DECISION TREE TUTORIAL

simplilearn

Decision Tree Tutorial

simplilearn

Decision Tree Tutorial

simplilearn

Decision Tree Tutorial

simplilearn

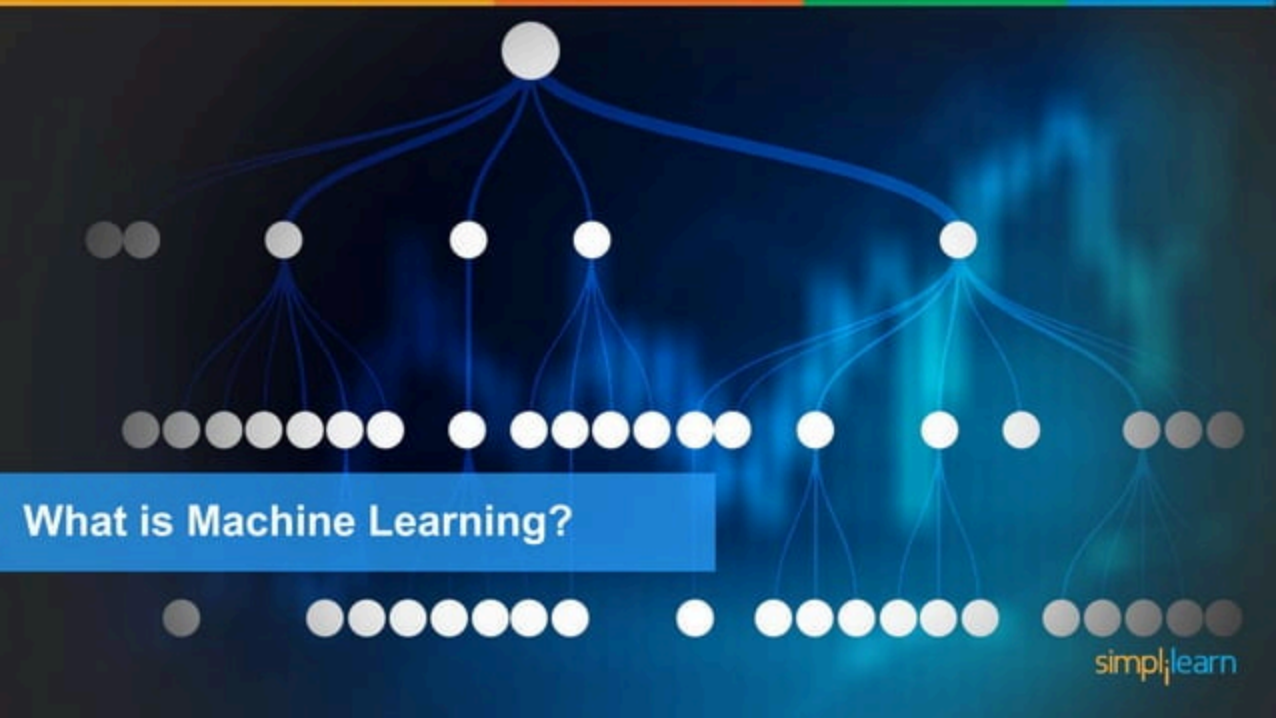
Decision Tree Tutorial

Decision Tree Tutorial

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

simplilearn


What is Machine Learning?
simplilearn

What is Machine Learning?

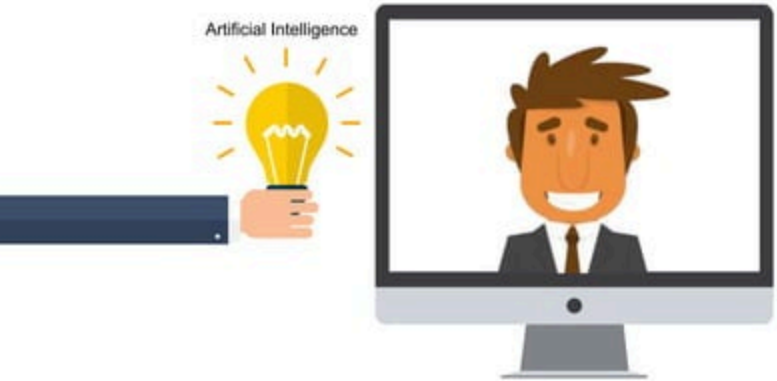
What is Machine Learning?

simplilearn

What is Machine Learning?

simplilearn

What is Machine Learning?

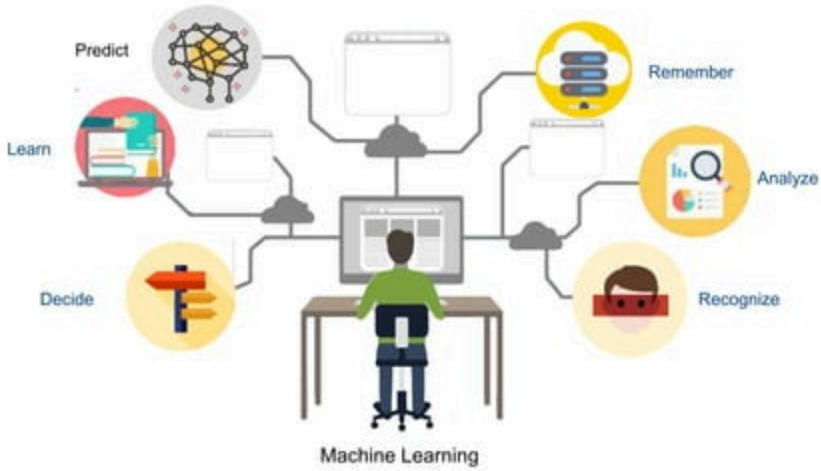


simplilearn

What is Machine Learning?

