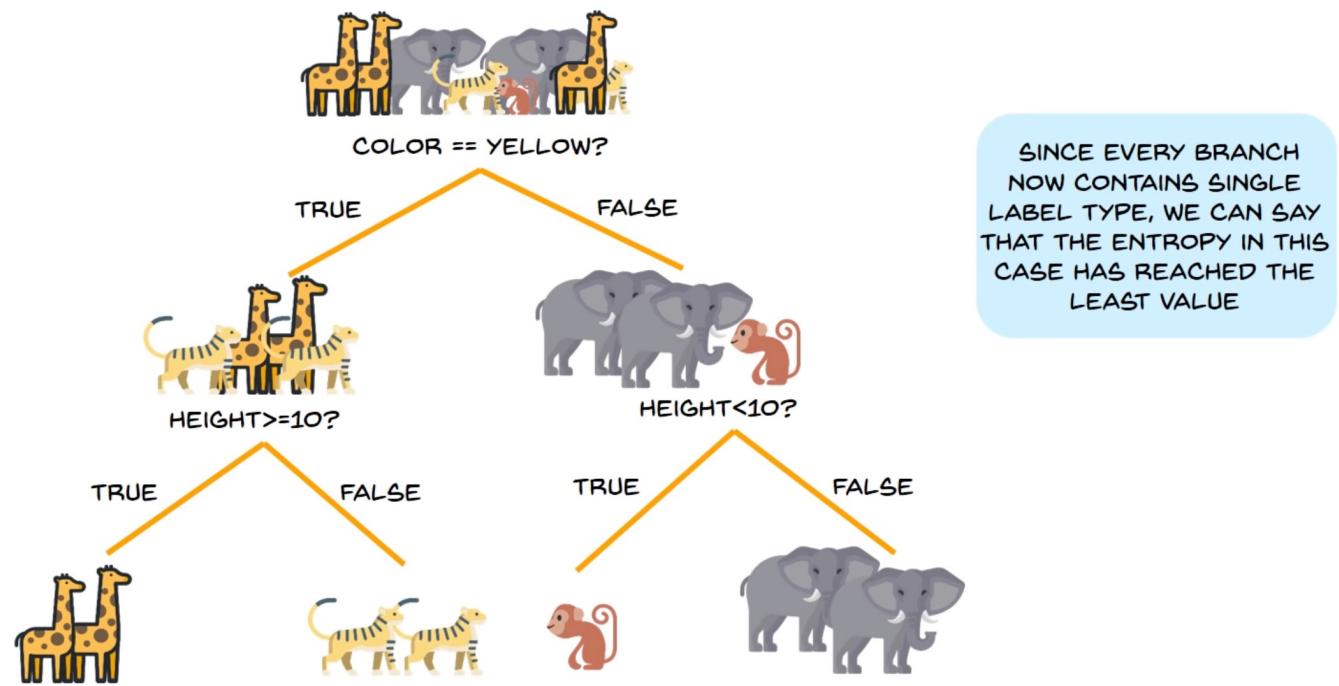
HOWEVER WE STILL
NEED SOME
SPLITTING AT BOTH
THE BRANCHES TO
ATTAIN AN ENTROPY
VALUE EQUAL TO
ZERO



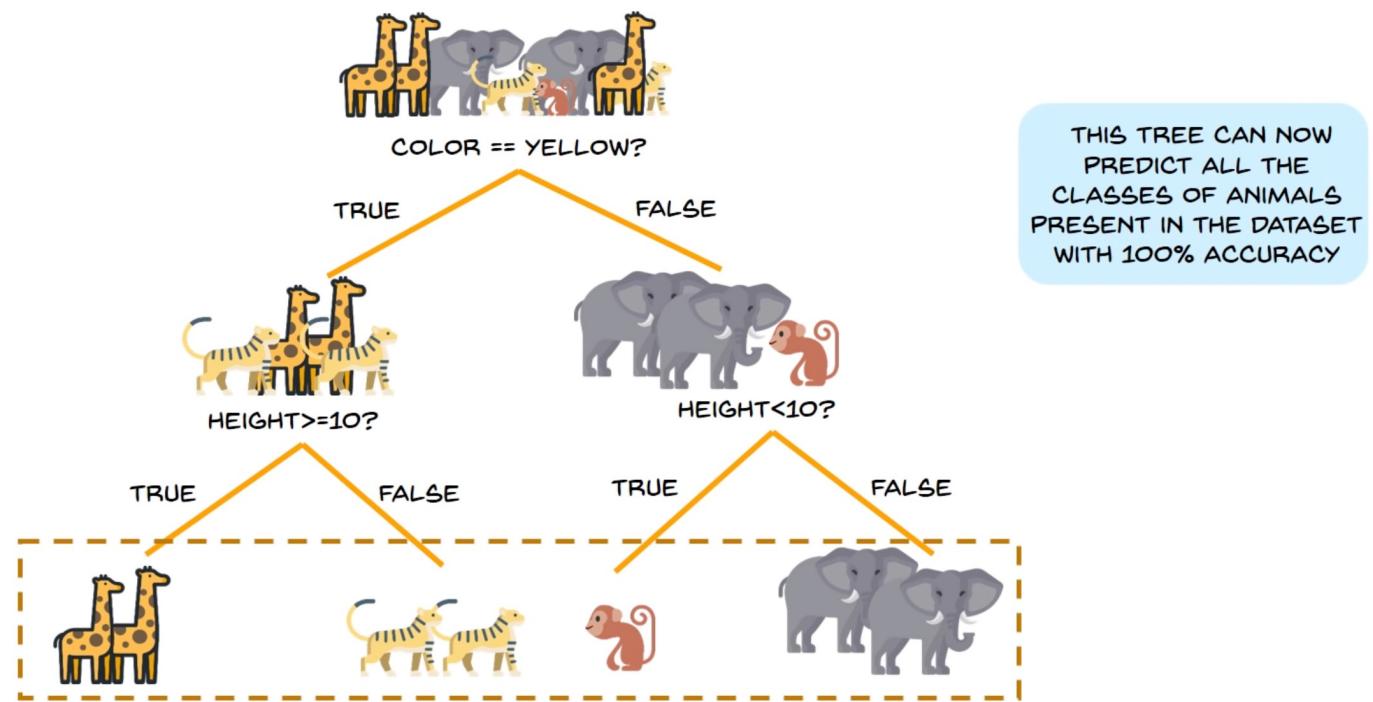
How does a Decision Tree work ?



