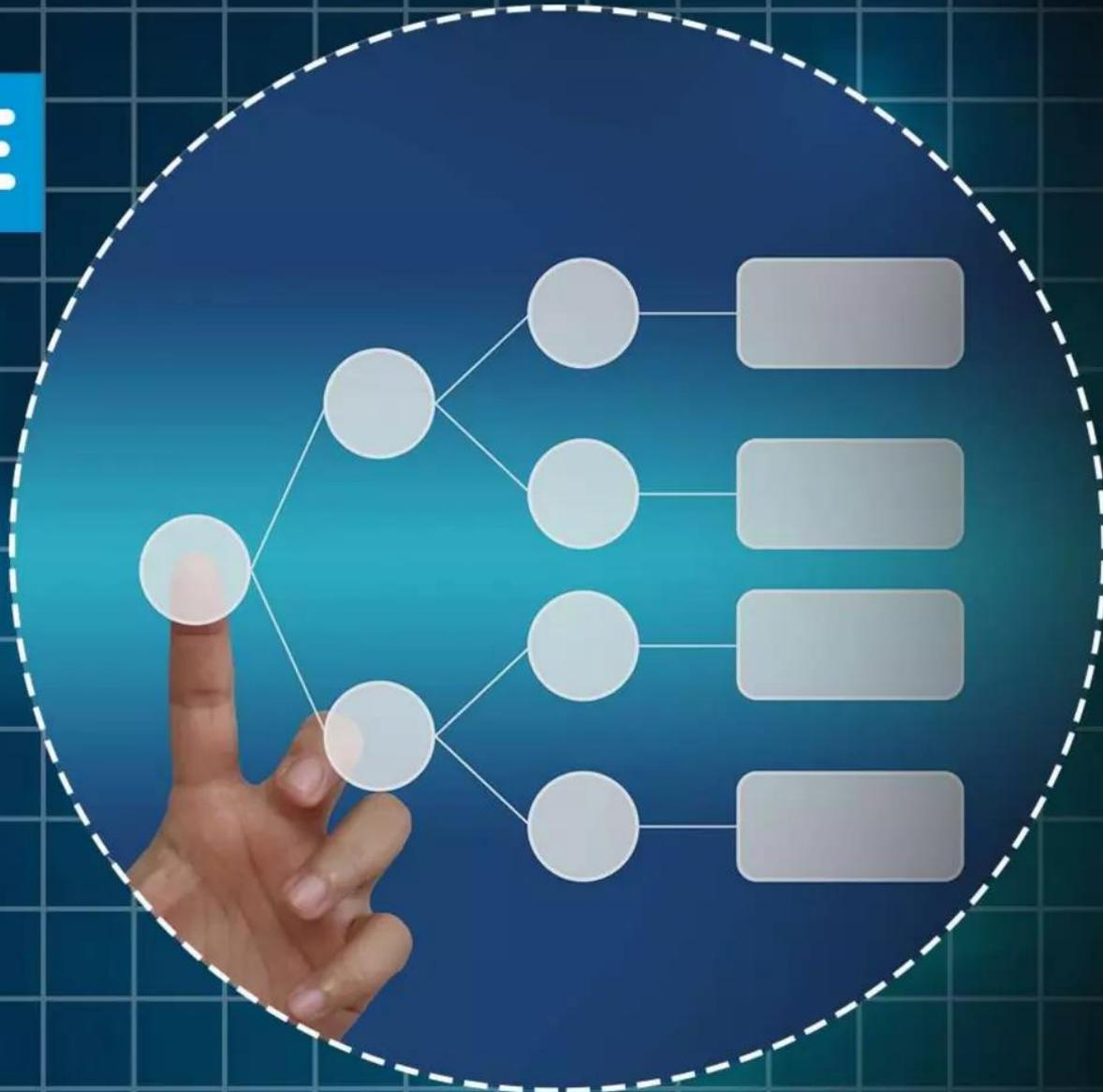


DECISION TREE TUTORIAL



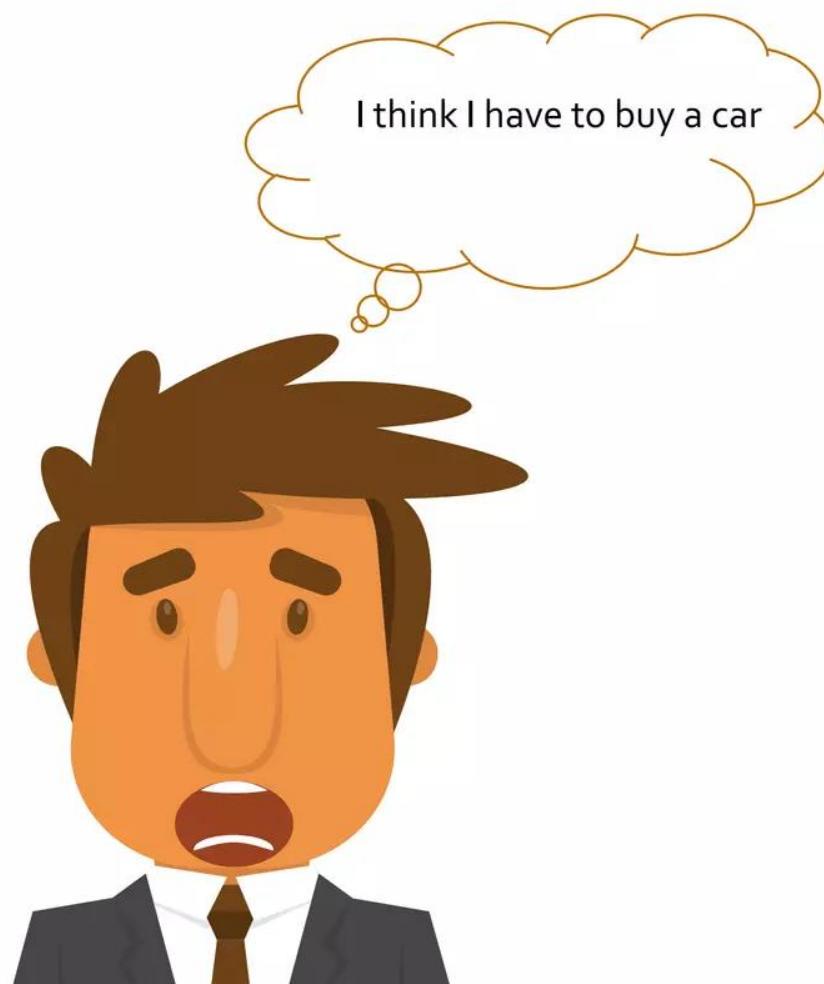
simplilearn



Decision Tree Tutorial



Decision Tree Tutorial



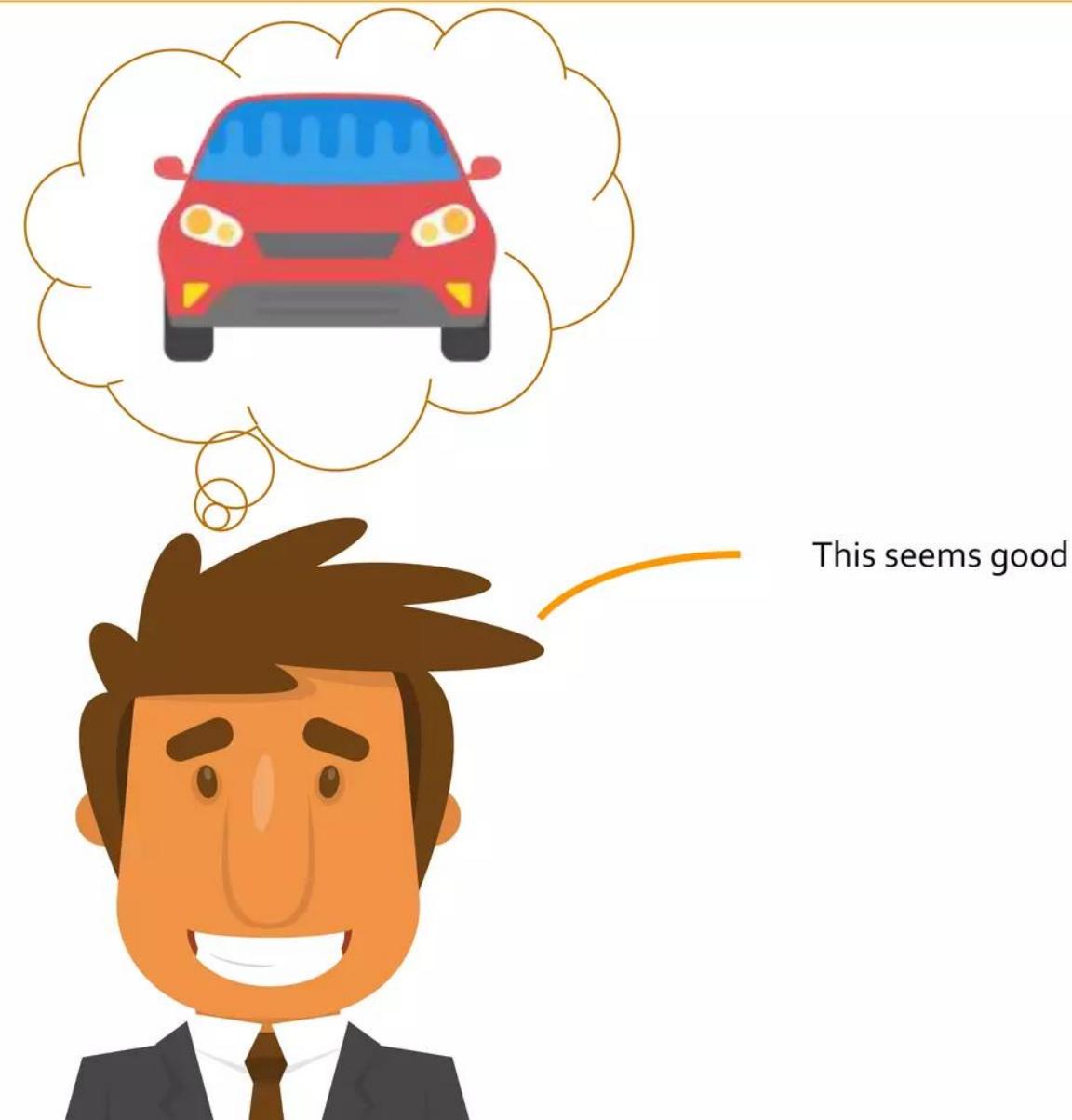
Decision Tree Tutorial



Decision Tree Tutorial



Decision Tree Tutorial

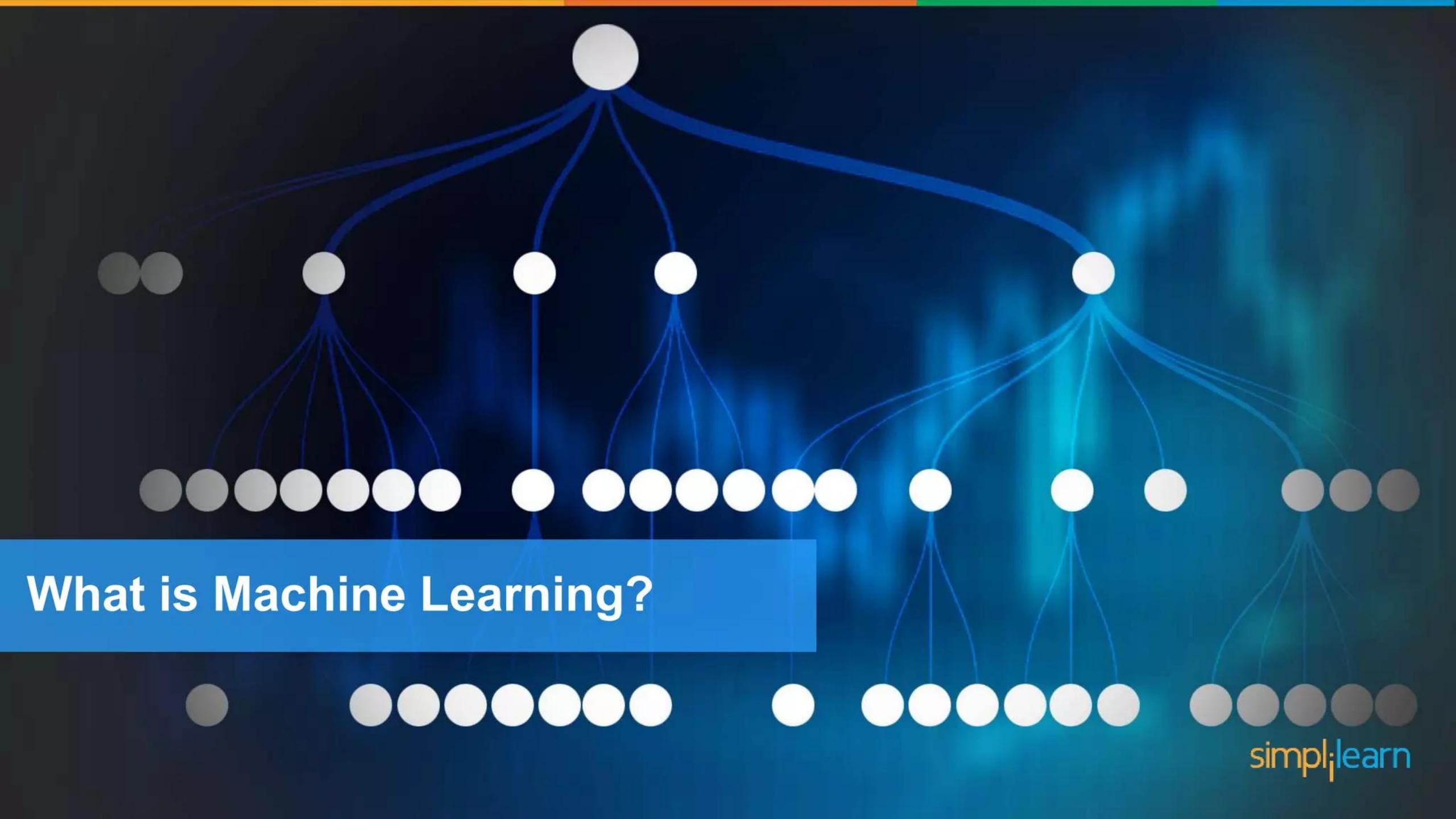


This seems good

What's in it for you?

- ▶ What is Machine Learning?
- ▶ Types of Machine Learning
- ▶ Problems in Machine Learning
- ▶ What is Decision Tree?
- ▶ What are the problems a Decision Tree solves?
- ▶ Advantages of Decision Tree
- ▶ Disadvantages of Decision Tree
- ▶ How does Decision Tree work?
- ▶ Use Case – Loan repayment prediction





What is Machine Learning?

What is Machine Learning?



What is Machine Learning?



What is Machine Learning?