simplilearn

What is Machine Learning?



simplilearn

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

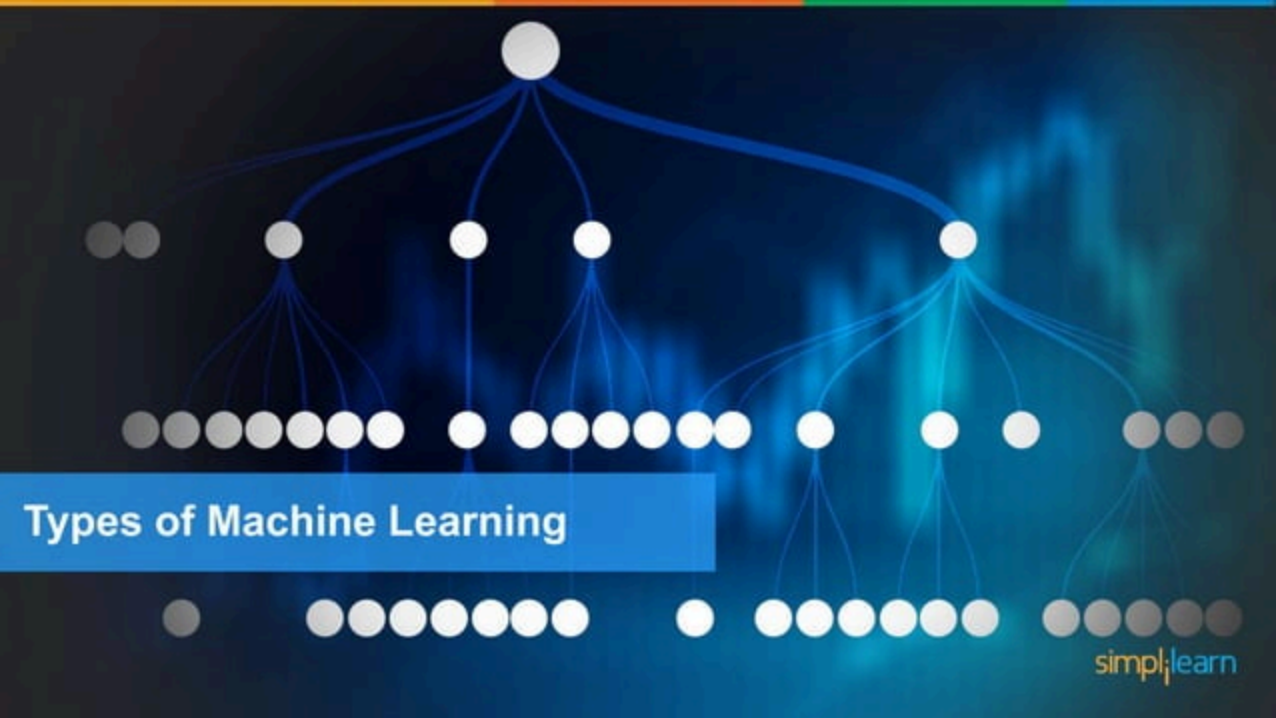
What is Machine Learning?

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
simplilearn

Types of Machine Learning

Supervised Learning

Types of Machine Learning

Supervised Learning

Unsupervised Learning

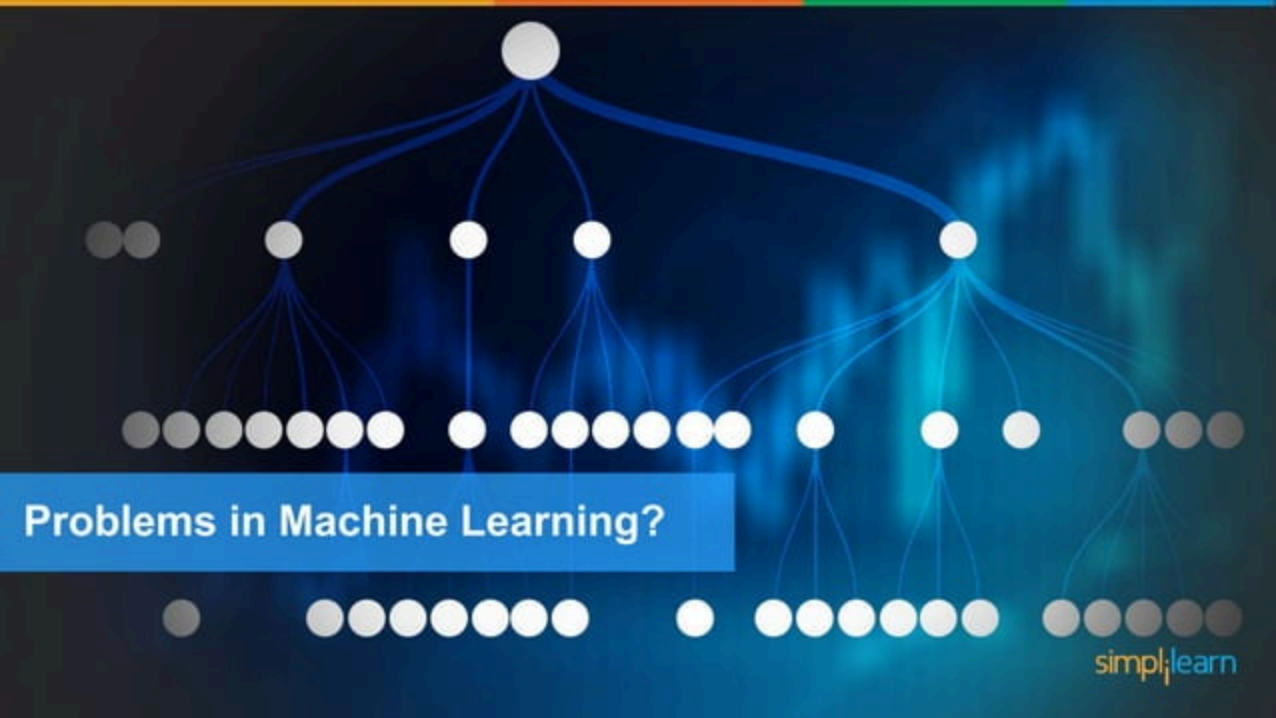
Types of Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