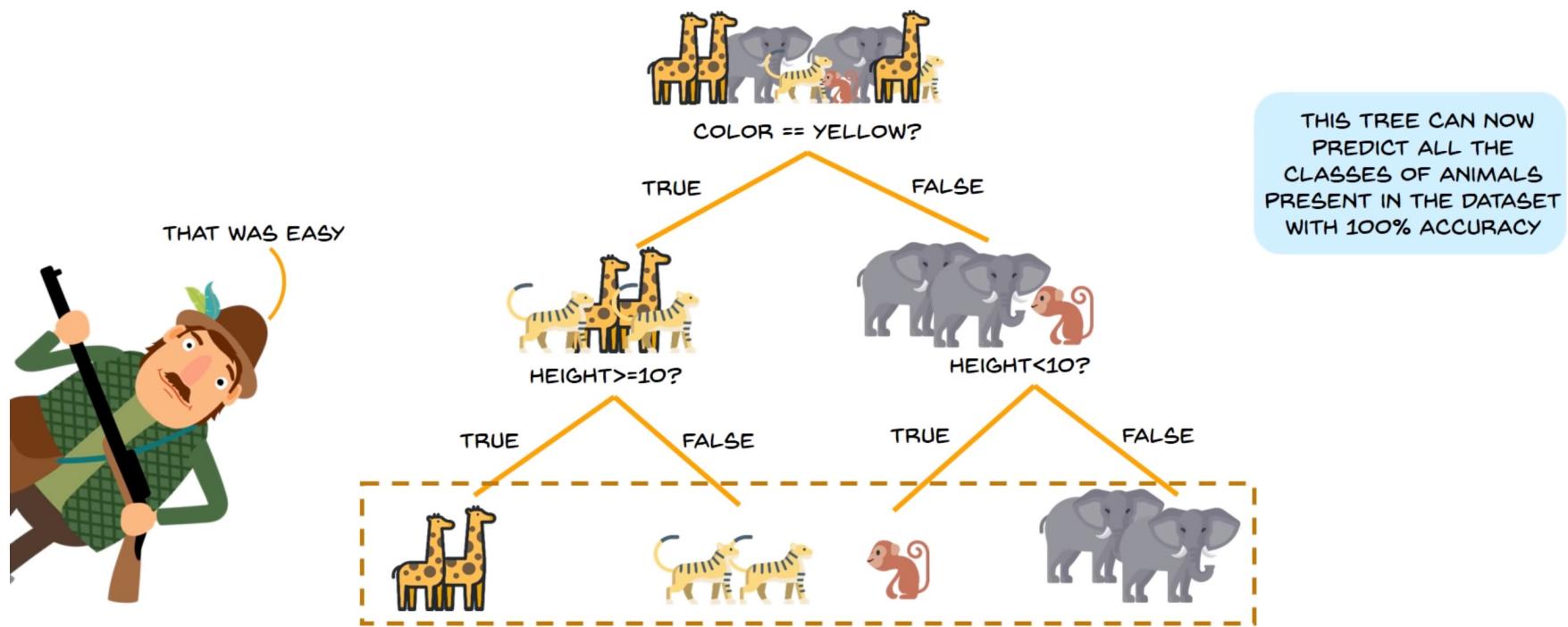
How does a Decision Tree work ?



How does a Decision Tree work ?



How does a Decision Tree work ?

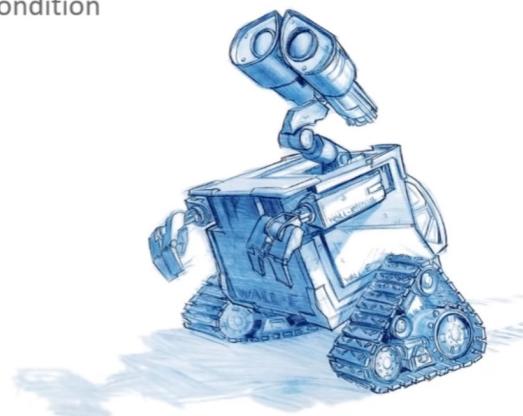


What is Classification?

“Classification is the process of dividing the datasets into different categories or groups by adding label”

- **Note:** It adds the data point to a particular

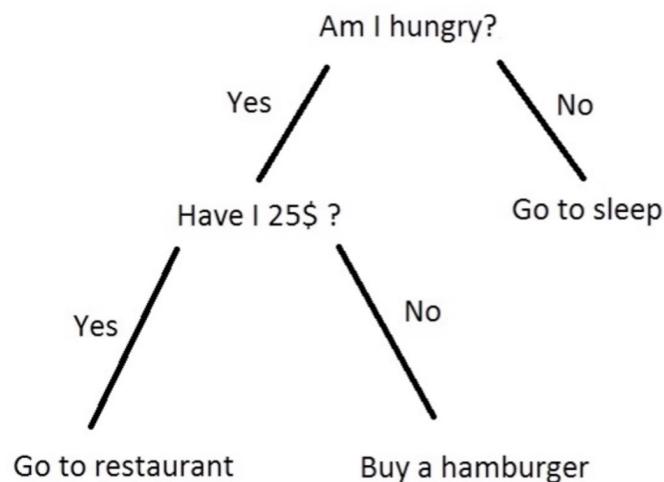
labelled group on the basis of some condition ”



Types of Classification

Decision Tree

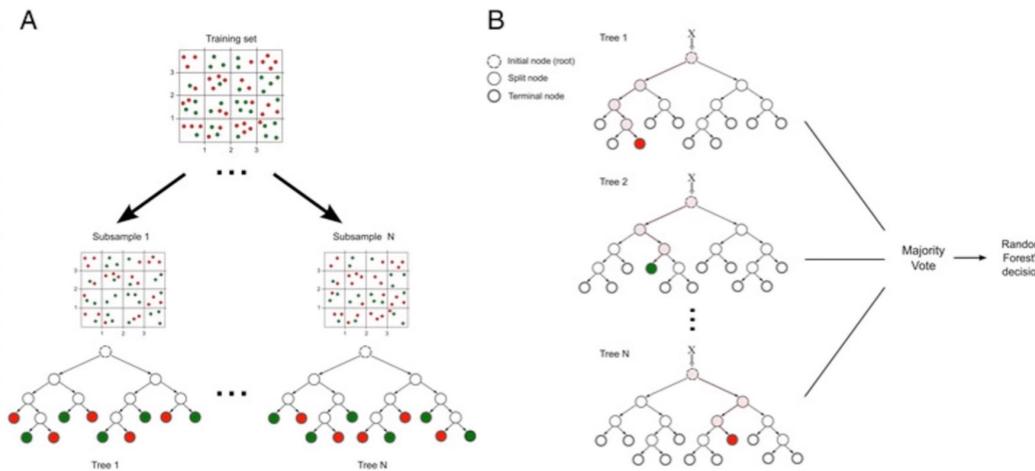
- Graphical representation of all the possible solutions to a decision
- Decisions are based on some conditions
- Decision made can be easily explained



Types of Classification

Random Forest

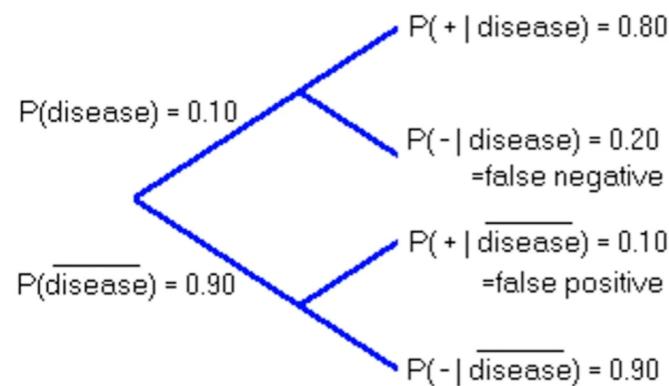
- Builds multiple decision trees and merges them together
- More accurate and stable prediction
- Random decision forests correct for decision trees' habit of overfitting to their training set
- Trained with the “bagging” method



Types of Classification

Naïve Bayes

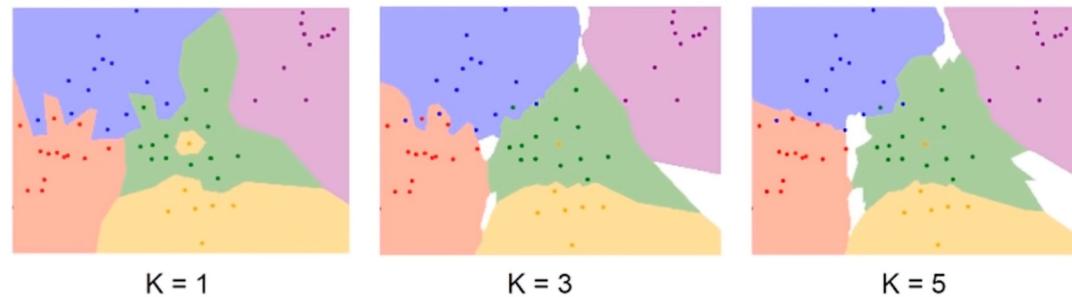
- Classification technique based on Bayes' Theorem
- Assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature



Types of Classification

KNN

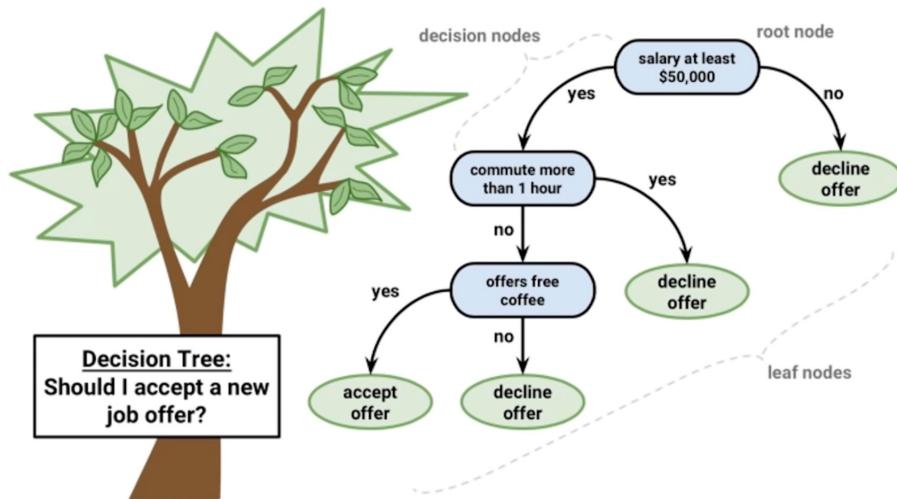
- Stores all the available cases and classifies new cases based on a similarity measure
- The “K” in KNN algorithm is the nearest neighbors we wish to take vote from.



What is decision tree?

Decision Tree

“A decisión tree es a graphical representation of all the possible solutions to a decision based on certain condicions”



Dataset

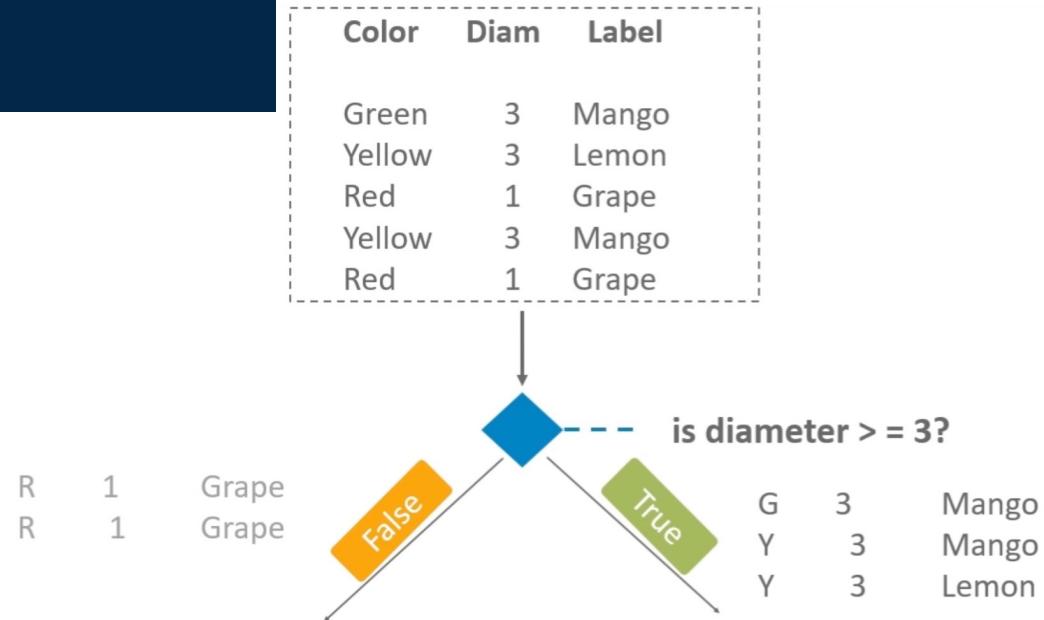
| Colour | Diameter | Label |
|--------|----------|-------|
| Green | 3 | Mango |
| Yellow | 3 | Mango |
| Red | 1 | Grape |
| Red | 1 | Grape |
| Yellow | 3 | Lemon |

Dataset

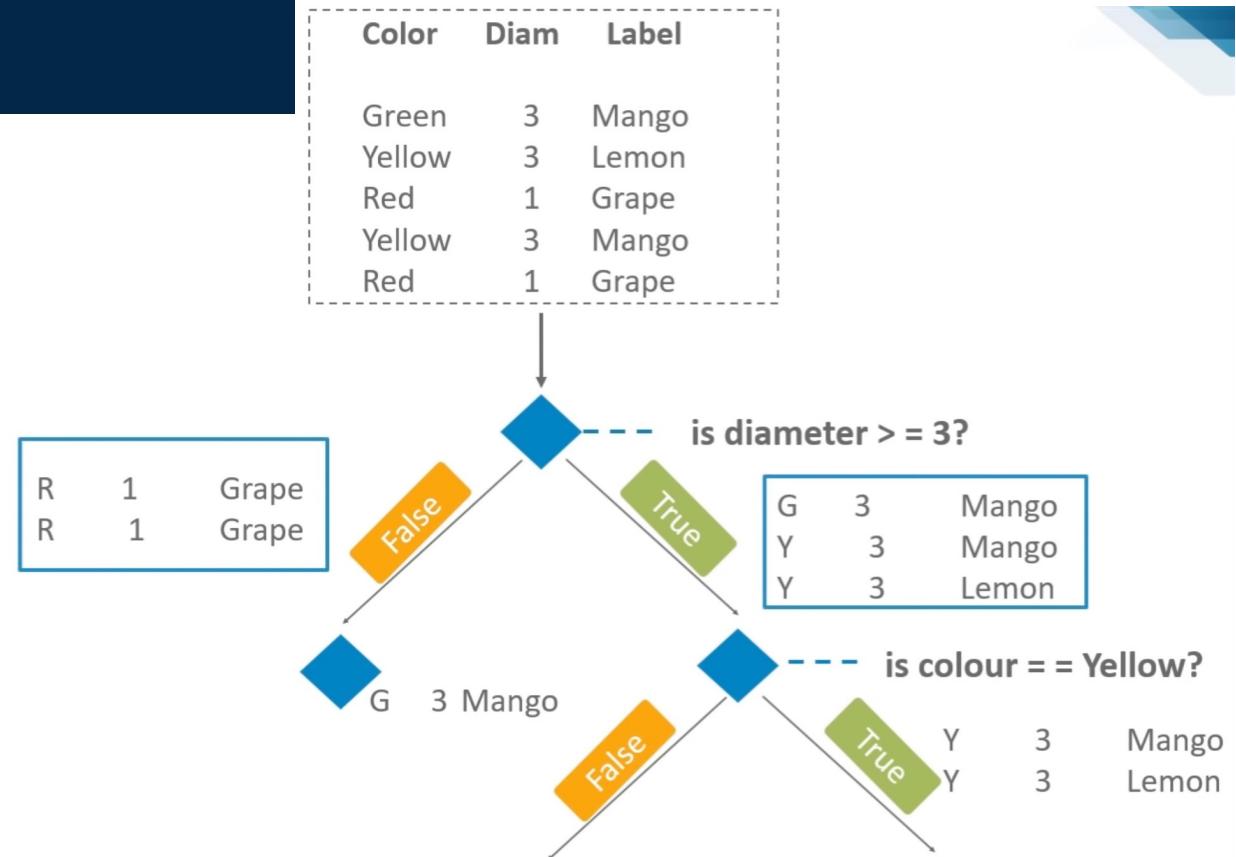
| Color | Diam | Label |
|--------|------|-------|
| Green | 3 | Mango |
| Yellow | 3 | Lemon |
| Red | 1 | Grape |
| Yellow | 3 | Mango |
| Red | 1 | Grape |