Artificial Intelligence



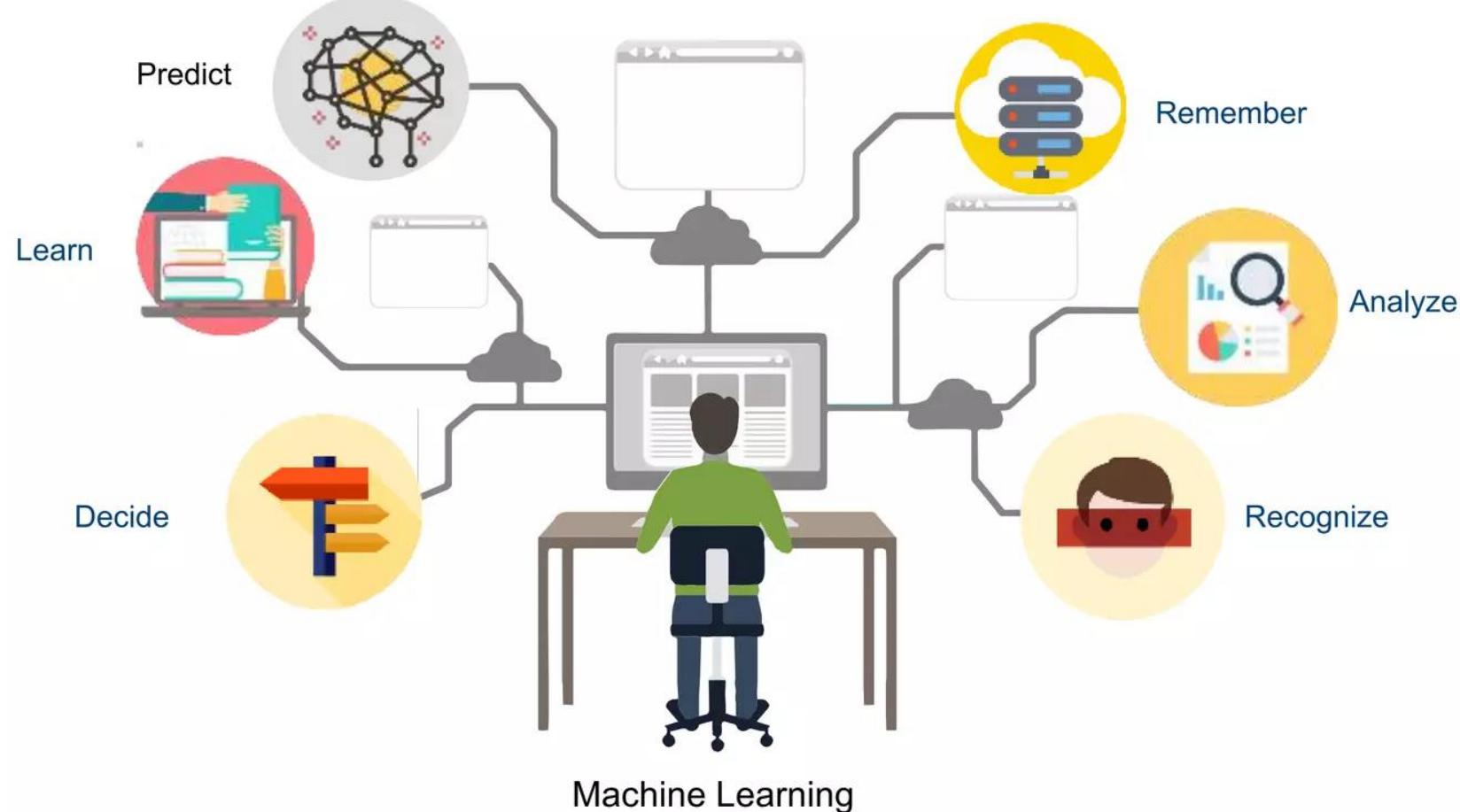
What is Machine Learning?



What is Machine Learning?



What is Machine Learning?



What is Machine Learning?

Machine Learning is an application of Artificial Intelligence wherein the system gets the ability to automatically learn and improve based on experience



Ordinary system

What is Machine Learning?

Machine Learning is an application of Artificial Intelligence wherein the system gets the ability to automatically learn and improve based on experience



Ordinary system



With Artificial
Intelligence

What is Machine Learning?

Machine Learning is an application of Artificial Intelligence wherein the system gets the ability to automatically learn and improve based on experience



Ordinary system



Ability to learn and improve on its own

What is Machine Learning?

Machine Learning is an application of Artificial Intelligence wherein the system gets the ability to automatically learn and improve based on experience



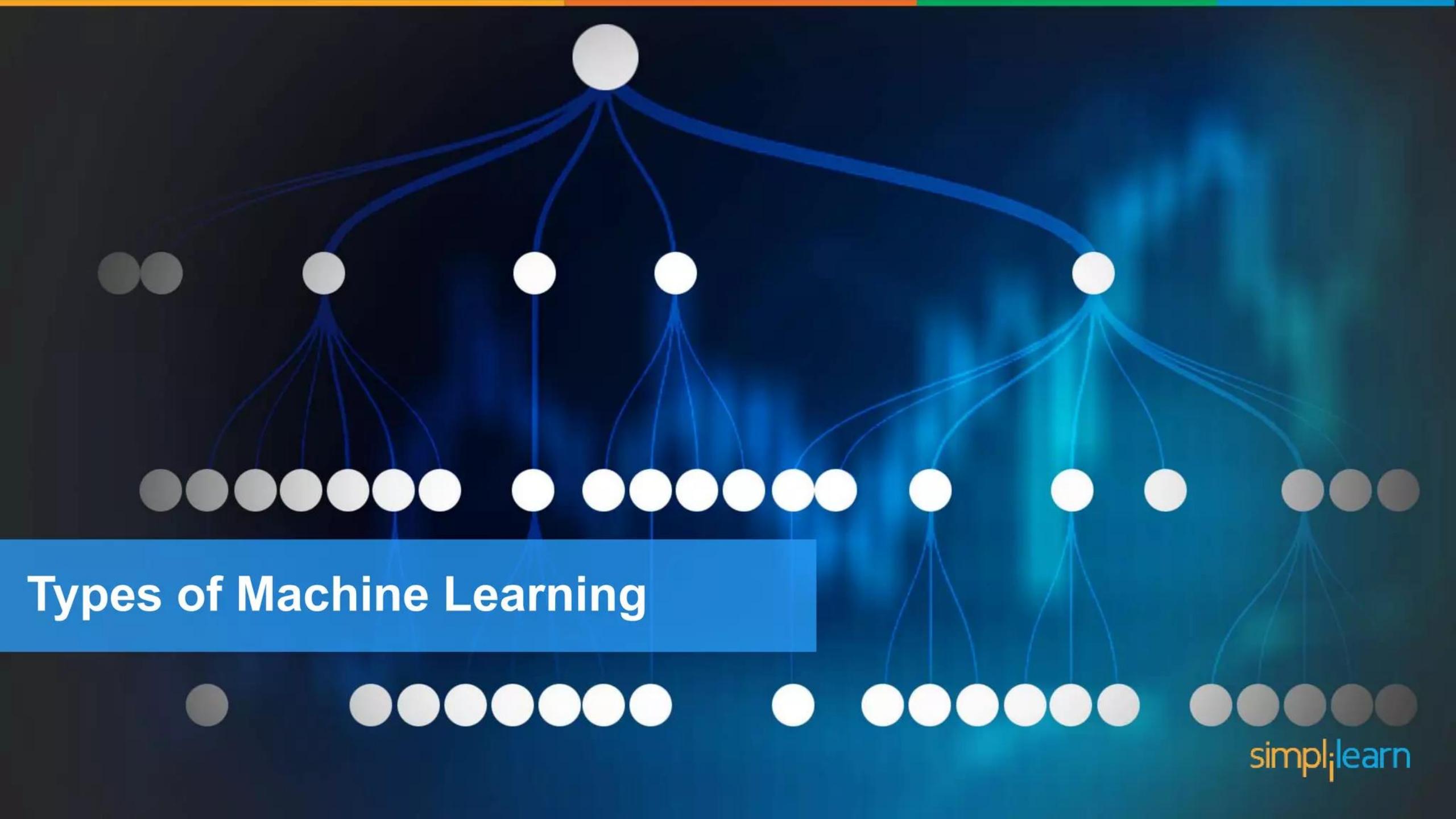
Ordinary system



Ability to learn and improve on its own



Machine Learning



Types of Machine Learning

Types of Machine Learning



Supervised Learning

Types of Machine Learning



Supervised Learning



Unsupervised Learning

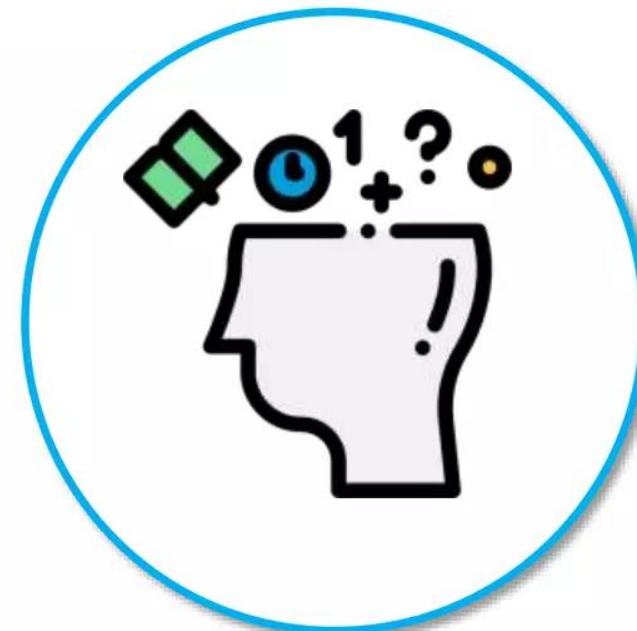
Types of Machine Learning



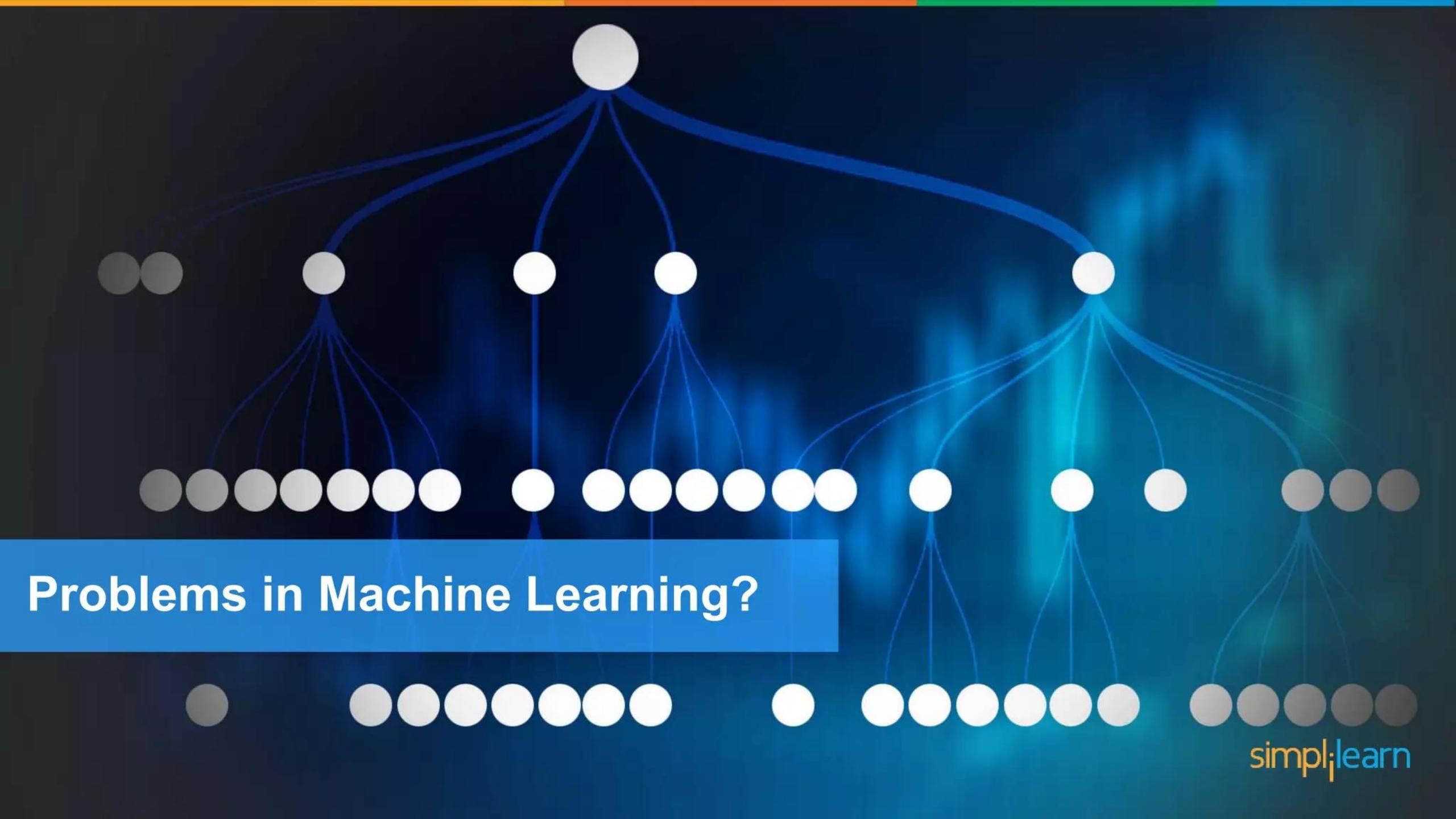
Supervised Learning



Unsupervised Learning

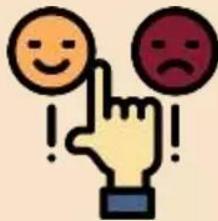


Reinforcement Learning



Problems in Machine Learning?

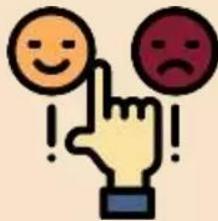
Problems in Machine Learning



Classification

Problems with categorical solutions like 'Yes' or 'No', 'True' or 'False', '1' or '0'

Problems in Machine Learning



Classification

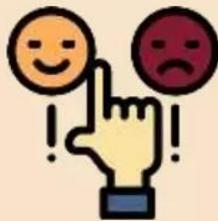
Problems with categorical solutions like 'Yes' or 'No', 'True' or 'False', '1' or '0'



Regression

Problems wherein continuous value needs to be predicted like 'Product Prices', 'Profit'

Problems in Machine Learning



Classification

Problems with categorical solutions like 'Yes' or 'No', 'True' or 'False', '1' or '0'



Regression

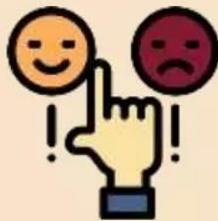
Problems wherein continuous value needs to be predicted like 'Product Prices', 'Profit'



Clustering

Problems wherein the data needs to be organized to find specific patterns like in the case of 'Product Recommendation'

Problems in Machine Learning



Classification

Problems with categorical solutions like 'Yes' or 'No', 'True' or 'False', '1' or '0'



Regression

Problems wherein continuous value needs to be predicted like 'Product Prices', 'Profit'



Clustering

Problems wherein the data needs to be organized to find specific patterns like in the case of 'Product Recommendation'