simplilearn

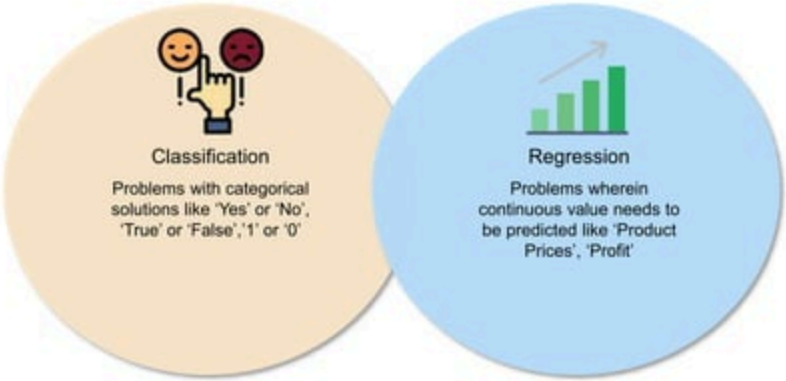

Problems in Machine Learning?
simplilearn

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

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

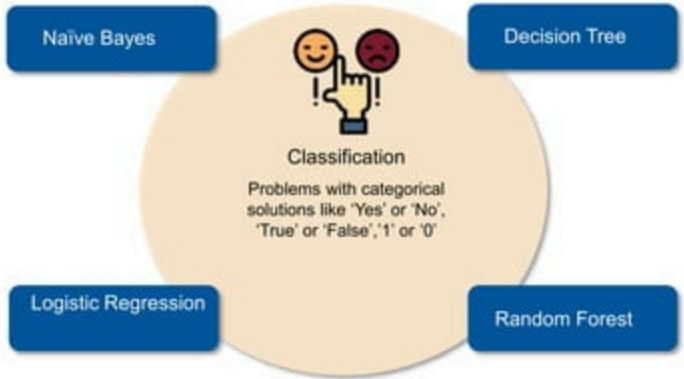
Problems wherein the data needs to be organized to find specific patterns like in the case of 'Product Recommendation'

Problems in Machine Learning

Classification

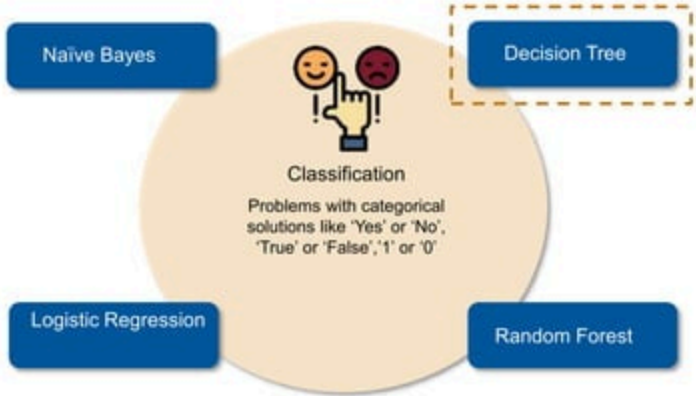
Problems with categorical solutions like 'Yes' or 'No', 'True' or 'False','1' or '0'