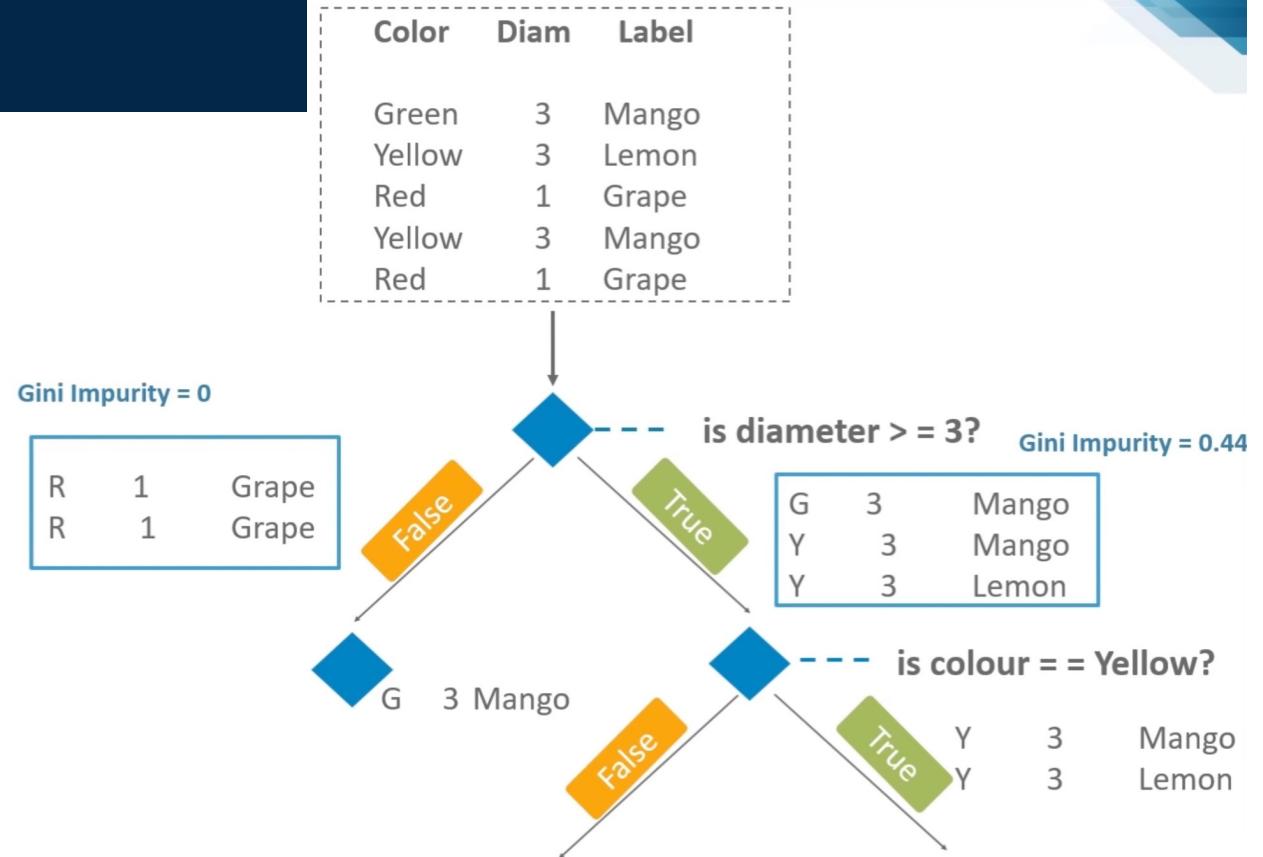
Dataset



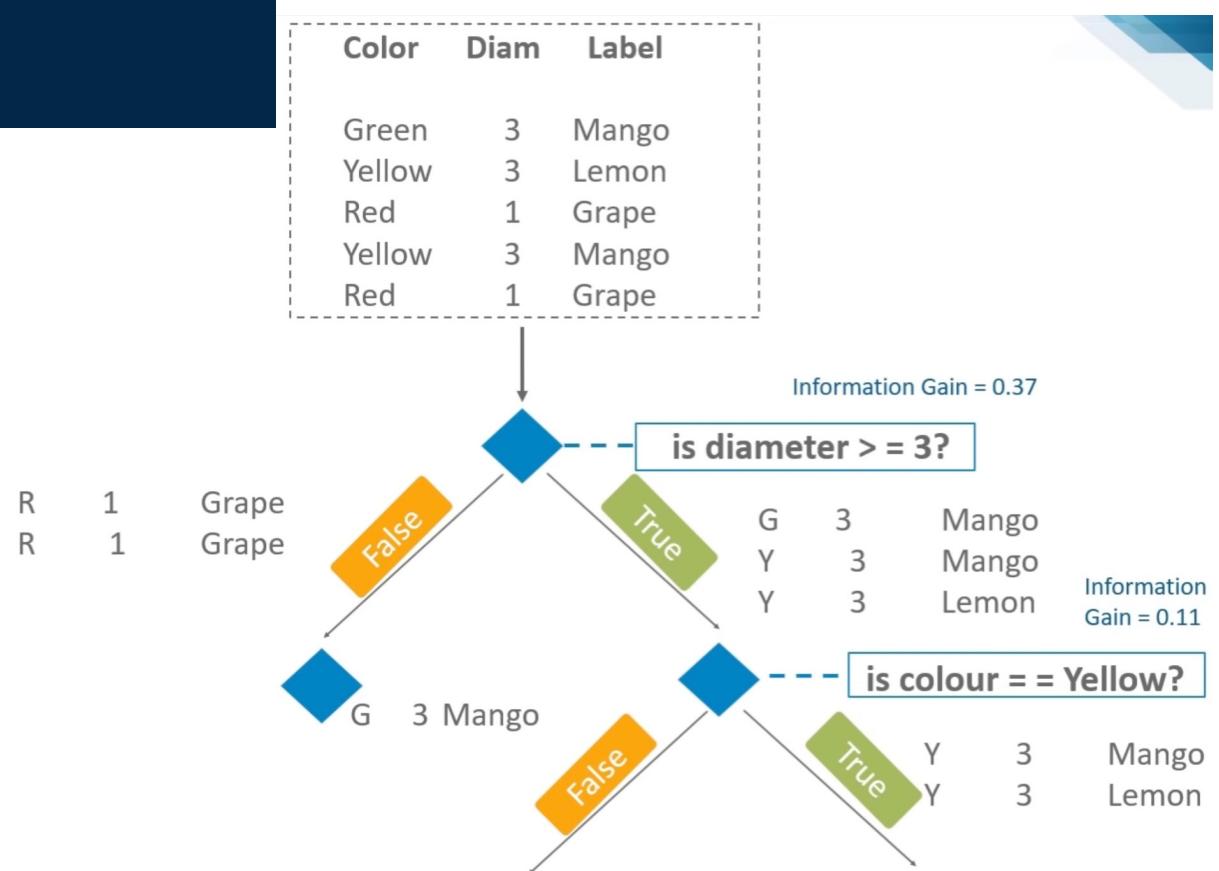
Dataset



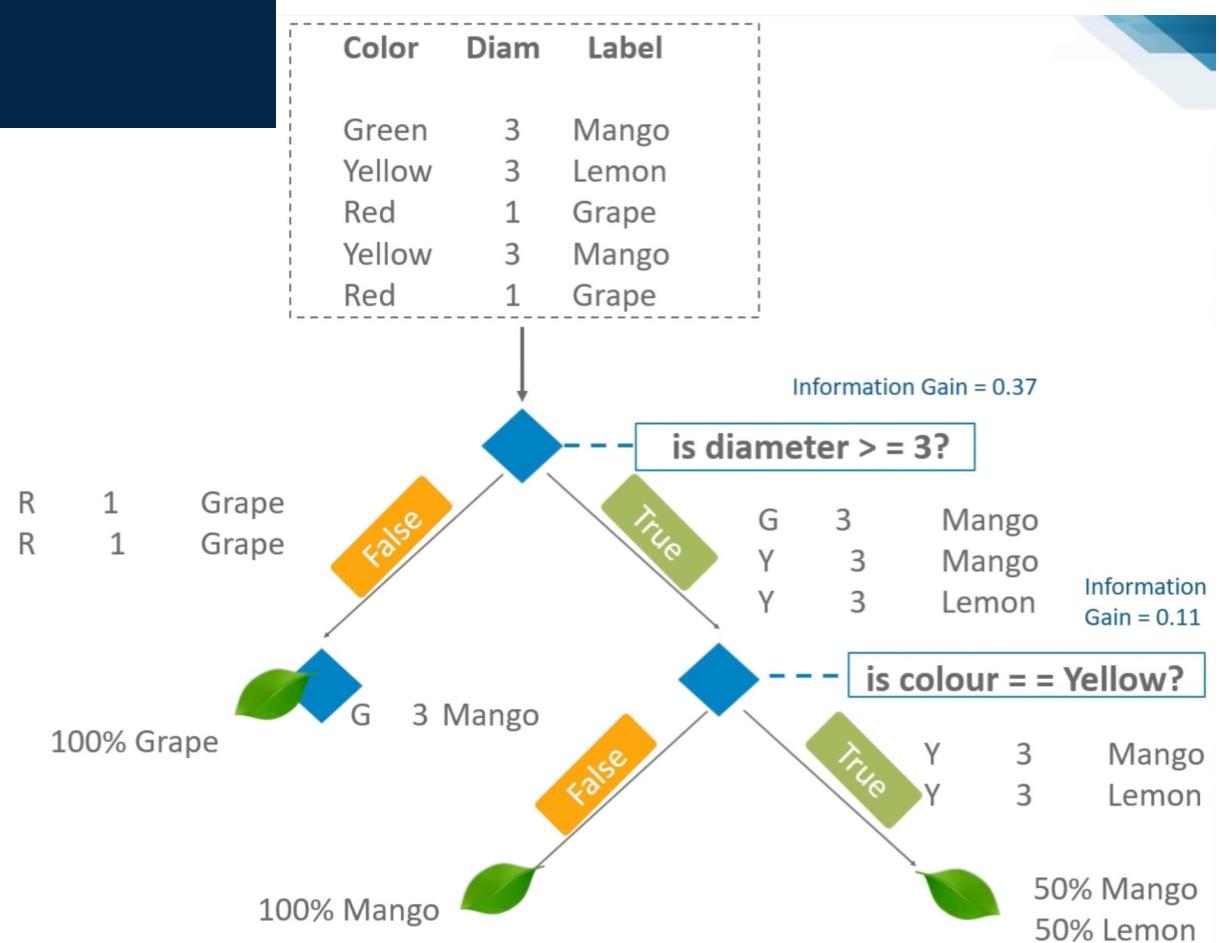
Dataset



Dataset



Dataset



Dataset

Green 3 Mango

Is the colour green?

Yellow 3 Lemon

Is the diameter ≥ 3

Yellow 3 Mango

Is the colour yellow

TRUE

False

Dataset

Is the colour green?