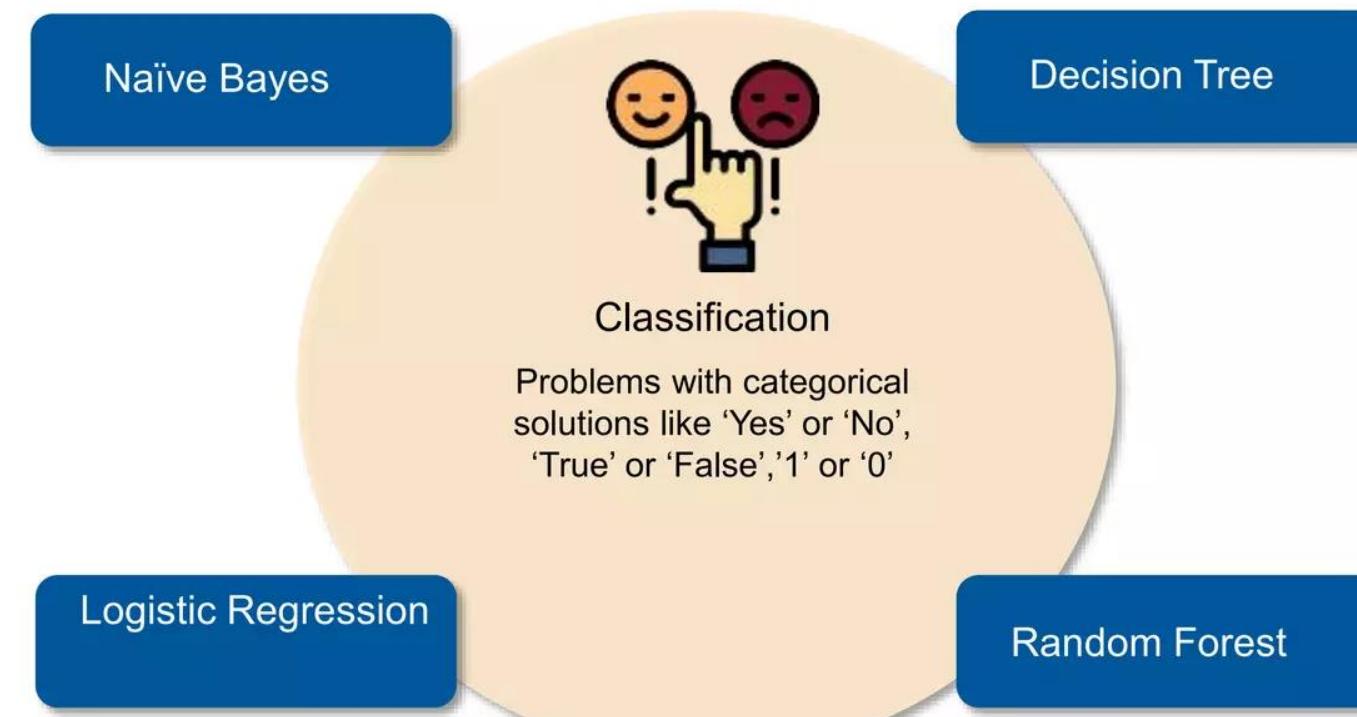
Problems in Machine Learning



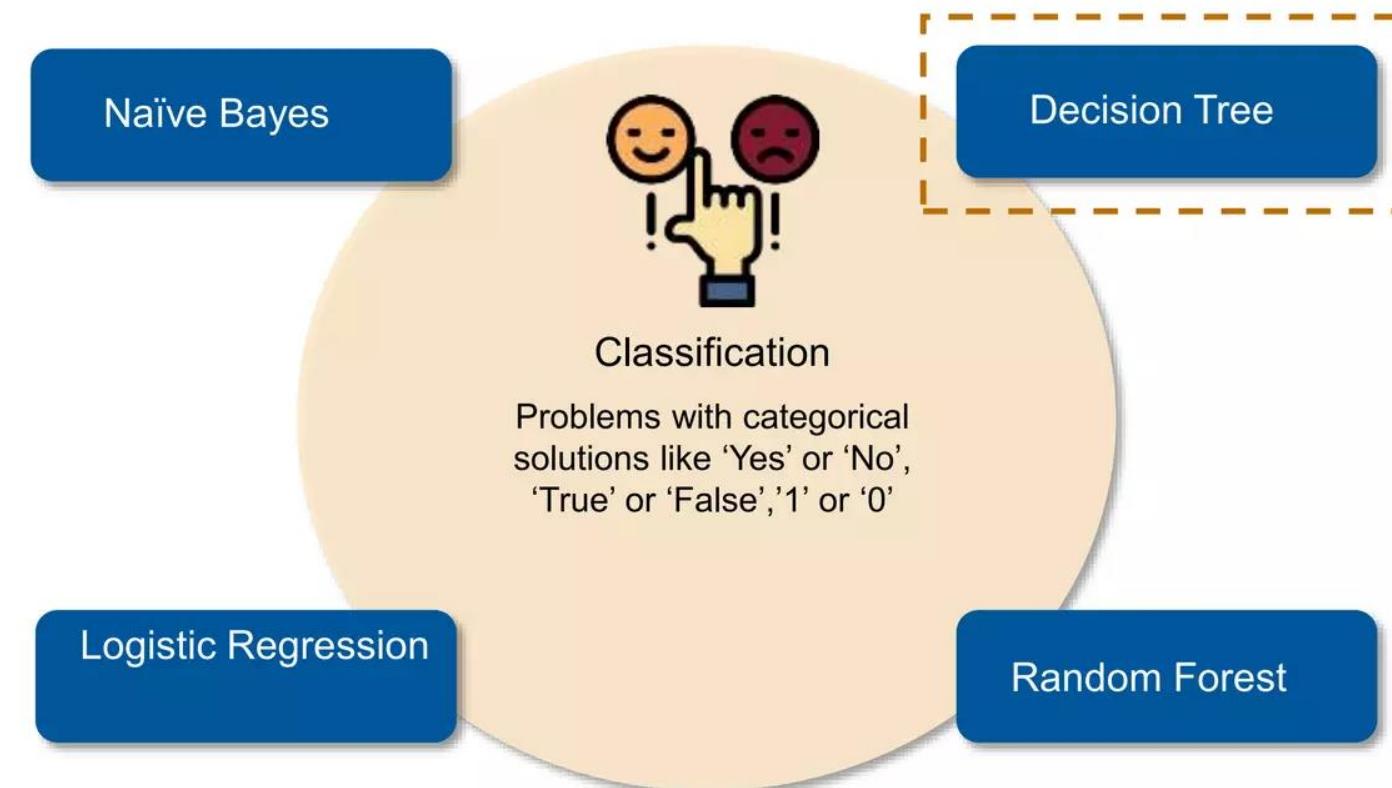
Classification

Problems with categorical solutions like 'Yes' or 'No', 'True' or 'False', '1' or '0'

Problems in Machine Learning



Problems in Machine Learning





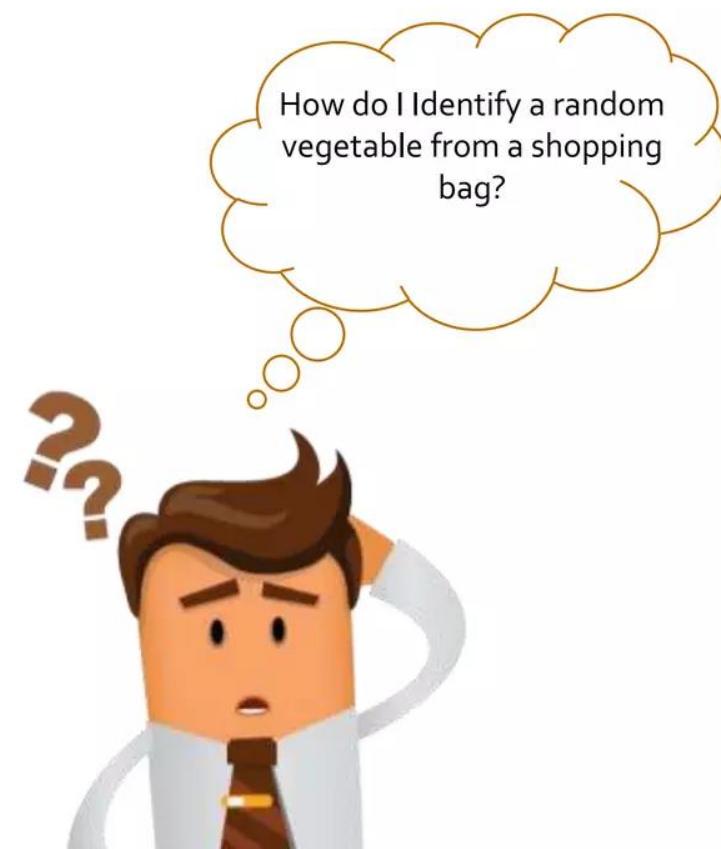
What is Decision Tree?

What is 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

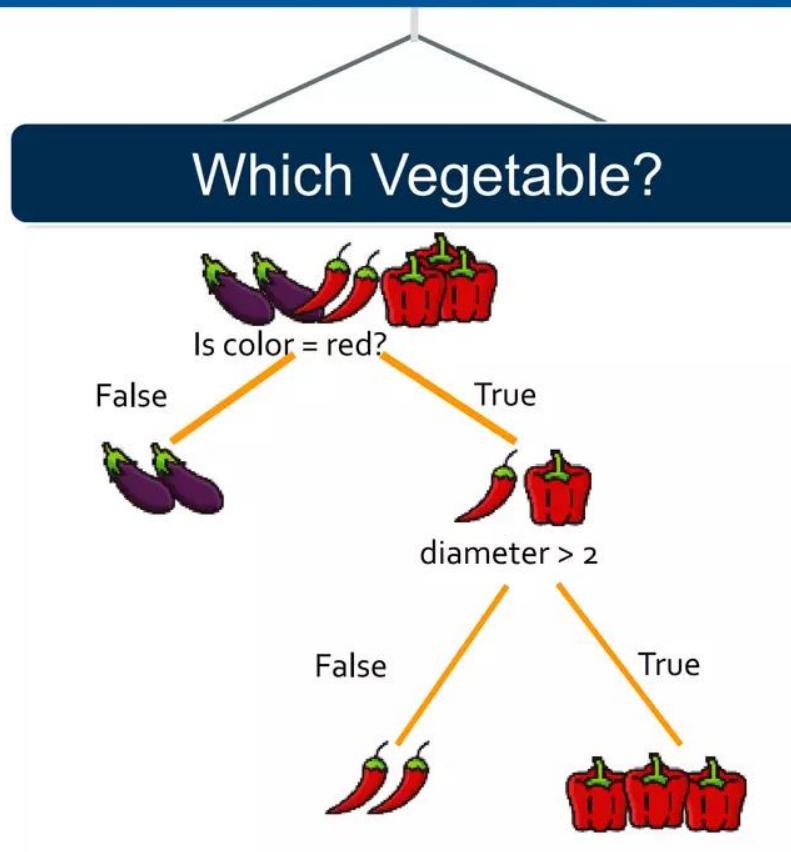
What is Decision Tree?

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



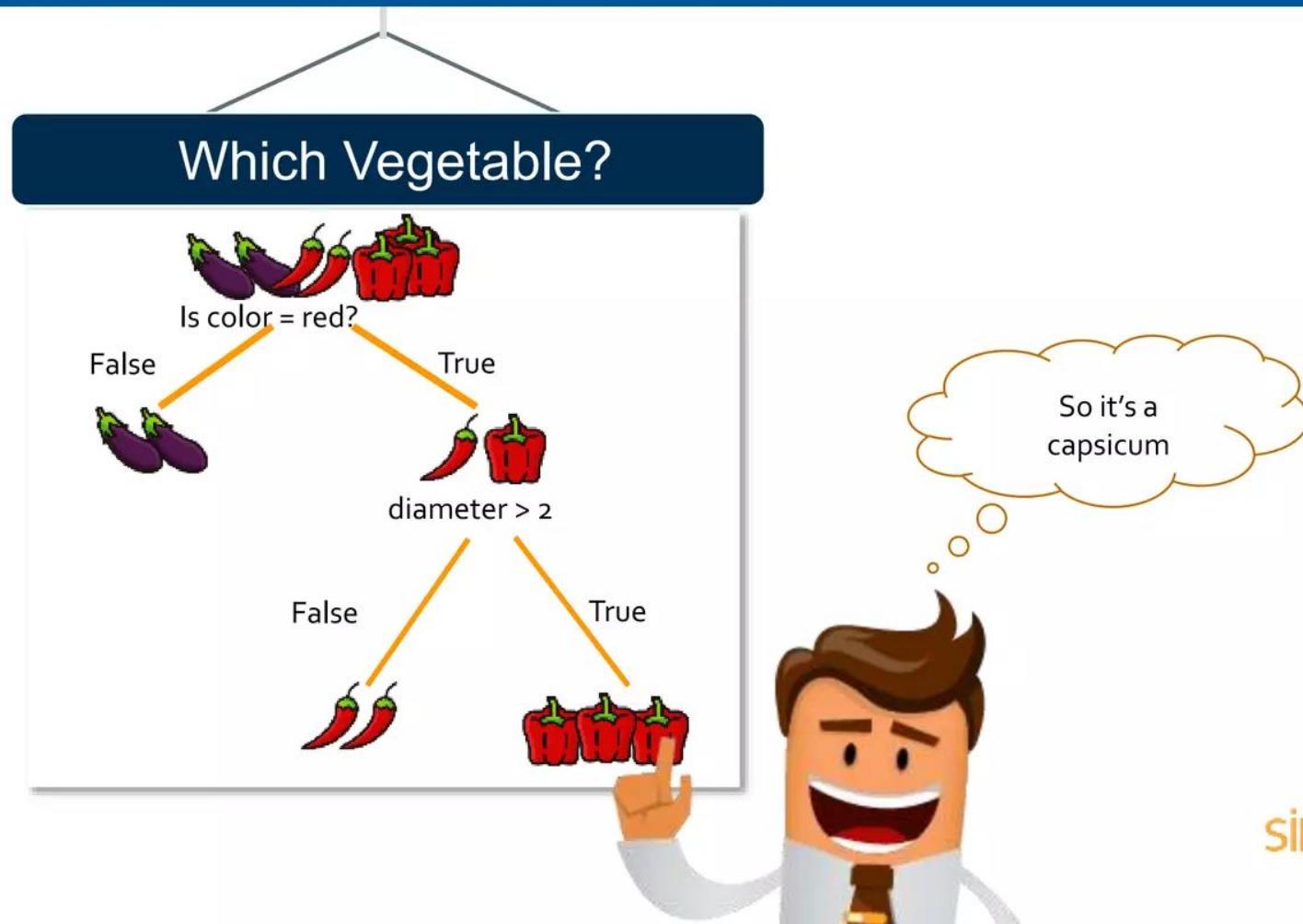
What is Decision Tree?