Problems in Machine Learning

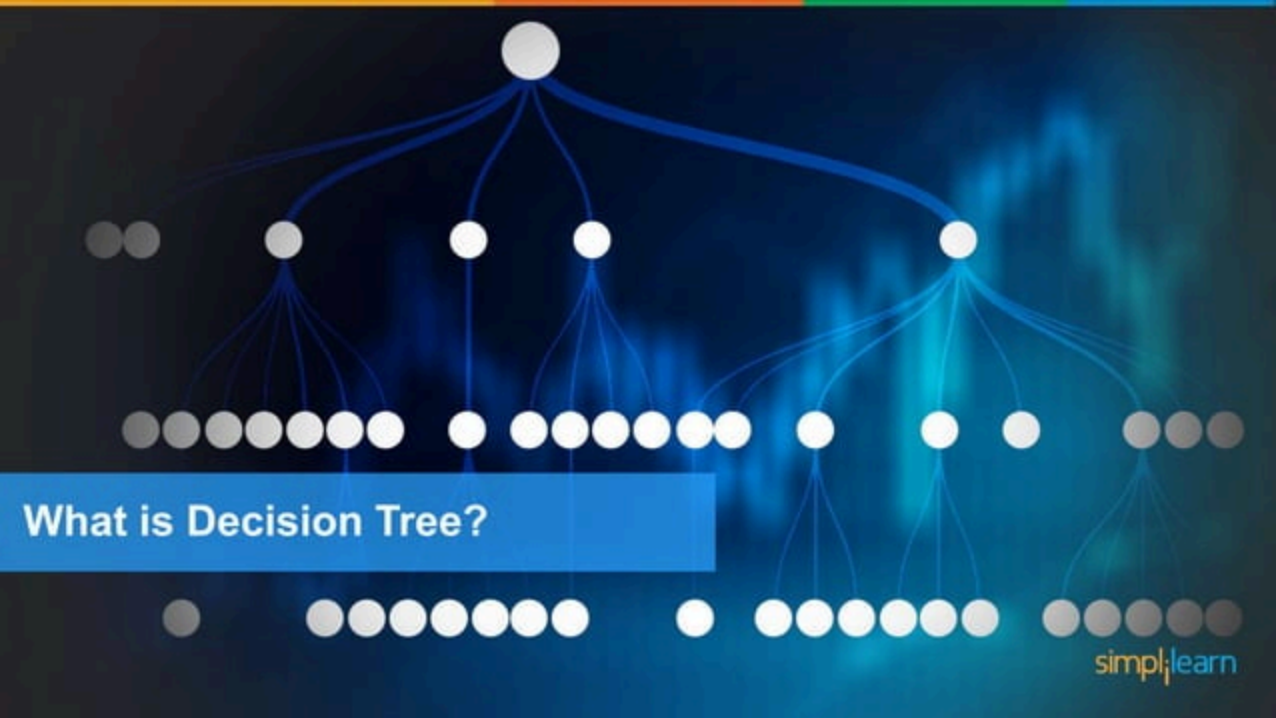


simplilearn

Problems in Machine Learning



simplilearn


What is Decision Tree?
simplilearn

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

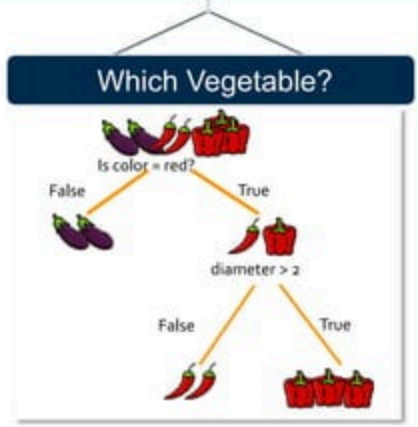
What is Decision Tree?

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

simplilearn

What is Decision Tree?

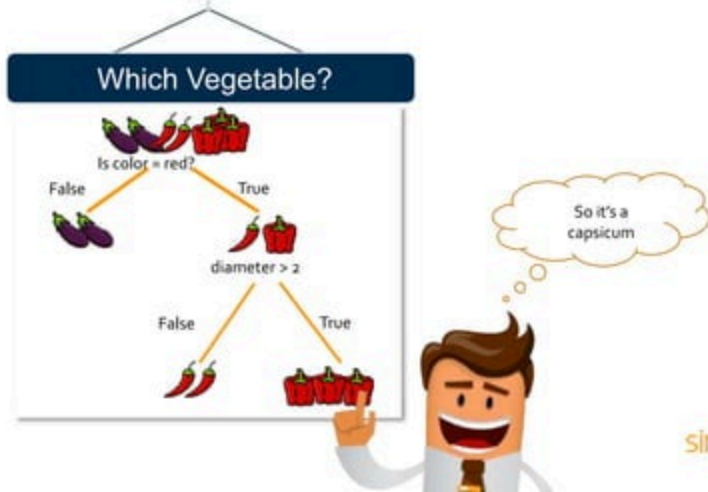
Decision Tree is a tree shaped diagram used to determine a course of action. Each branch of the tree represents a possible decision, occurrence or reaction