Is the diameter ≥ 3

Is the colour yellow

TRUE

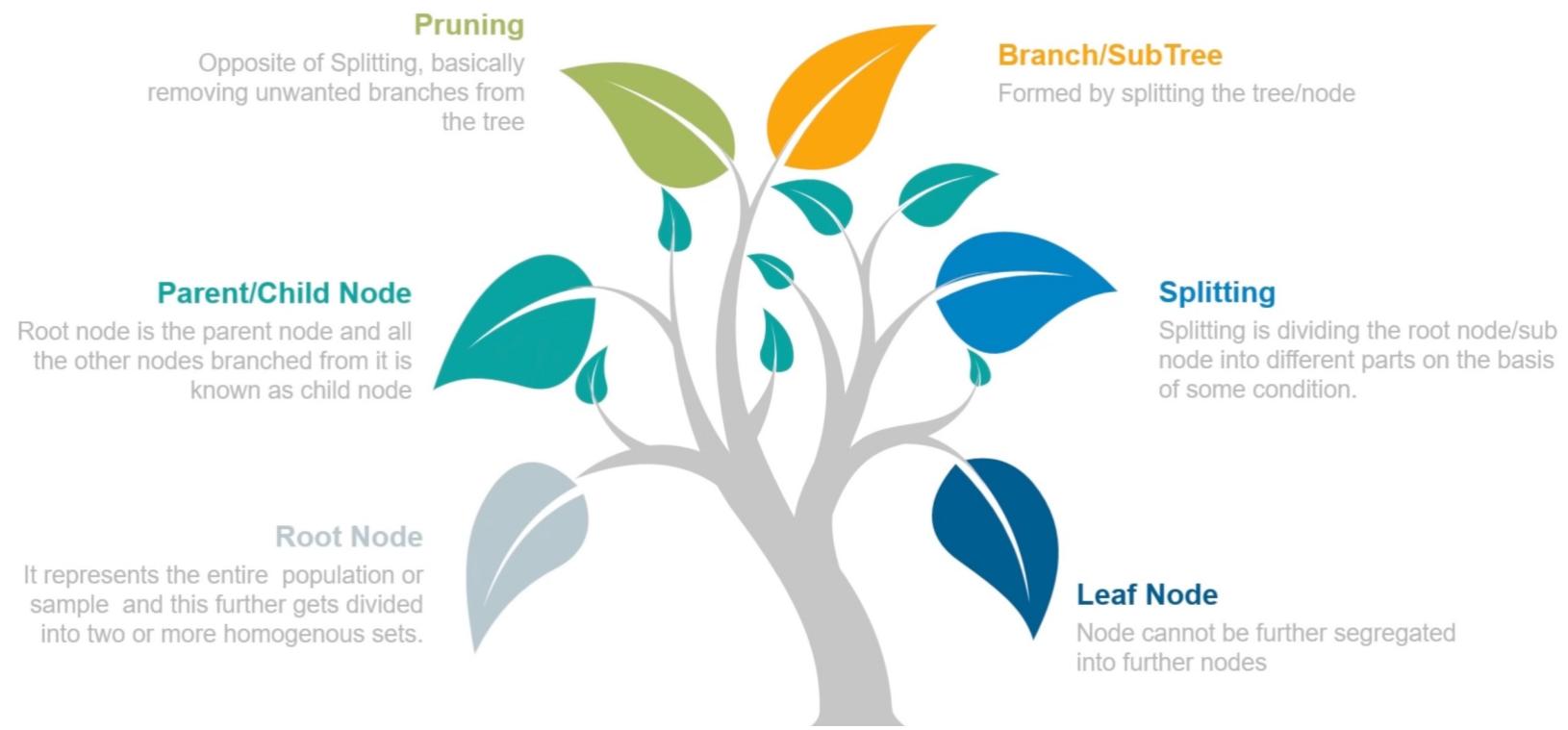
False

Green 3 Mango

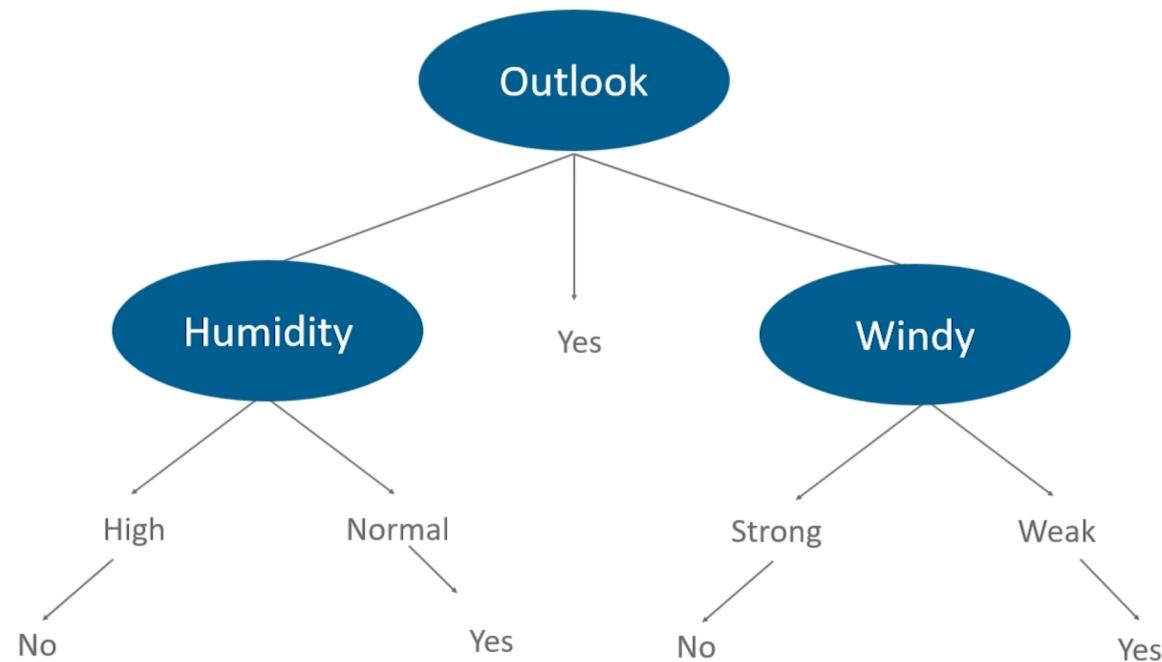
Yellow 3 Lemon

Yellow 3 Mango

Decision Tree Terminology



Let's First Visualize the Decision Tree



Learn about Tree Terminology

Which one among them
should you pick first?

| outlook | temp. | 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 |

Learn about Decision Tree

Answer: Determine the attribute that best classifies the training data

| outlook | temp. | 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 |

Learn about Decision Tree

But How do we choose
the best attribute?

Or

How does a tree decide
where to split?

| outlook | temp. | 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 |

How does A Tree Decide Where To Split?

Gini Index

The measure of impurity (or purity) used in building decision tree in CART is Gini Index

Chi Square

It is an algorithm to find out the statistical significance between the differences between sub-nodes and parent node



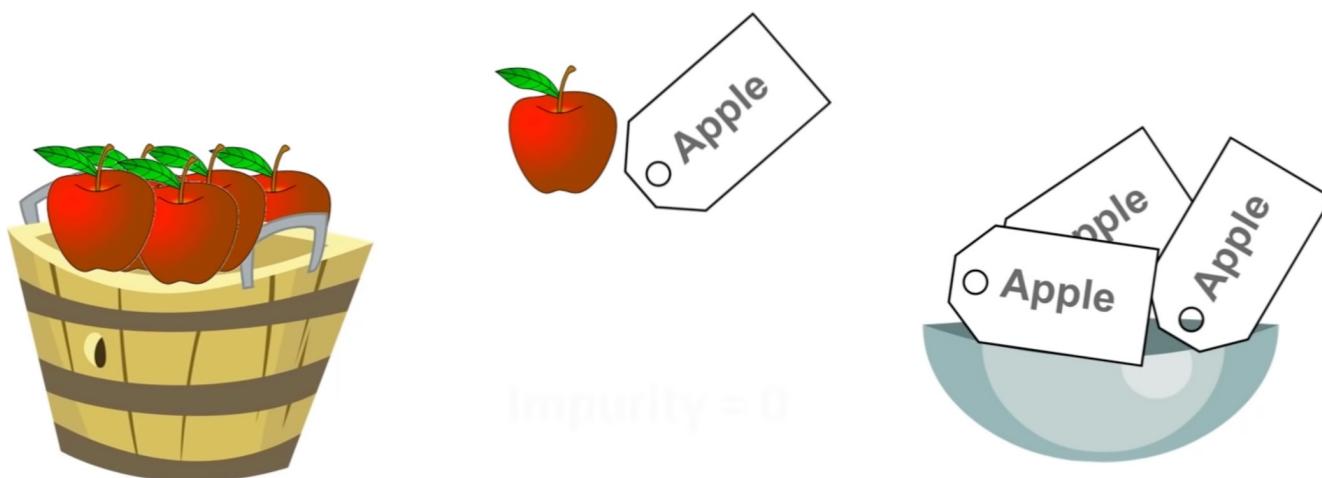
Information Gain

The information gain is the decrease in entropy after a dataset is split on the basis of an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain

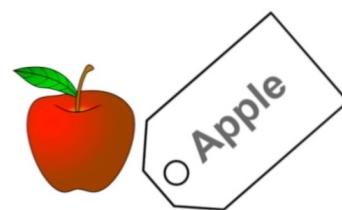
Reduction in Variance

Reduction in variance is an algorithm used for continuous target variables (regression problems). The split with lower variance is selected as the criteria to split the population