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 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





Problems that Decision Tree can solve

Problems that Decision Tree can solve

Classification

Regression



Problems that Decision Tree can solve



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

Problems that Decision Tree can solve

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



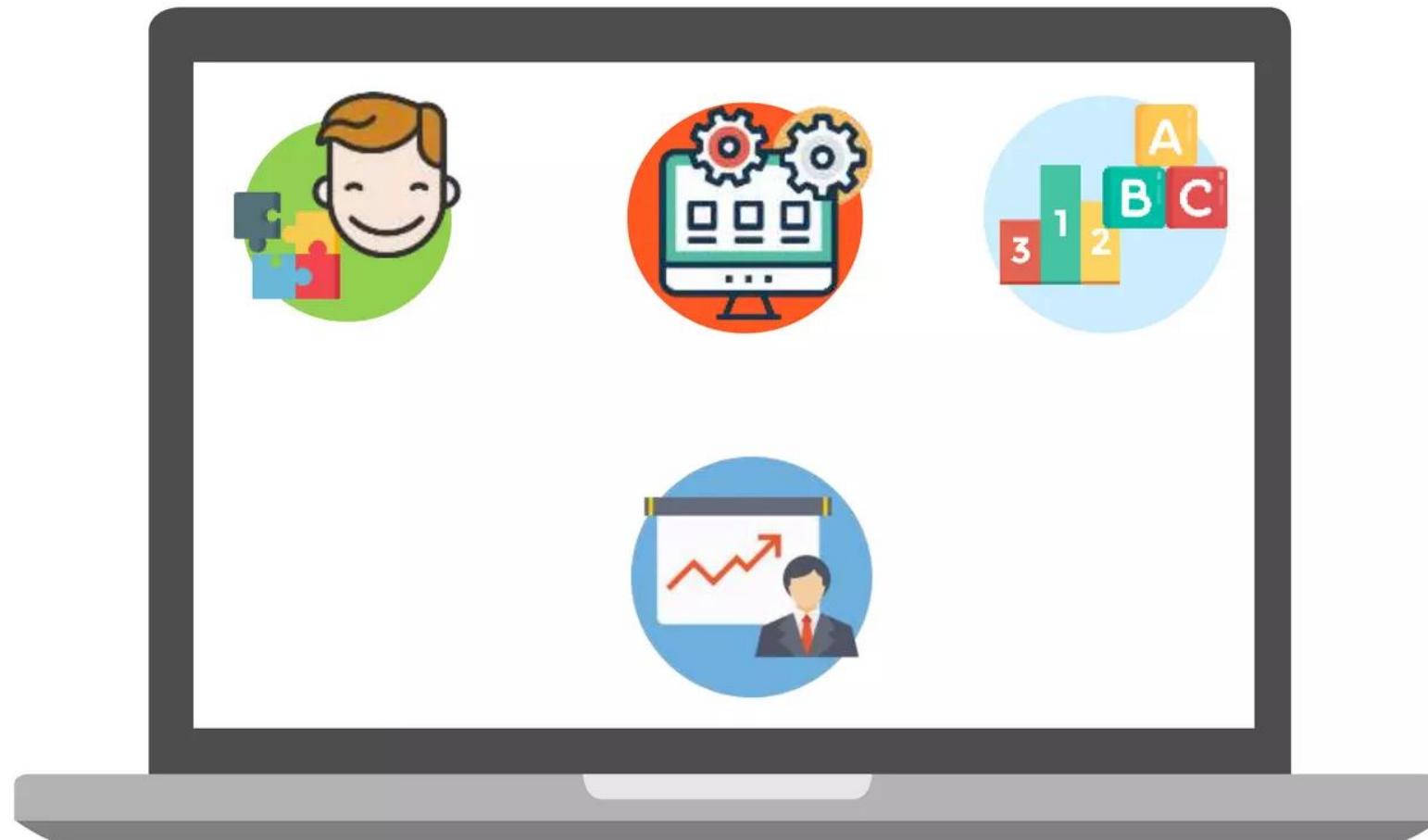
Regression

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.