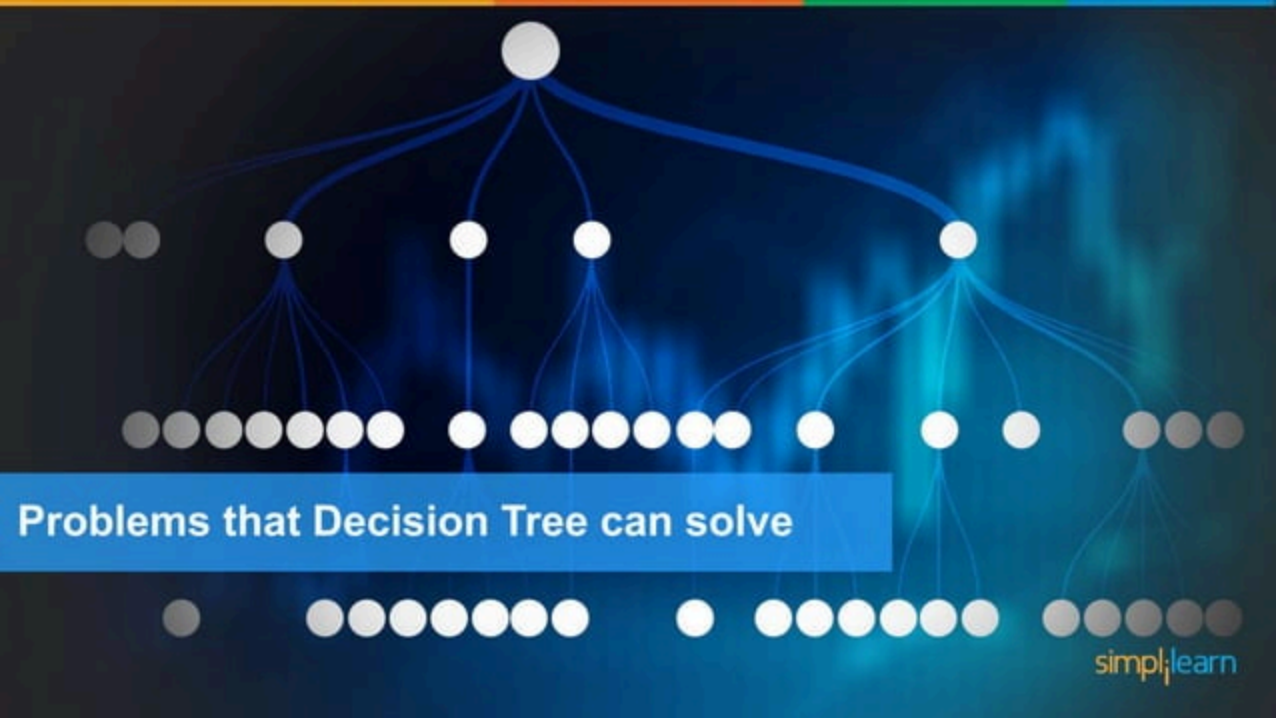
simplilearn

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



simplilearn


Problems that Decision Tree can solve
simplilearn

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

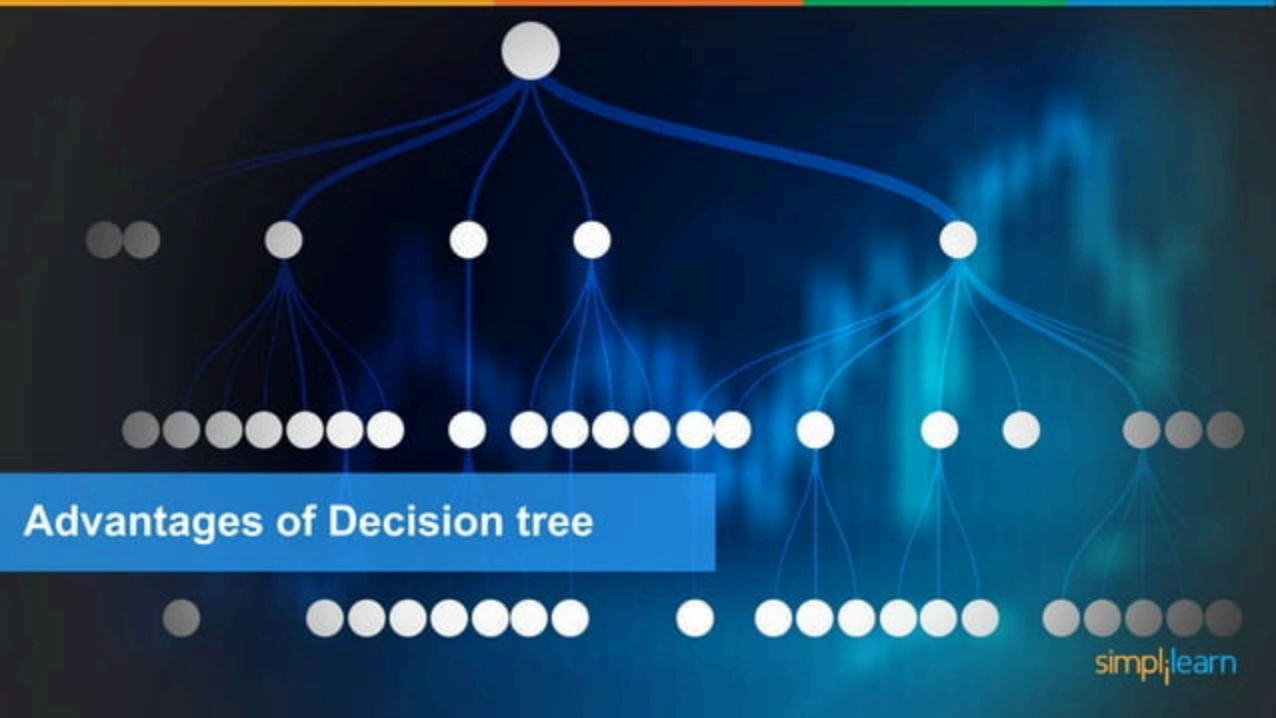
Problems that Decision Tree can solve

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

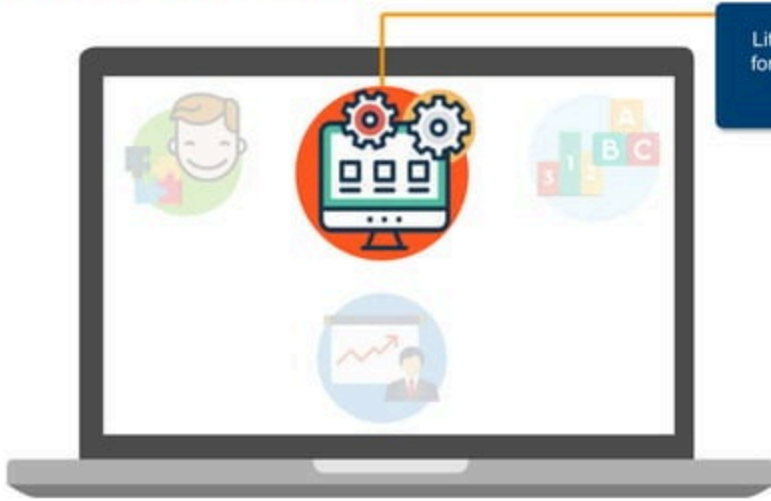
Regression tree is used when the target variable is numerical or continuous in nature. We fit a regression model to the target variable using each of the independent variables. Each split is made based on the sum of squared error.