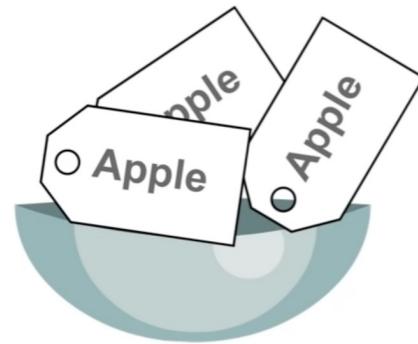
Let's First Understand What is Impurity



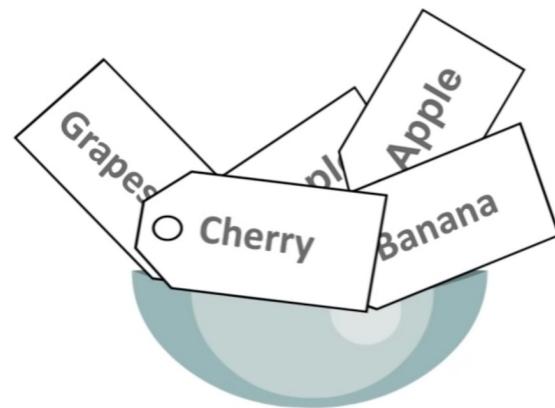
Let's First Understand What is Impurity



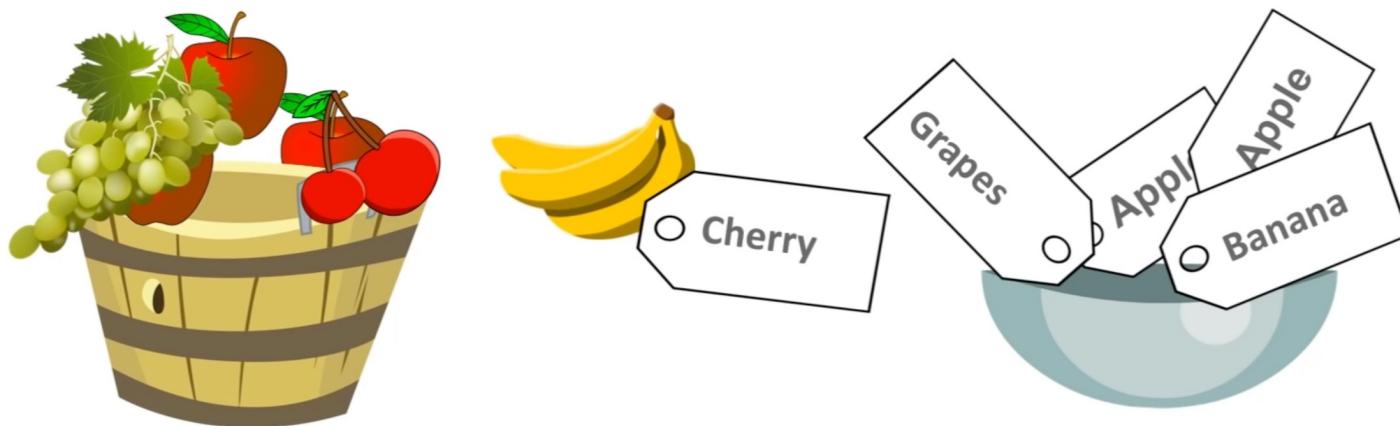
Impurity = 0



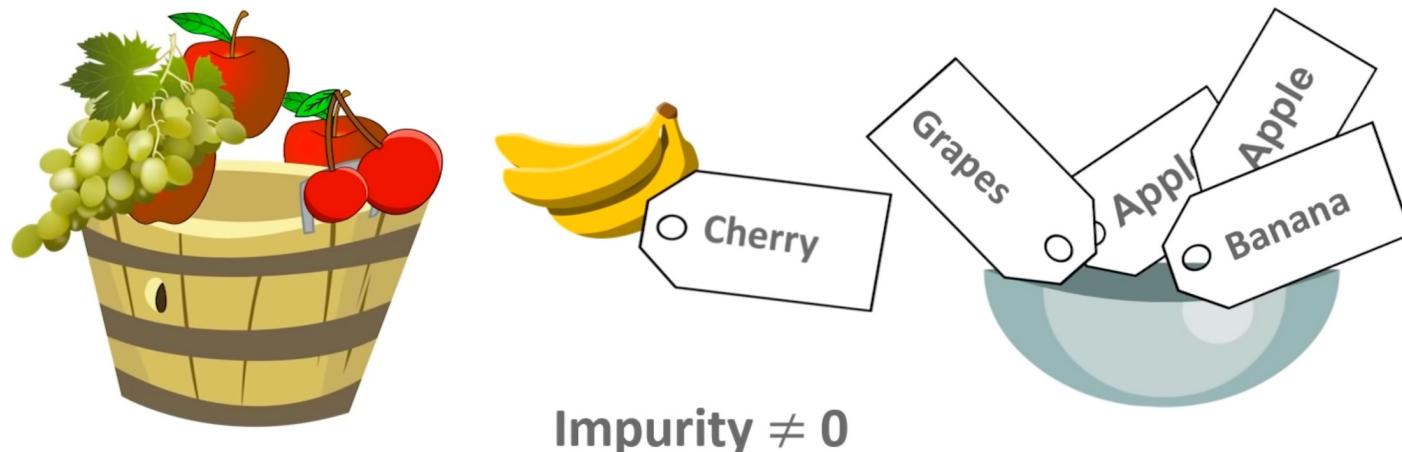
Let's First Understand What is Impurity



Let's First Understand What is Impurity

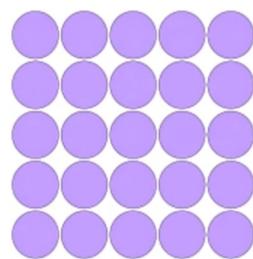


Let's First Understand What is Impurity

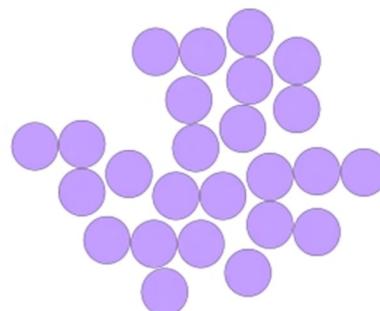


What is Entropy?

- Defines randomness in the data
- **Entropy** is just a metric which measures the impurity or
- The first step to solve the problem of a decision tree

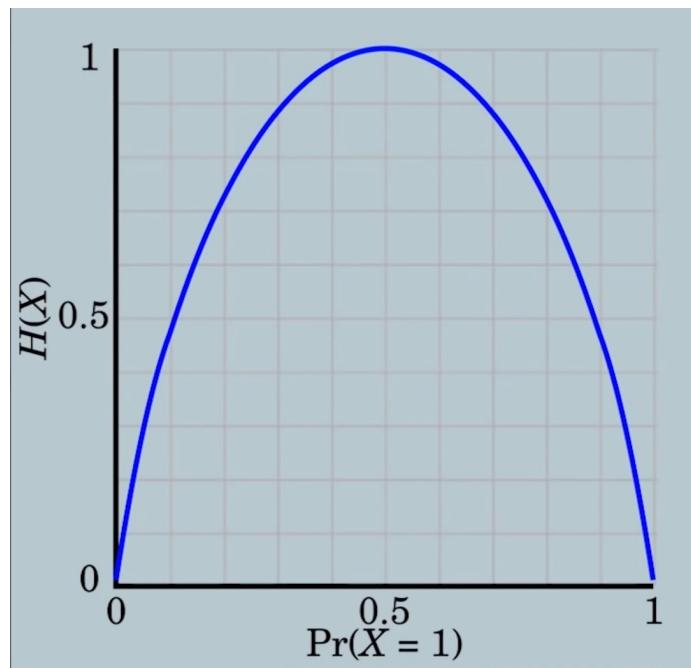


Low Entropy



High Entropy

What is Entropy?



$$\text{Entropy}(S) = -P(\text{yes}) \log_2 P(\text{yes}) - P(\text{no}) \log_2 P(\text{no})$$

Where,

- S is the total sample space,
- $P(\text{yes})$ is probability of yes

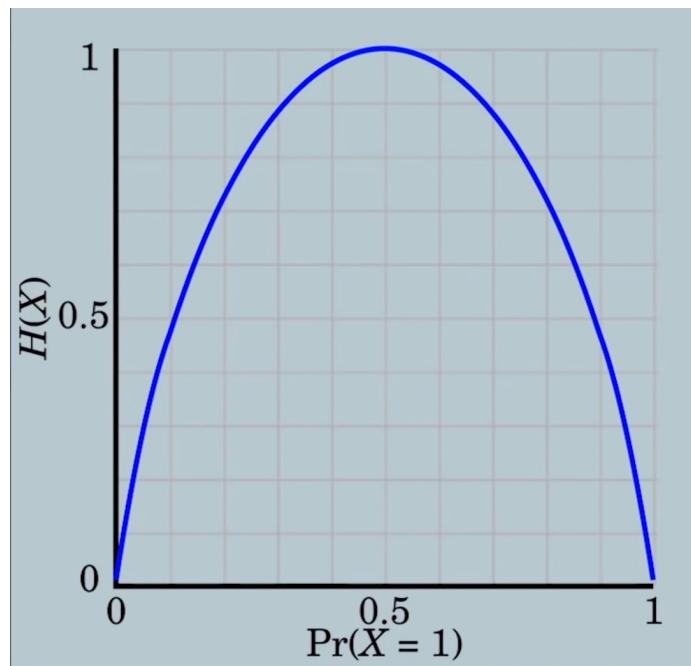
If number of yes = number of no ie $P(S) = 0.5$

$$\Rightarrow \text{Entropy}(S) = 1$$

If it contains all yes or all no ie $P(S) = 1$ or 0

$$\Rightarrow \text{Entropy}(S) = 0$$

What is Entropy?