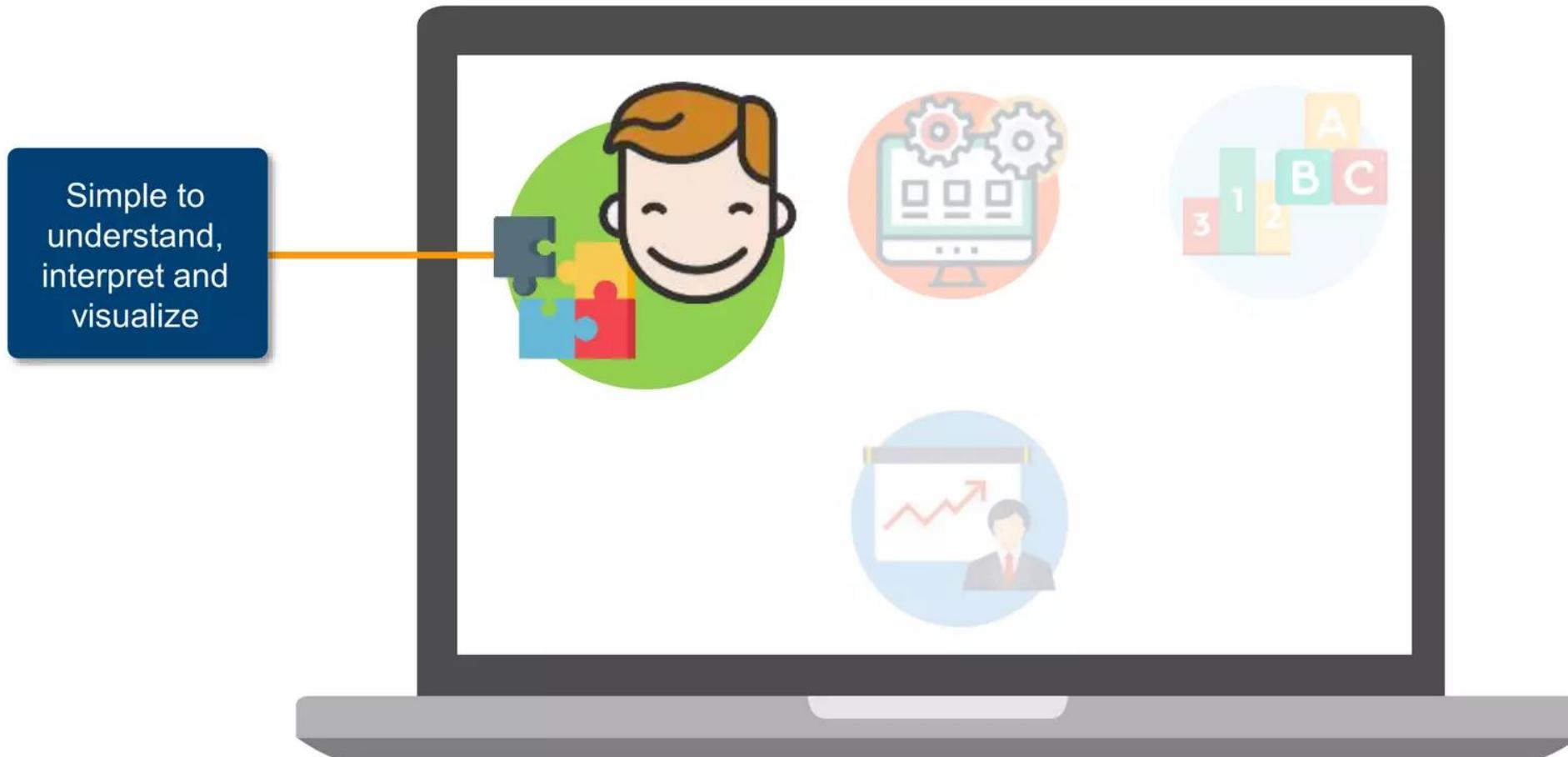


Advantages of Decision tree

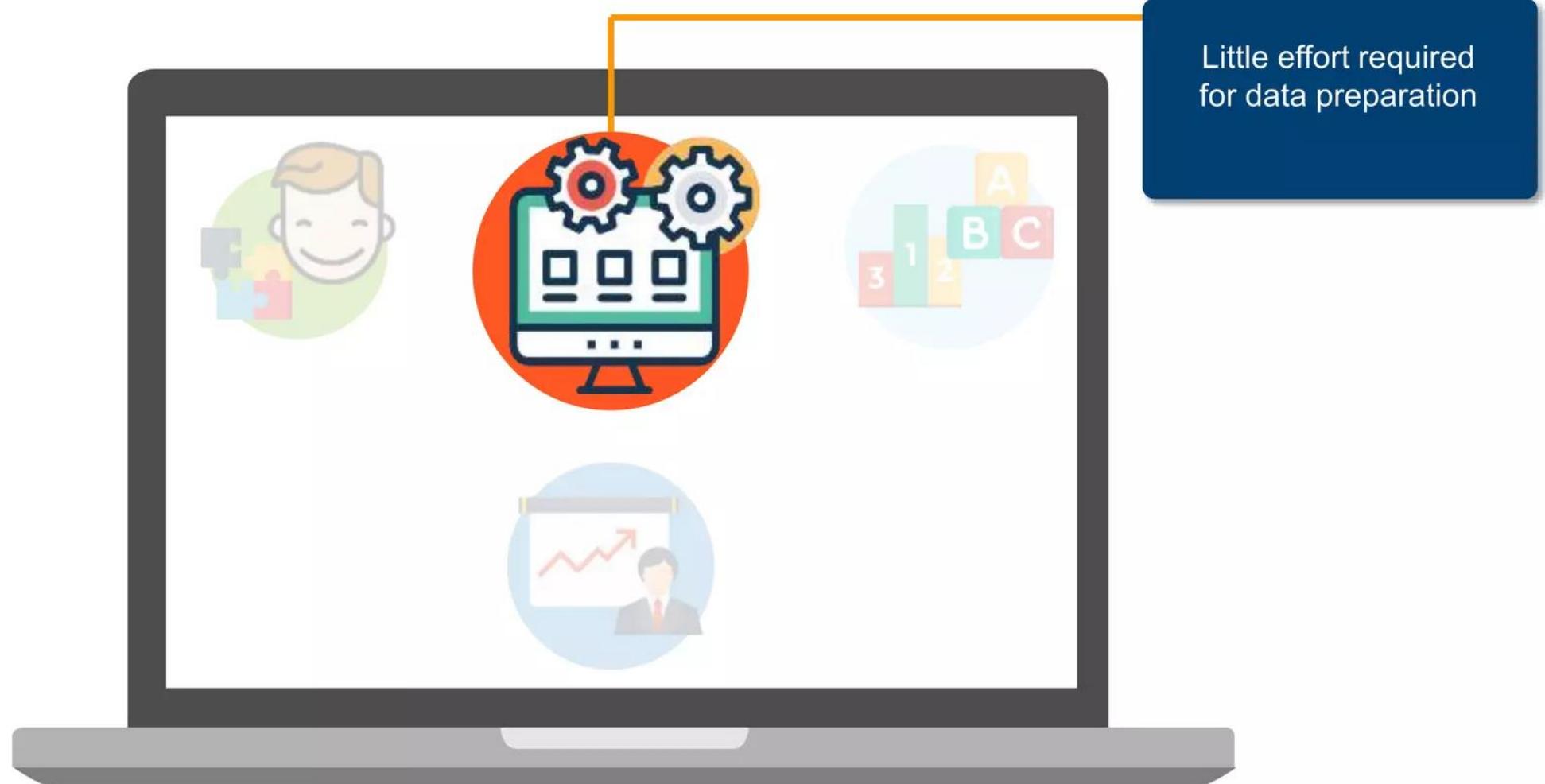
Advantages of Decision Tree



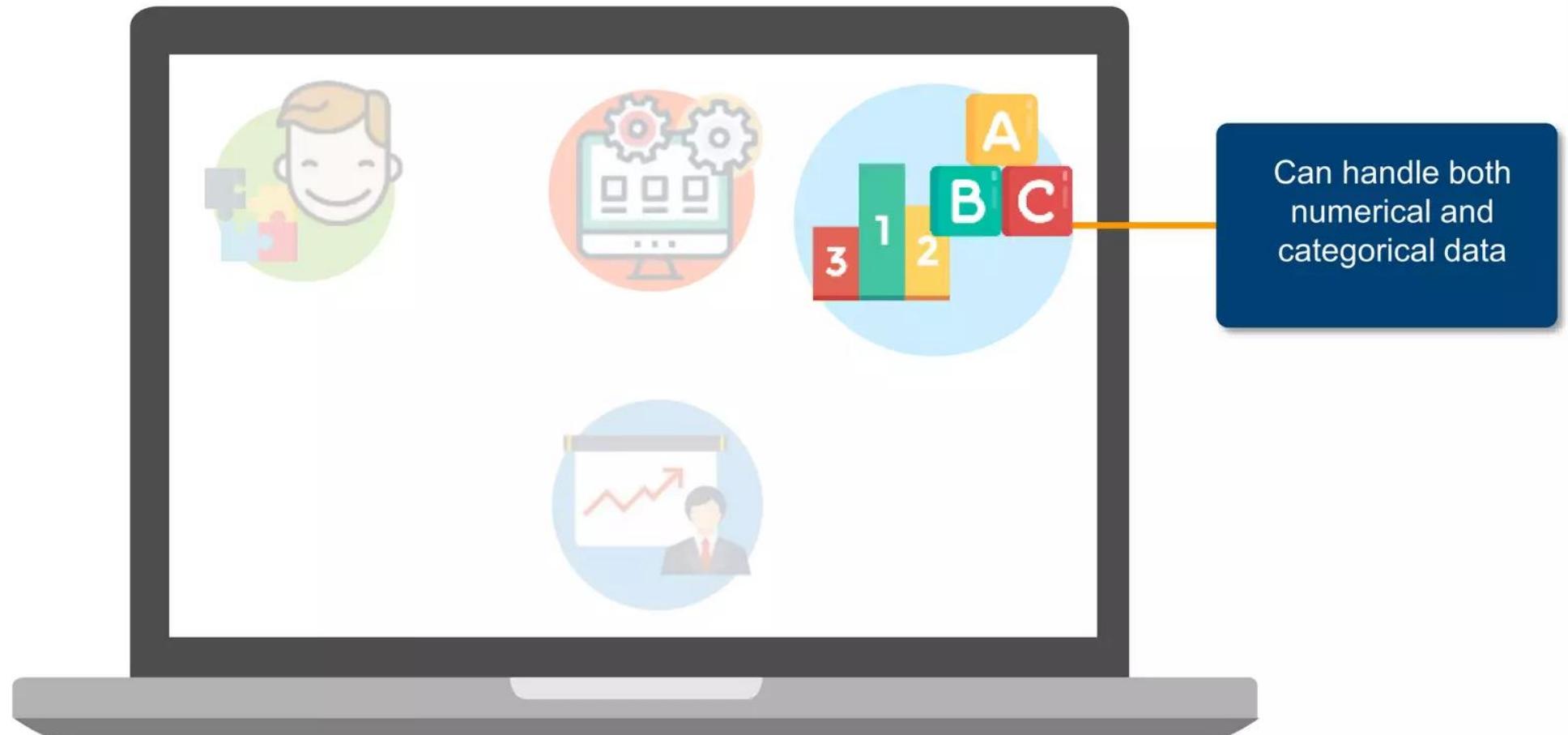
Advantages of Decision Tree



Advantages of Decision Tree



Advantages of Decision Tree



Advantages of Decision Tree





Disadvantages of Decision Tree

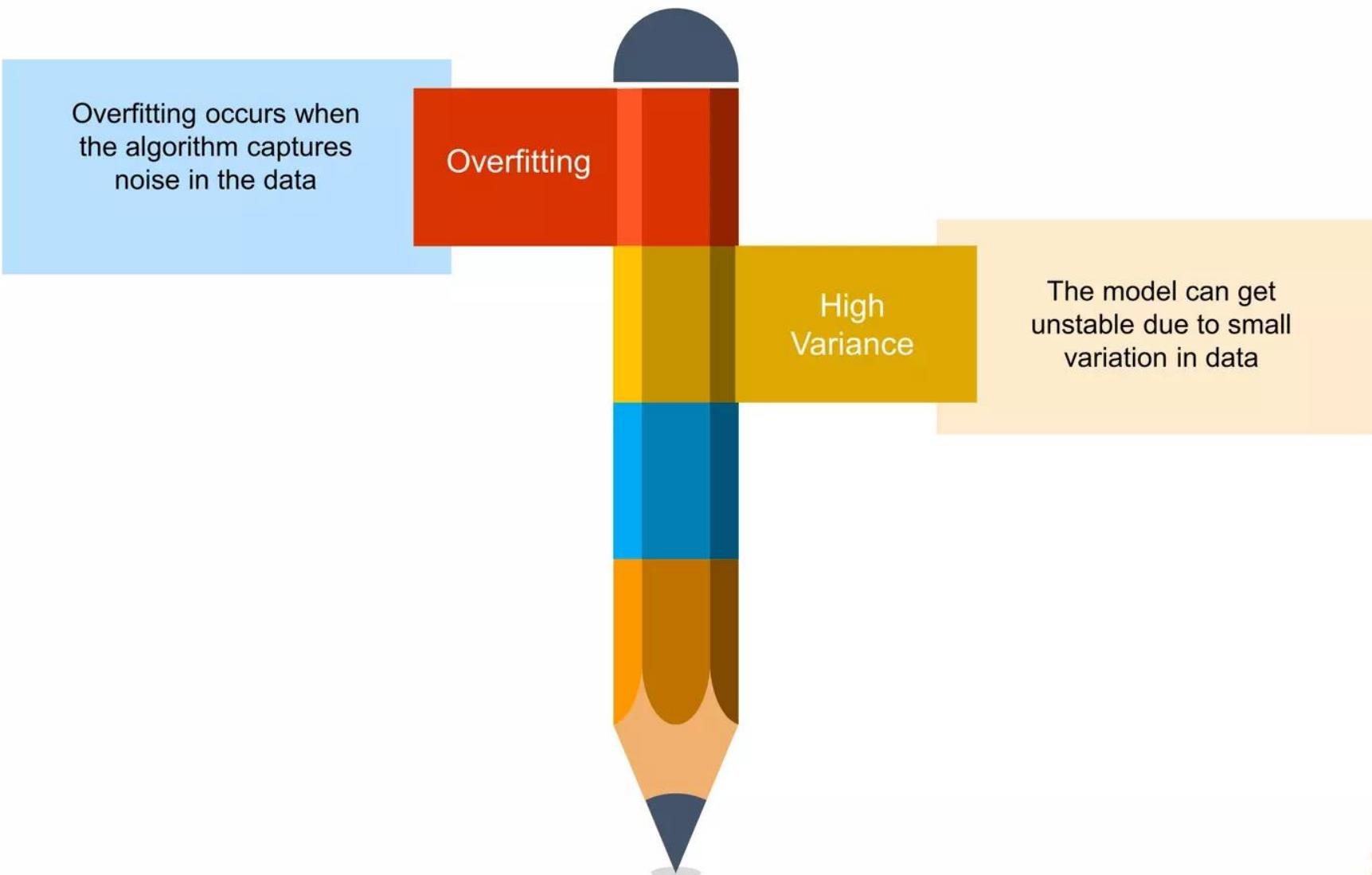
Disadvantages of Decision Tree

Overfitting occurs when
the algorithm captures
noise in the data

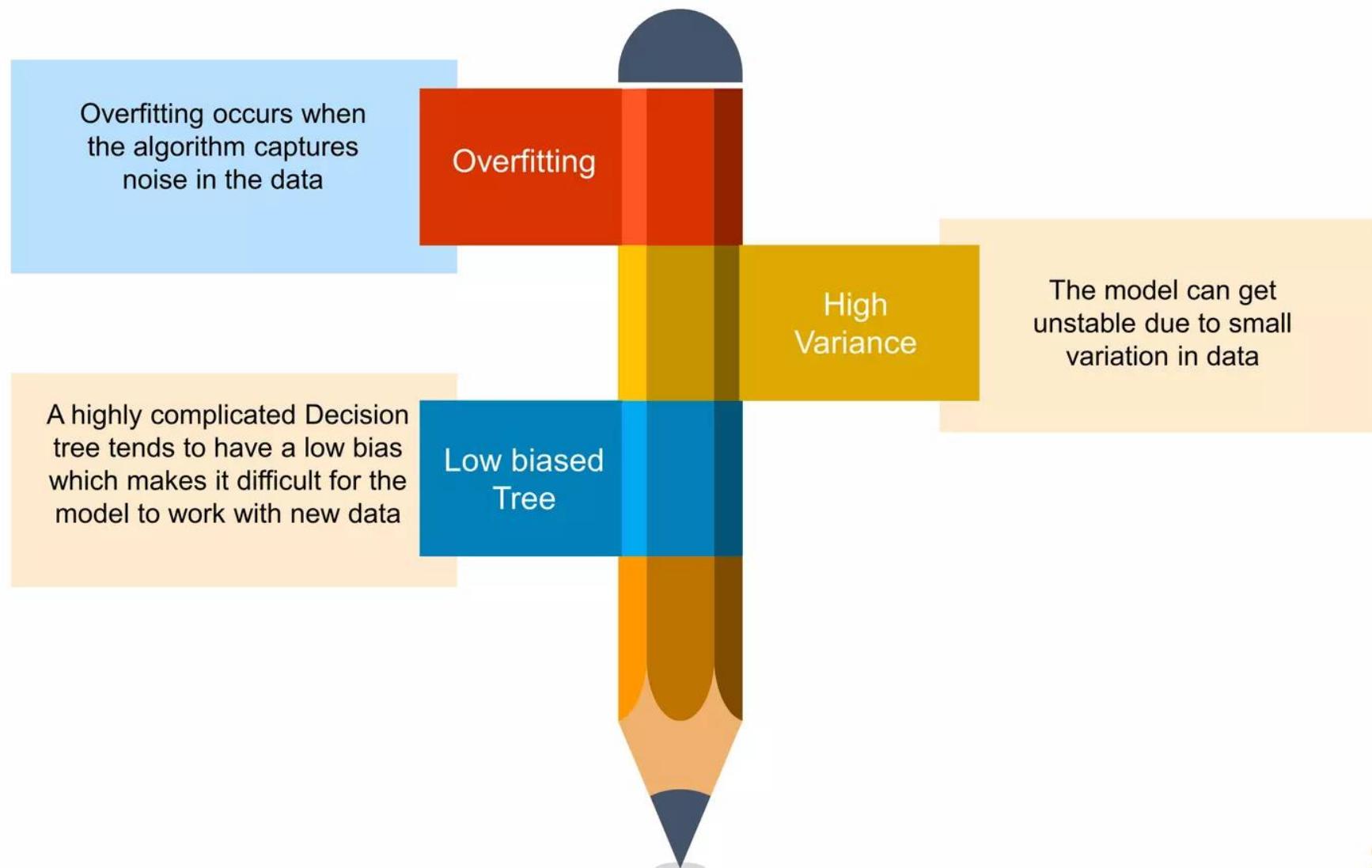
Overfitting



Disadvantages of Decision Tree



Disadvantages of Decision Tree





Decision Tree – Important Terms

Decision Tree – Important Terms



Decision Tree – Important Terms

Entropy

Entropy is the measure of randomness or unpredictability in the dataset

Example



High entropy

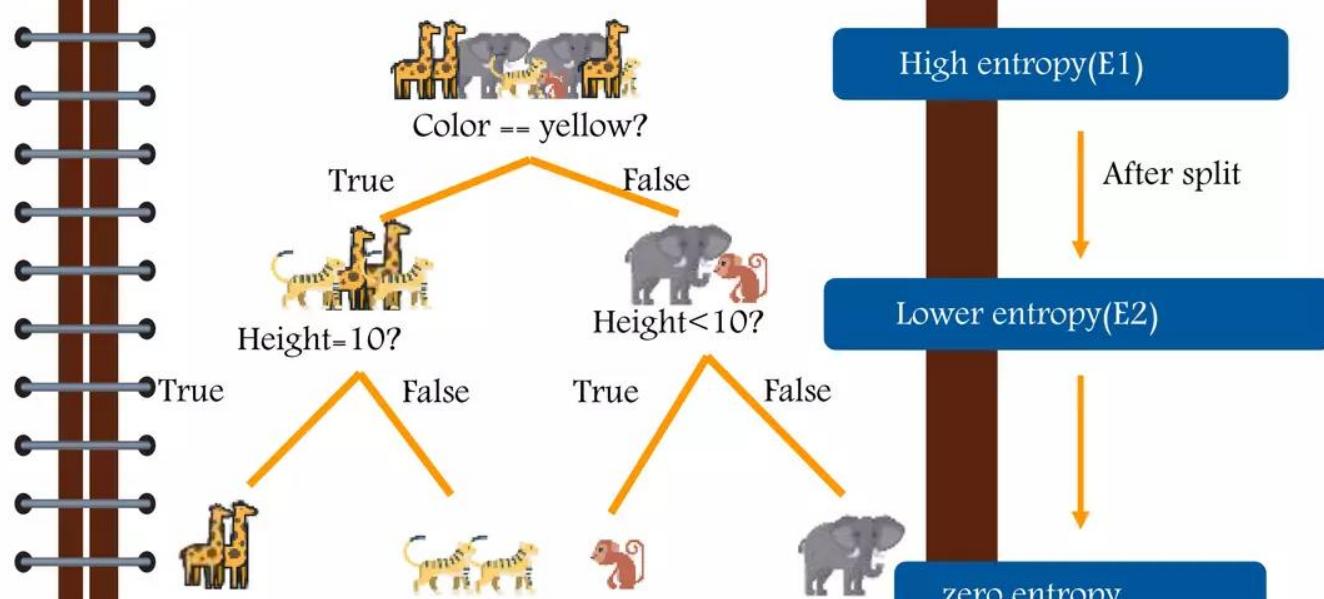
This Dataset has a very high entropy

Decision Tree – Important Terms

Entropy

Entropy is the measure of randomness or unpredictability in the dataset

Example

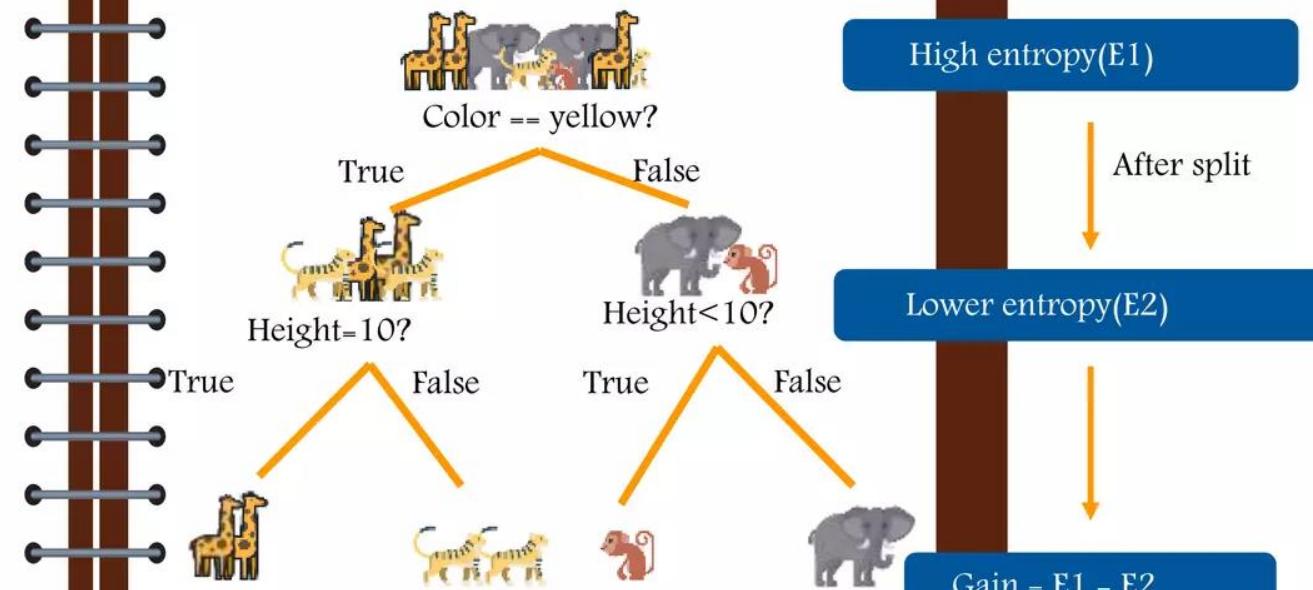


Decision Tree – Important Terms

Information gain

It is the measure of decrease in entropy after the dataset is split

Example

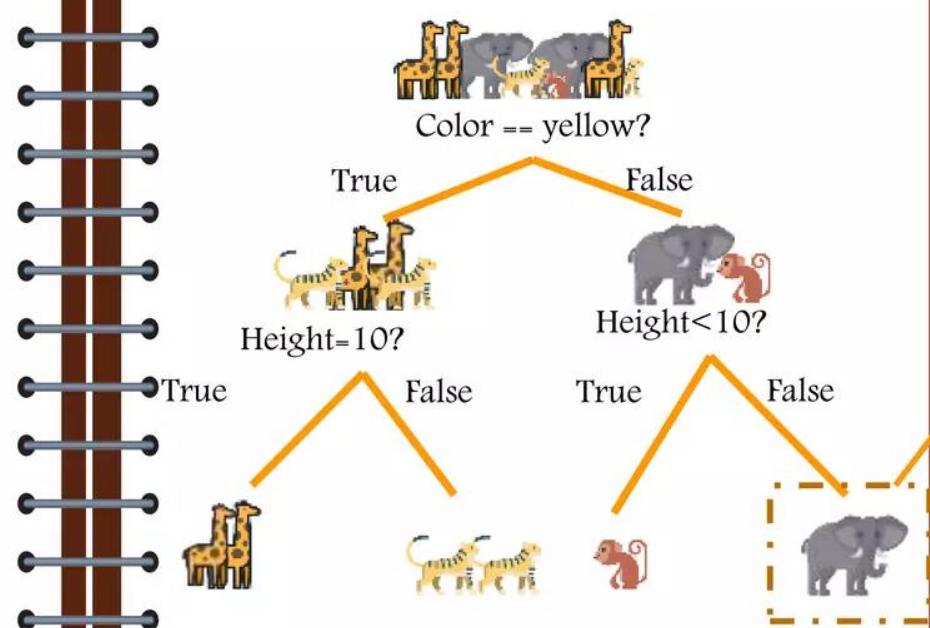


Decision Tree – Important Terms

Leaf Node

Leaf node carries the classification or the decision

Example



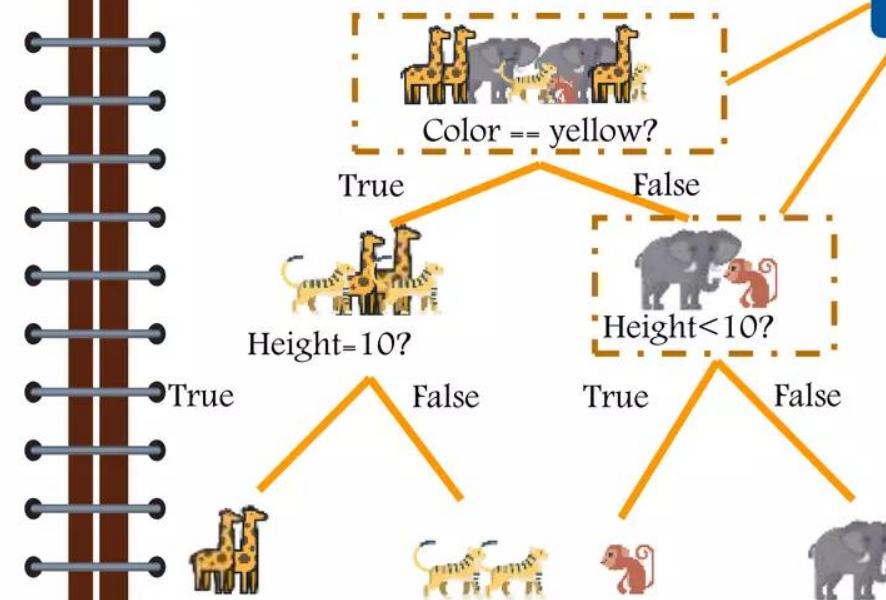
Leaf Node

Decision Tree – Important Terms

Decision Node

Decision node has two or more branches

Example



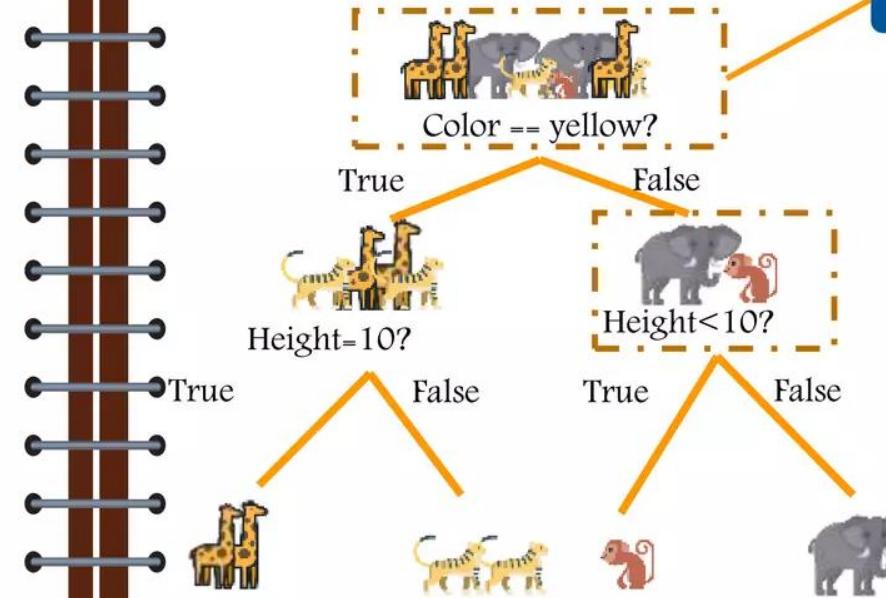
decision Node

Decision Tree – Important Terms

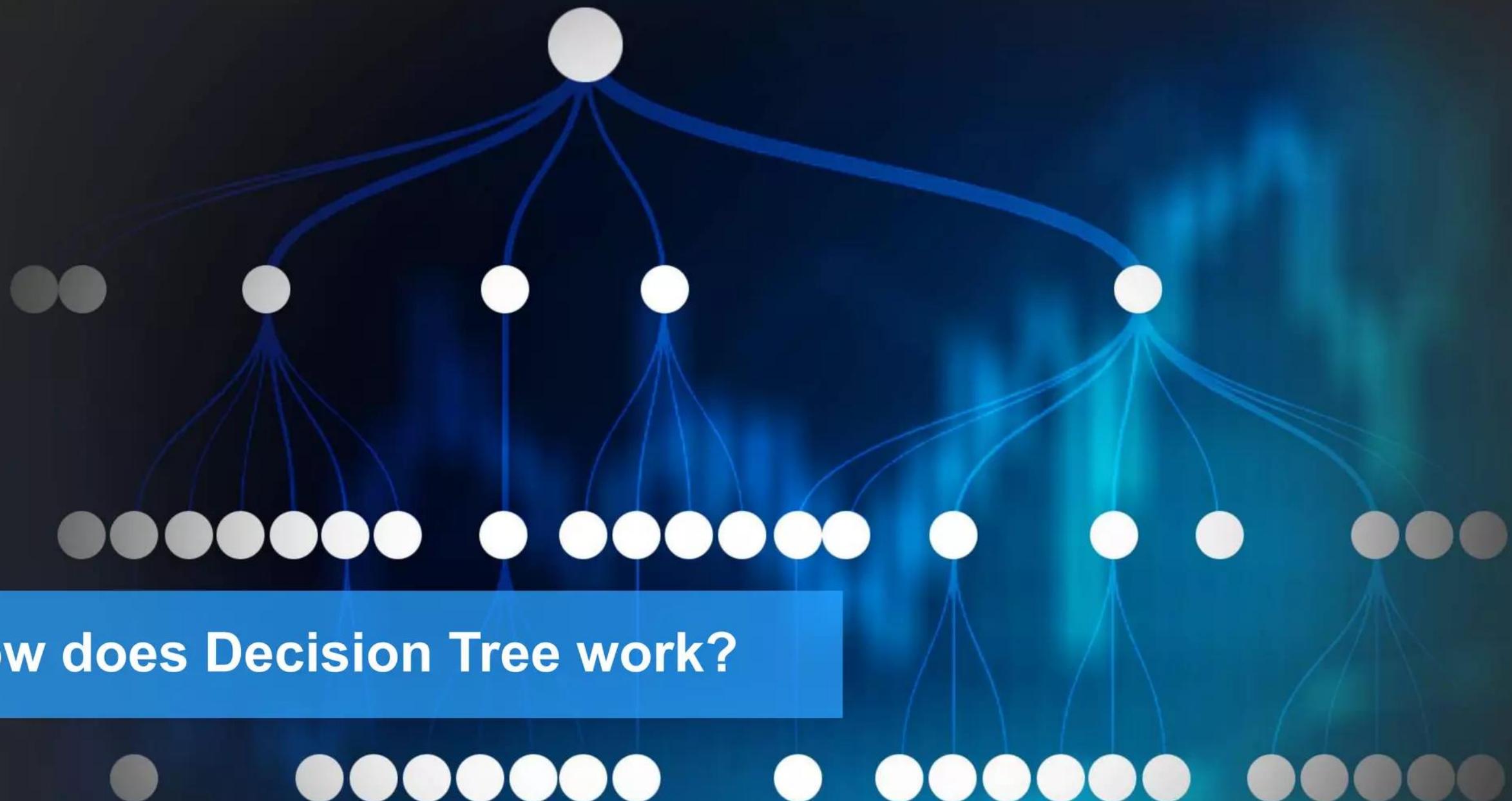
Root Node

The top most Decision node is known as the Root node

Example



Root Node



How does Decision Tree work?