Advantages of Decision tree
simplilearn

Advantages of Decision Tree

Advantages of Decision Tree

Advantages of Decision Tree

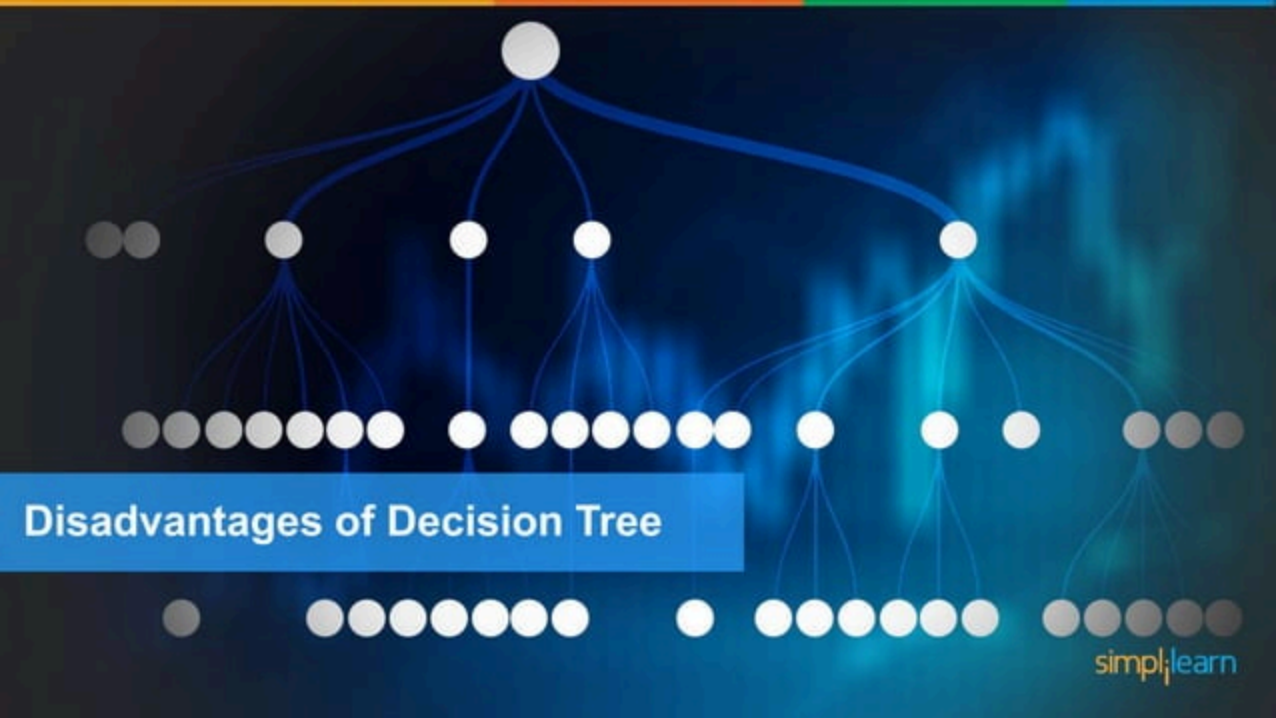


Advantages of Decision Tree

simplilearn

Advantages of Decision Tree

simplilearn

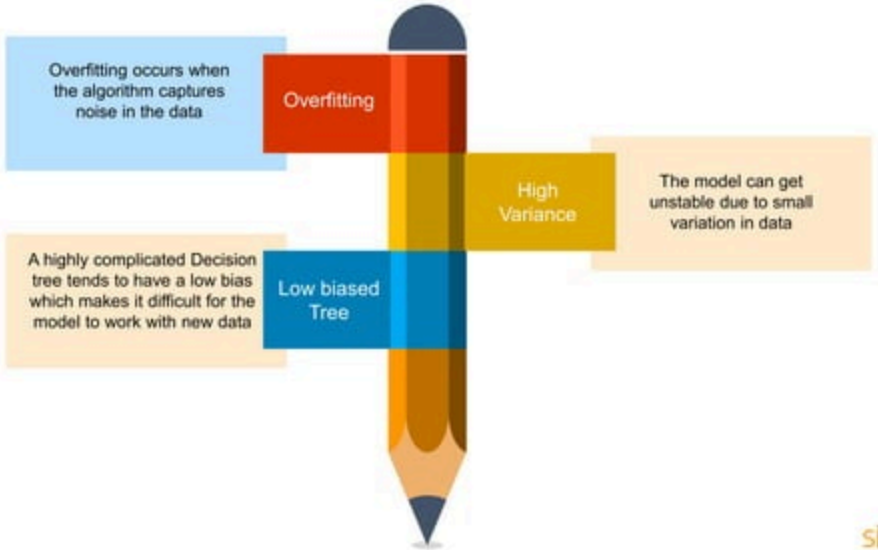

Disadvantages of Decision Tree
simplilearn