$$\text{Entropy}(s) = -P(\text{yes}) \log_2 P(\text{yes}) - P(\text{no}) \log_2 P(\text{no})$$

Where,

- S is the total sample space,
- $P(\text{yes})$ is probability of yes

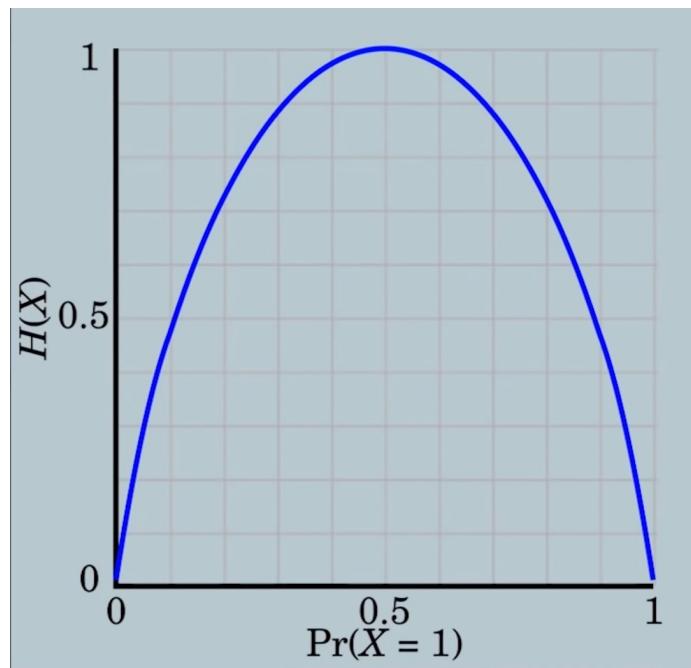
If number of yes = number of no ie $P(S) = 0.5$

$$\Rightarrow \text{Entropy}(s) = 1$$

If it contains all yes or all no ie $P(S) = 1$ or 0

$$\Rightarrow \text{Entropy}(s) = 0$$

What is Entropy?



$$E(S) = -P(\text{Yes}) \log_2 P(\text{Yes})$$

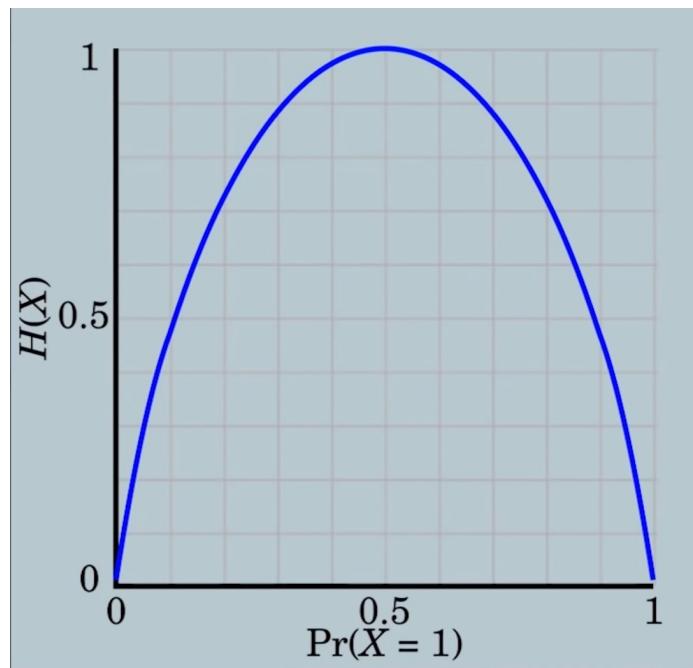
When $P(\text{Yes}) = P(\text{No}) = 0.5$ ie YES + NO = Total Sample(S)

$$E(S) = 0.5 \log_2 0.5 - 0.5 \log_2 0.5$$

$$E(S) = 0.5(\log_2 0.5 - \log_2 0.5)$$

$$E(S) = 1$$

What is Entropy?



$$E(S) = -P(\text{Yes}) \log_2 P(\text{Yes})$$

When $P(\text{Yes}) = 1$ ie YES = Total Sample(S)

$$E(S) = 1 \log_2 1$$

$$E(S) = 0$$

$$E(S) = -P(\text{No}) \log_2 P(\text{No})$$

When $P(\text{No}) = 1$ ie No = Total Sample(S)

$$E(S) = 1 \log_2 1$$

$$E(S) = 0$$

What is Information Gain?

- Measures the reduction in entropy
- Decides which attribute should be selected as the decision node

If S is our total collection,

Information Gain = Entropy(S) – [(Weighted Avg) x Entropy(each feature)]

Step 1: Compute the entropy for the Data set

Out of 14 instances we have 9 YES and 5 NO

So we have the formula,

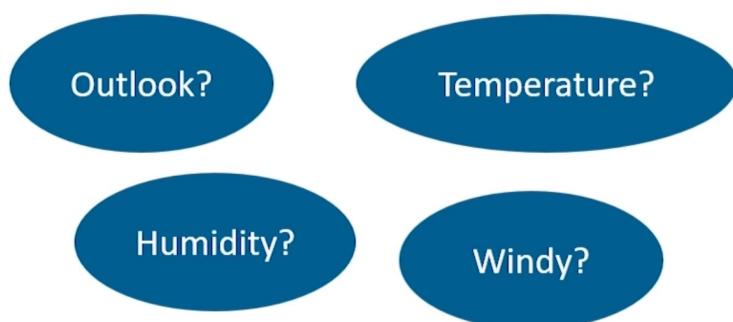
$$E(S) = -P(Yes) \log_2 P(Yes) - P(No) \log_2 P(No)$$

$$E(S) = - (9/14) * \log_2 9/14 - (5/14) * \log_2 5/14$$

$$E(S) = 0.41 + 0.53 = 0.94$$

| | outlook | temp. | humidity | windy | play |
|-----|----------|-------|----------|-------|------|
| D1 | sunny | hot | high | false | no |
| D2 | sunny | hot | high | true | no |
| D3 | overcast | hot | high | false | yes |
| D4 | rainy | mild | high | false | yes |
| D5 | rainy | cool | normal | false | yes |
| D6 | rainy | cool | normal | true | no |
| D7 | overcast | cool | normal | true | yes |
| D8 | sunny | mild | high | false | no |
| D9 | sunny | cool | normal | false | yes |
| D10 | rainy | mild | normal | false | yes |
| D11 | sunny | mild | normal | true | yes |
| D12 | overcast | mild | high | true | yes |
| D13 | overcast | hot | normal | false | yes |
| D14 | rainy | mild | high | true | no |

Which Node To Select As Root Node?



| outlook | temp. | 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 |



Machine Learning

Decision Trees

Phd. César Astudillo | Facultad de Ingeniería