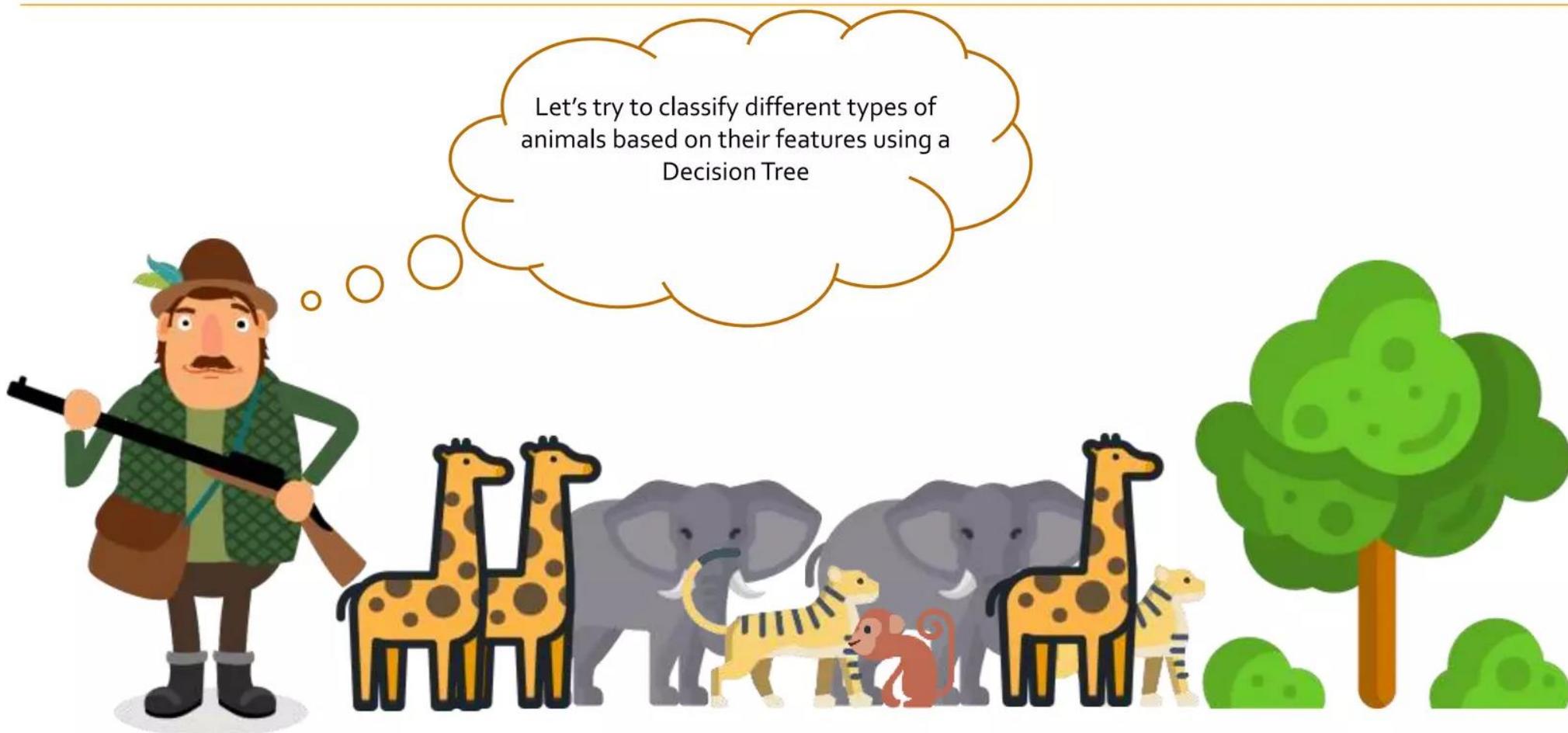
How does a Decision Tree work?



How does a Decision Tree work?



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?

The dataset is looking quite messy and the entropy is high in this case



Training Dataset

Color	Height	Label
grey	10	elephant
Yellow	10	giraffe
brown	3	Monkey
grey	10	elephant
Yellow	4	Tiger

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?

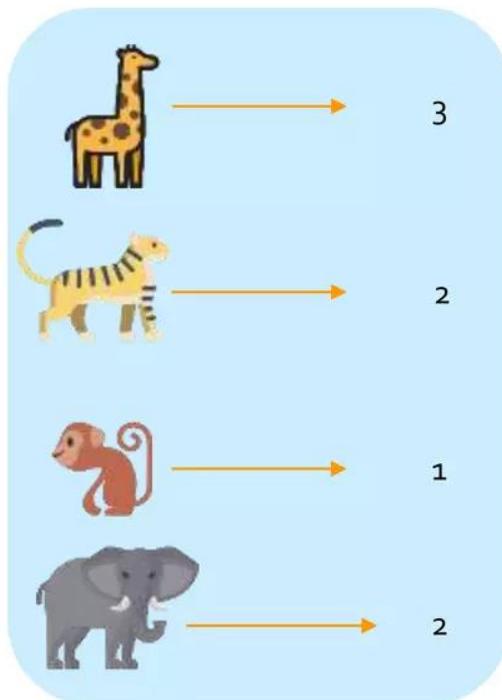
Formula for entropy

$$\sum_{i=1}^k P(value_i) \cdot \log_2(P(value_i))$$

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

$$\sum_{i=1}^k P(value_i) \cdot \log_2(P(value_i))$$



How does a Decision Tree work?

Let's use the formula

$$\sum_{i=1}^k P(value_i) \cdot \log_2(P(value_i))$$

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

$$\text{Entropy}=0.571$$



How does a Decision Tree work?

Let's use the formula

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

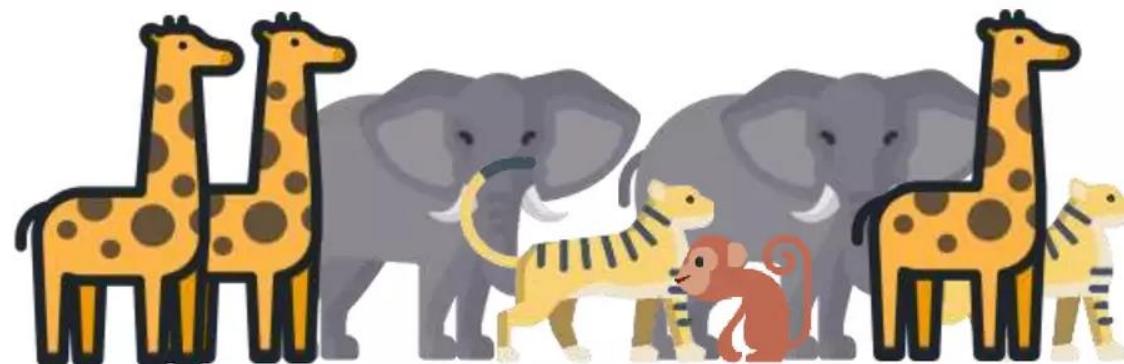
$$\text{Entropy} = \left(\frac{3}{8}\right) \log_2\left(\frac{3}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right) + \left(\frac{1}{8}\right) \log_2\left(\frac{1}{8}\right) + \left(\frac{2}{8}\right) \log_2\left(\frac{2}{8}\right)$$

$$\text{Entropy}=0.571$$



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

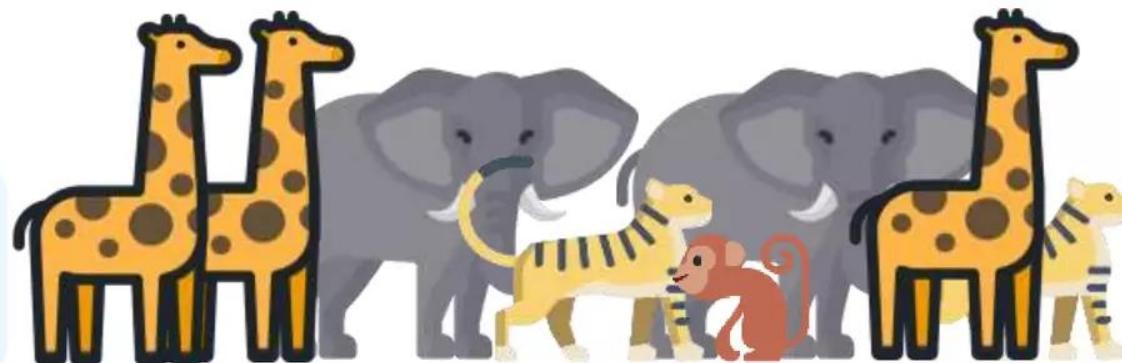
How does a Decision Tree work?

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?

Conditions

Color==Yellow?

Height \geq 10

Color== Brown?

Color==Grey

Diameter $<$ 10

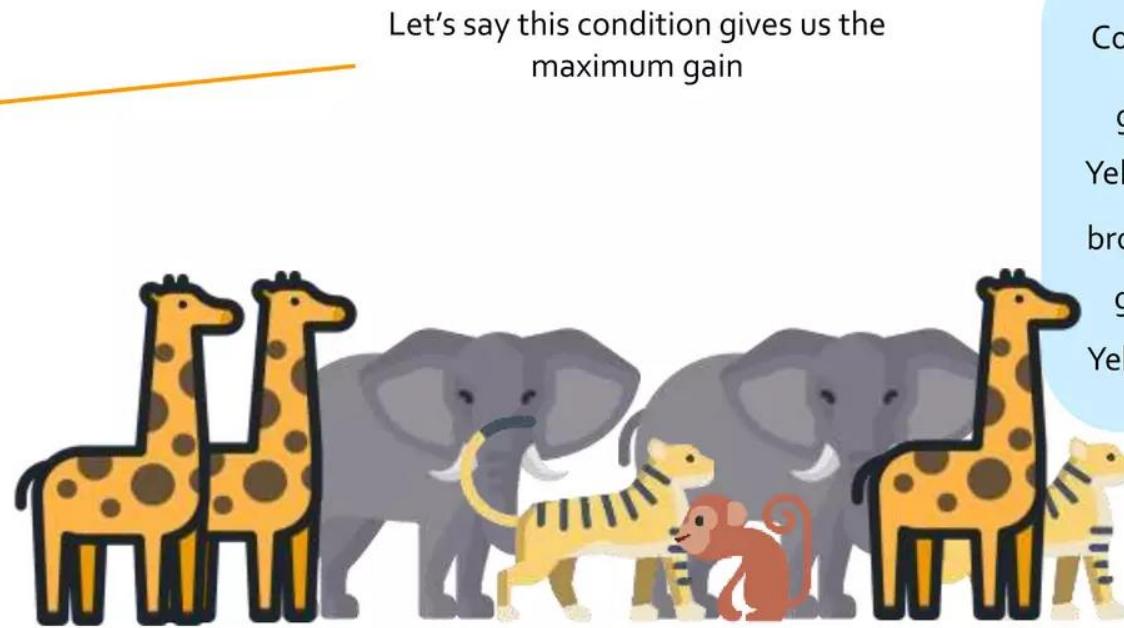
Training Dataset

Color	Height	Label
grey	10	elephant
Yellow	10	giraffe
brown	3	Monkey
grey	10	elephant
Yellow	4	Tiger



How does a Decision Tree work?