Disadvantages of Decision Tree

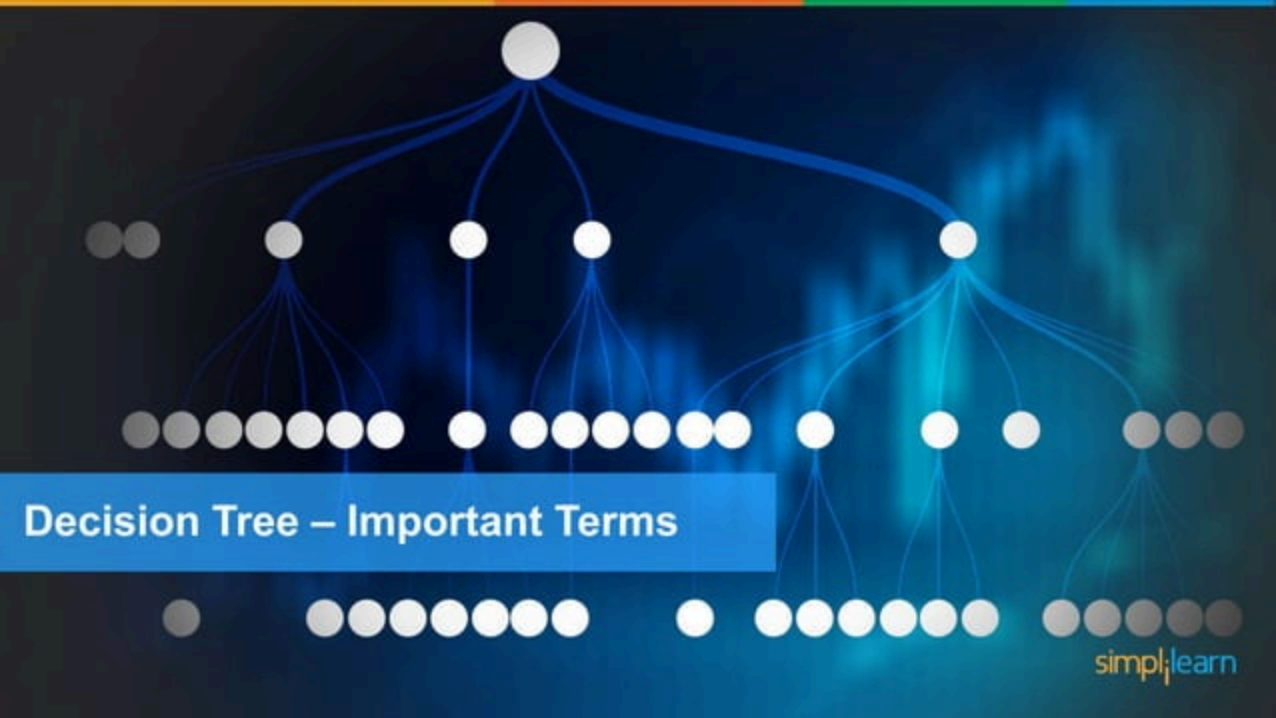
Disadvantages of Decision Tree

simplilearn

Disadvantages of Decision Tree



simplilearn


Decision Tree – Important Terms
simplilearn

Decision Tree – Important Terms

simplilearn

Decision Tree – Important Terms

Entropy

Entropy is the measure of randomness or unpredictability in the dataset