Conditions
Color==Yellow?
Height>=10
Color== Brown?
Color==Grey
Diameter<10

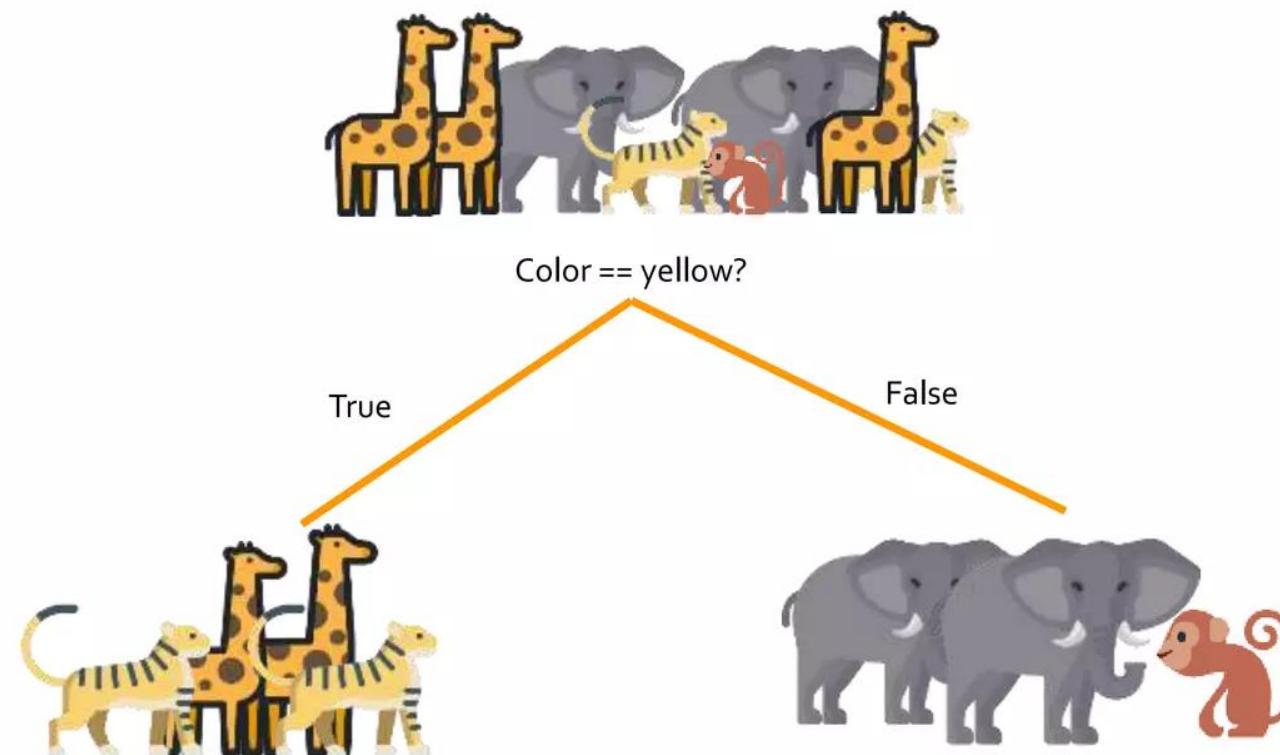


Let's say this condition gives us the maximum gain

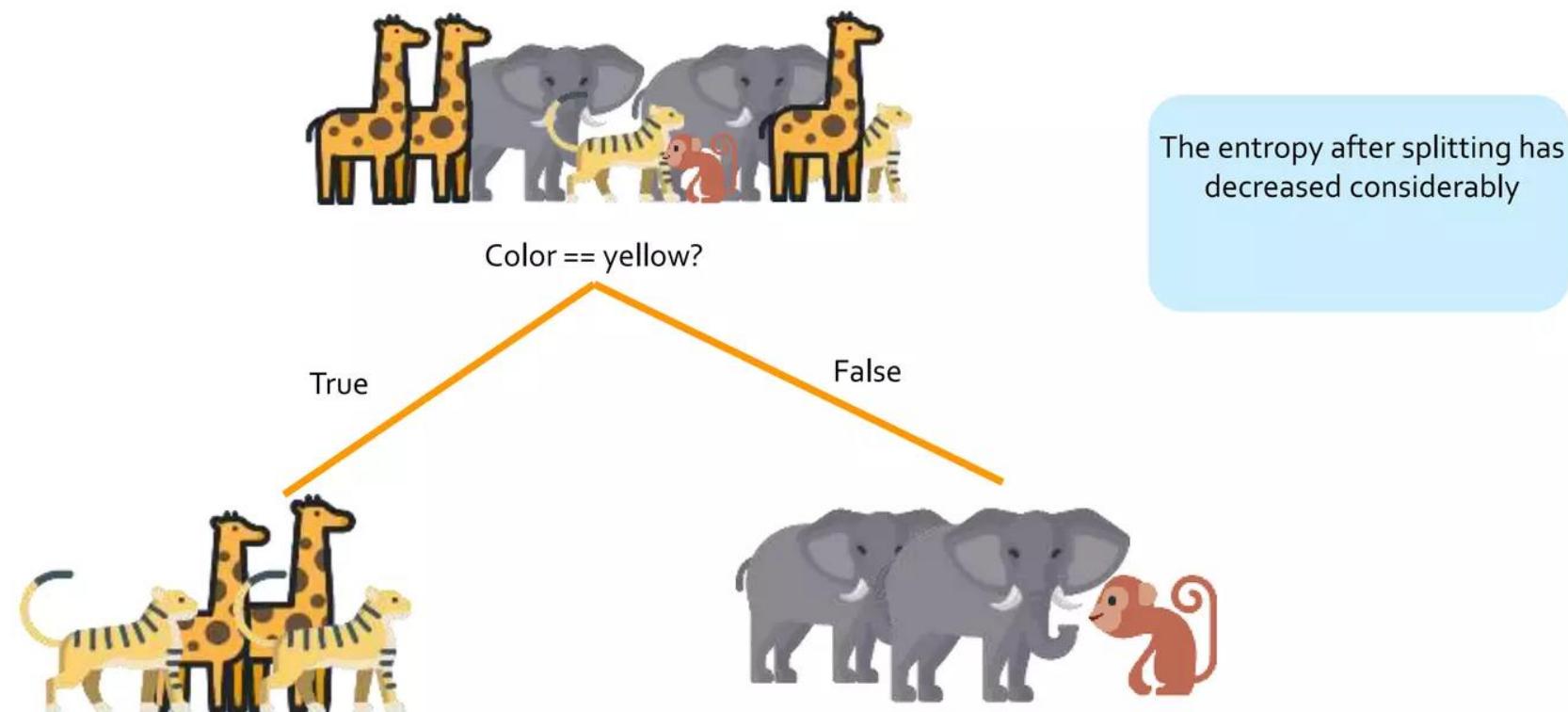
Training Dataset		
Color	Height	Label
grey	10	elephant
Yellow	10	giraffe
brown	3	Monkey
grey	10	elephant
Yellow	4	Tiger

How does a Decision Tree work?

We split the data



How does a Decision Tree work?



How does a Decision Tree work?

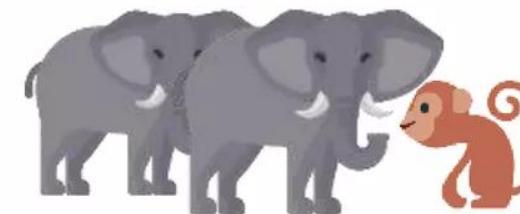
however we still need some splitting at both the branches to attain an entropy value equal to zero



Color == yellow?

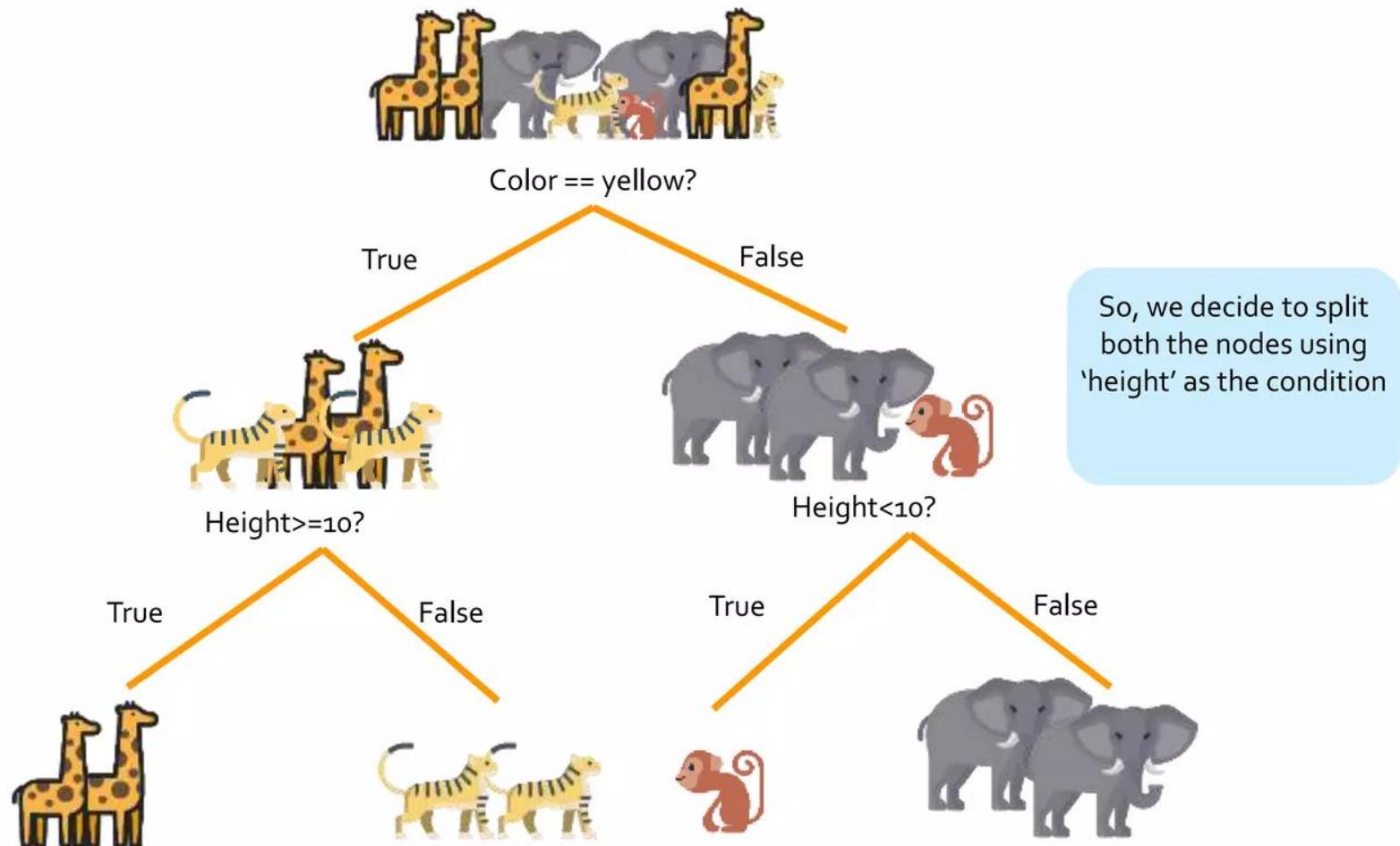
True

False

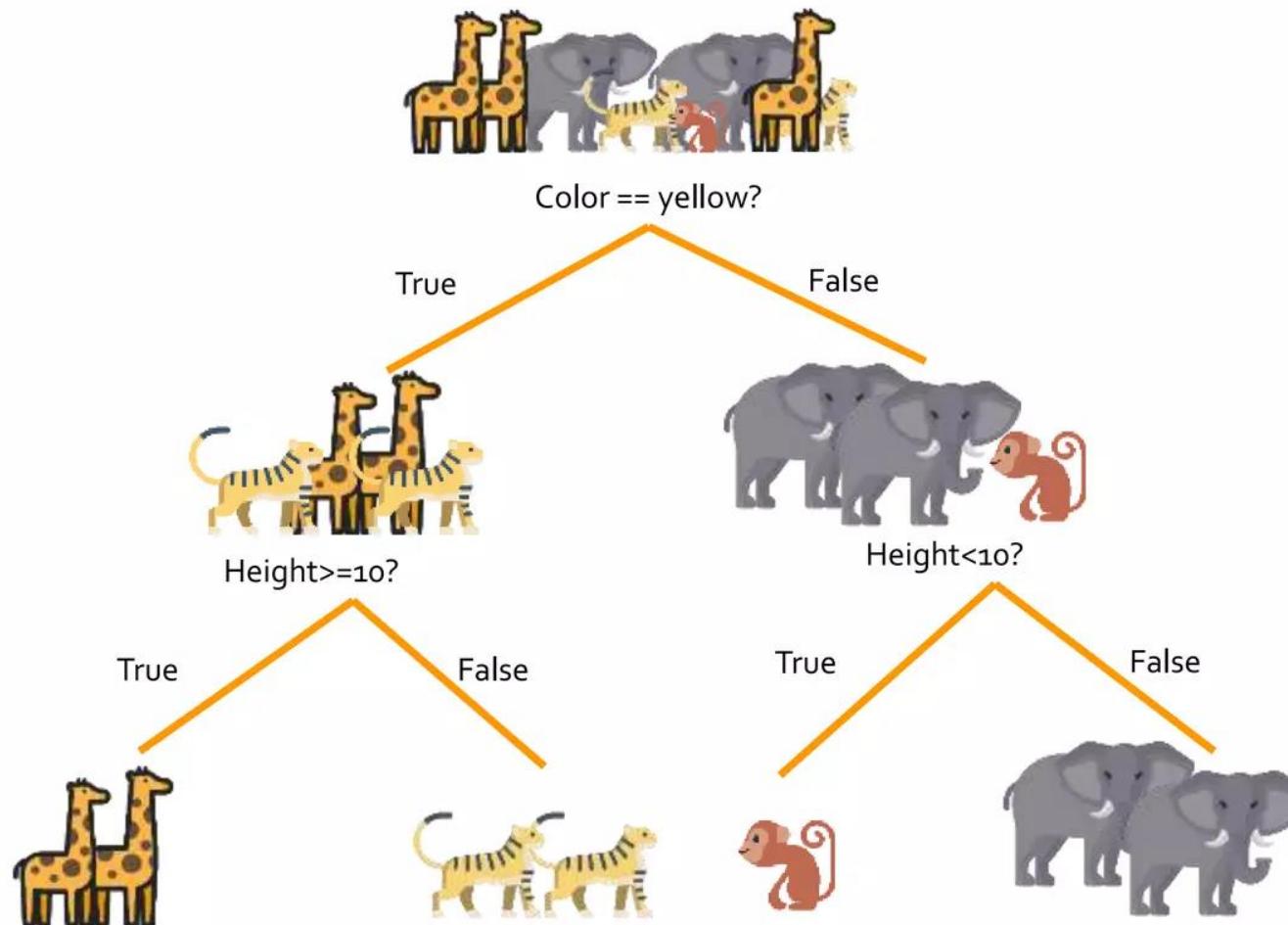


The entropy after splitting has decreased considerably

How does a Decision Tree work?

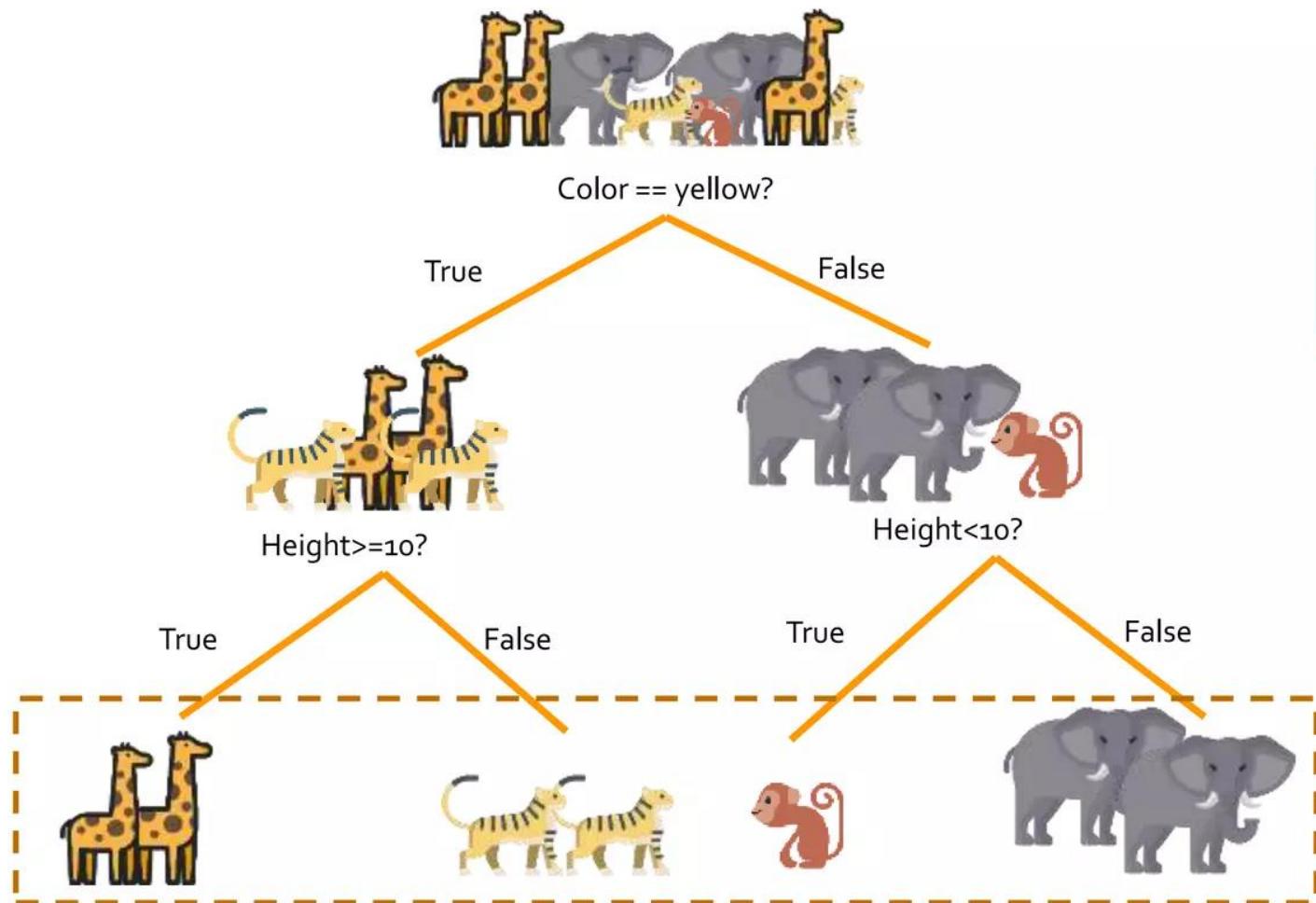


How does a Decision Tree work?



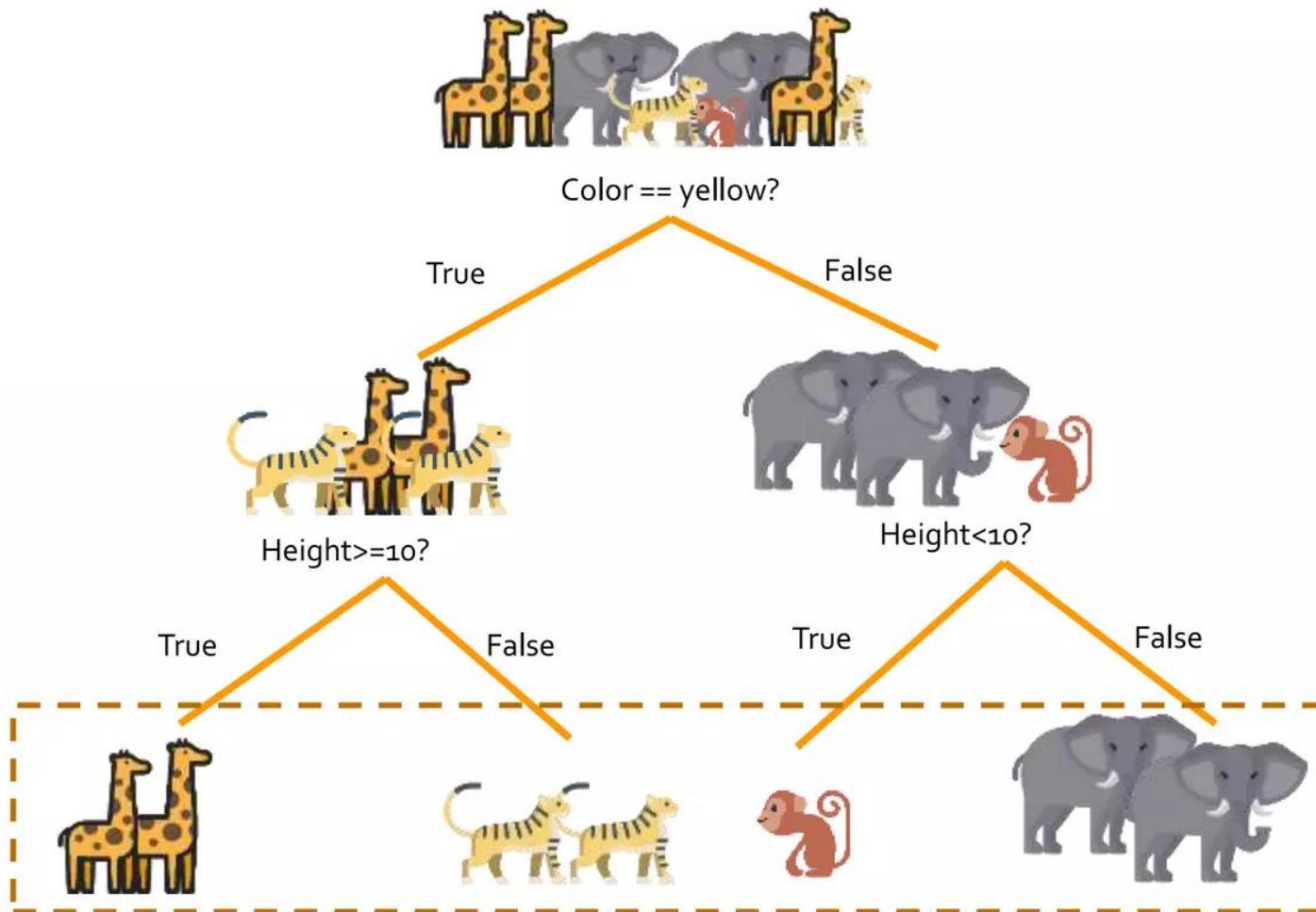
since every branch now contains single label type, we can say that the entropy in this case has reached the least value

How does a Decision Tree work?

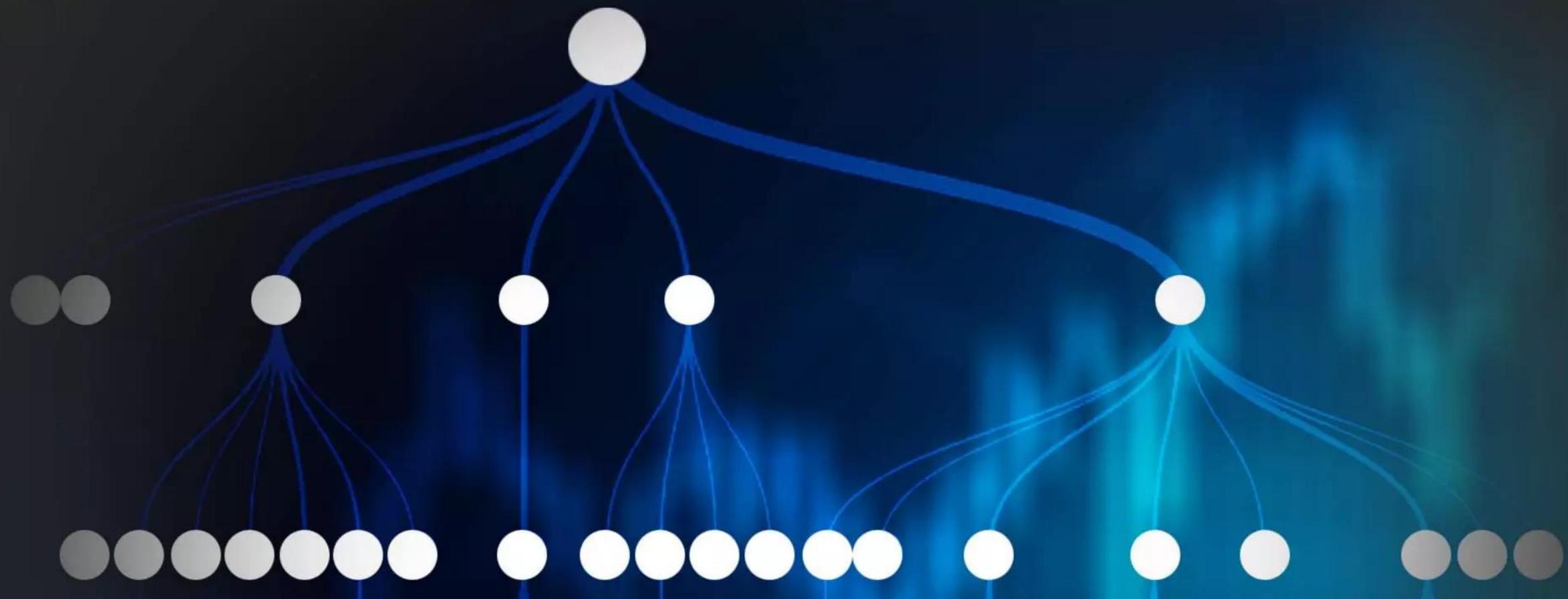


This Tree can now predict all the classes of animals present in the dataset with 100% accuracy

How does a Decision Tree work?



This Tree can now predict all the classes of animals present in the dataset with 100% accuracy



Use Case – Loan Repayment Prediction

Use Case – Loan Repayment prediction



Use Case – Problem Statement



Problem statement

To predict if a customer will repay loan amount or not using Decision Tree algorithm in python

Use Case – Implementation

```
#import the necessary packages
import numpy as np
import pandas as pd
from sklearn.cross_validation import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
from sklearn import tree

#Loading data file
balance_data =pd.read_csv('C:/Users/anirban.dey/Desktop/data_2.csv',
sep= ',', header= 0)
```



Use Case – Implementation

```
#import the necessary packages
print ("Dataset Length:: ", len(balance_data))
print ("Dataset Shape:: ", balance_data.shape)
```

```
Dataset Length::  
Dataset Shape::  
Out[166]: (None, (1000, 5))
```



simplilearn

Use Case – Implementation

```
print ("Dataset:: ")  
balance_data.head()
```

Dataset::

Out[167]:

	Result	Initial payment	Last payment	Credit Score	House Number
0	Yes	201	10018	250	3046
1	Yes	205	10016	395	3044
2	Yes	257	10129	109	3251
3	Yes	246	10064	324	3137
4	Yes	117	10115	496	3094



Use Case – Implementation

```
#Separating the Target variable
X = balance_data.values[:, 1:5]
Y = balance_data.values[:,0]

#Spliting Dataset into Test and Train
X_train, X_test, y_train, y_test = train_test_split( X, Y, test_size = 0.3,
random_state = 100)

#Function to perform training with Entropy
clf_entropy = DecisionTreeClassifier(criterion = "entropy", random_state = 100,
 max_depth=3, min_samples_leaf=5)
clf_entropy.fit(X_train, y_train)
```

```
Out[170]: DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=3,
 max_features=None, max_leaf_nodes=None,
 min_impurity_decrease=0.0, min_impurity_split=None,
 min_samples_leaf=5, min_samples_split=2,
 min_weight_fraction_leaf=0.0, presort=False, random_state=100,
 splitter='best')
```



Use Case – Implementation

```
#Function to make Predictions  
y_pred_en = clf_entropy.predict(X_test)  
y_pred_en
```



Use Case – Implementation

```
#Checking Accuracy  
print ("Accuracy is "), accuracy_score(y_test,y_pred)*100
```

Accuracy is

Out[172]: (None, 94.66666666666671)



simplilearn

Use Case



So, we have created a model that uses decision tree algorithm to predict whether a customer will repay the loan or not

Use Case



The Accuracy of the model is 94.6%

Use Case



The bank can use this model to decide whether it should approve loan request from a particular customer or not

Key takeaways

The diagram illustrates the progression of a system. On the left, a box labeled 'Traditional system' contains a person icon. An arrow points from this box to a middle box labeled 'Ability to learn and improve over time'. This middle box contains a person icon with a lightbulb above their head, symbolizing thought or learning. Another arrow points from the middle box to the right, leading to a box labeled 'Machine Learning'. This final box contains a person icon surrounded by various icons representing data and connectivity, such as a smartphone, a laptop, a bar chart, and a network graph.

The diagram consists of three circular icons, each representing a type of machine learning:

- Supervised Learning:** An icon showing a person standing next to a large document or board, with two smaller figures at the bottom.
- Unsupervised Learning:** An icon showing a laptop displaying a scatter plot of data points, with a magnifying glass over it.
- Reinforcement Learning:** An icon showing a silhouette of a person's head with various symbols (diamonds, question marks, plus signs) floating around it.

Below each icon, the text labels "Supervised Learning", "Unsupervised Learning", and "Reinforcement Learning" are written in a light blue font.

Classification
Predicts categorical labels like "Yes" or "No", "True" or "False", etc.

Regression
Predicts continuous values like house price, weight, etc.

Clustering
Groups unlabeled data points that have similar responses to form the clusters like "Sports", "Inconsequential".

How does a Decision Tree work?

THE TREE CAN NOW PREDICT ALL THE CLASSES OF ANIMALS PRESENT IN THE DISTRICT WITH 100% ACCURACY

The image shows a man in a dark suit and tie standing on a sidewalk next to a road with a yellow street lamp. He is holding a light-colored smartphone in his right hand. A thought bubble originates from the phone, containing the following text:

I NEED TO FIND OUT IF MY
CUSTOMERS ARE GOING TO
RETURN THE LOAN THEY
TOOK FROM ANY BANK OR NOT