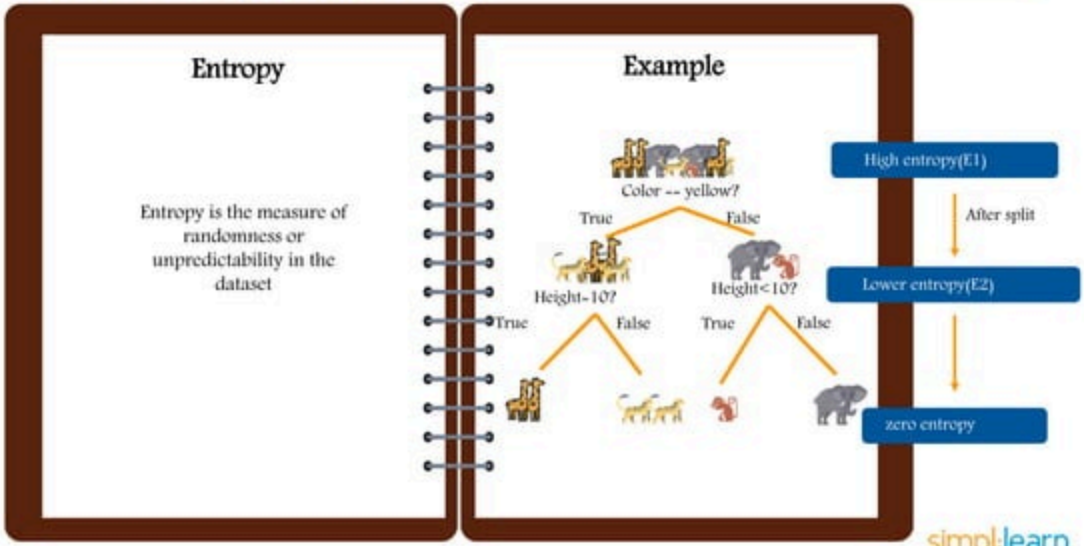
Example

High entropy

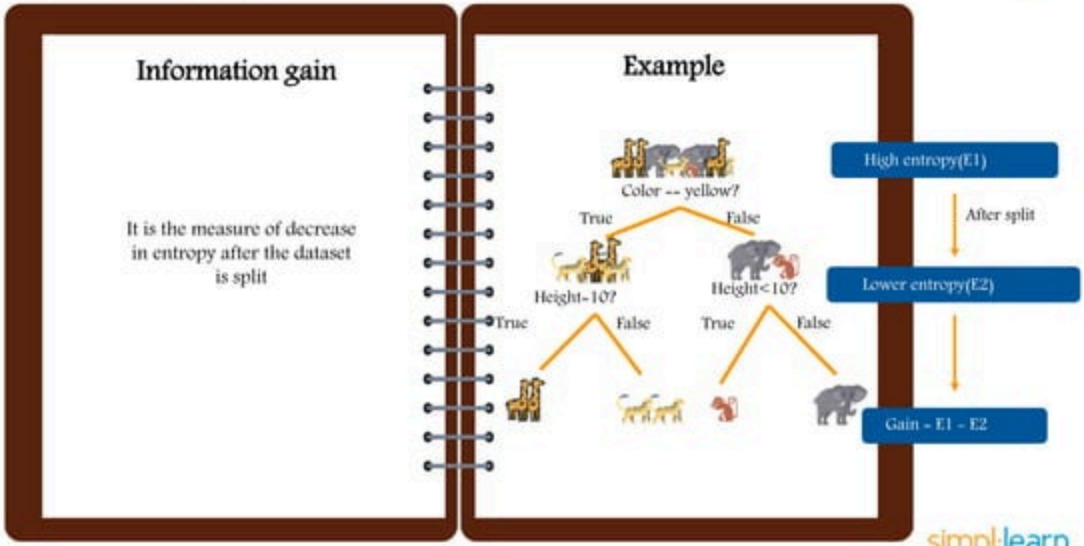
This Dataset has a very high entropy

simplilearn

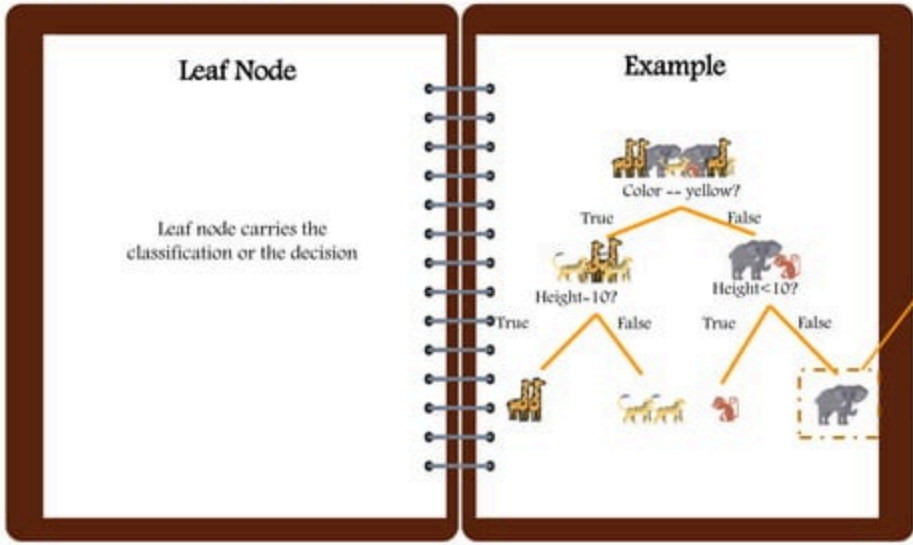
Decision Tree – Important Terms



Decision Tree – Important Terms

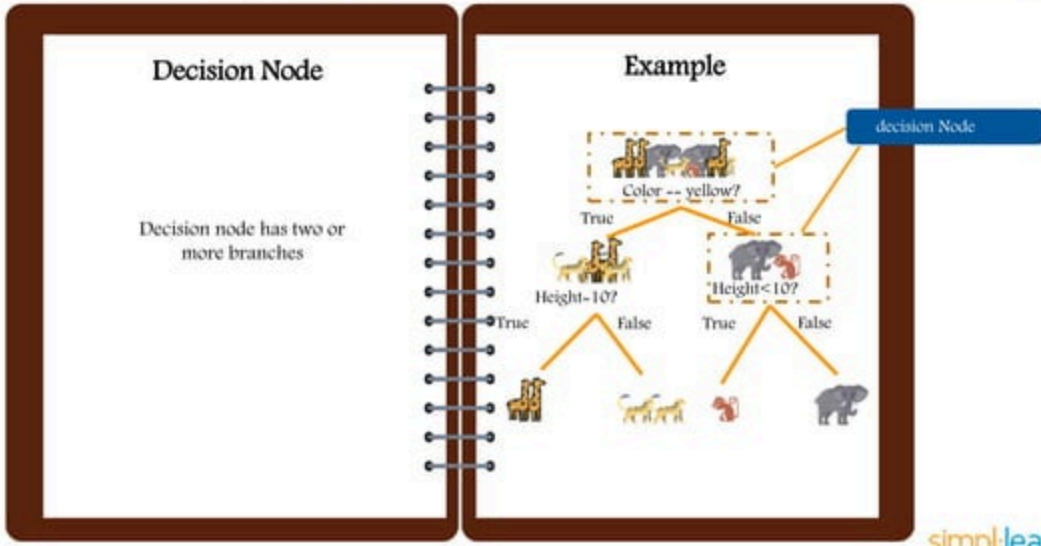


Decision Tree – Important Terms

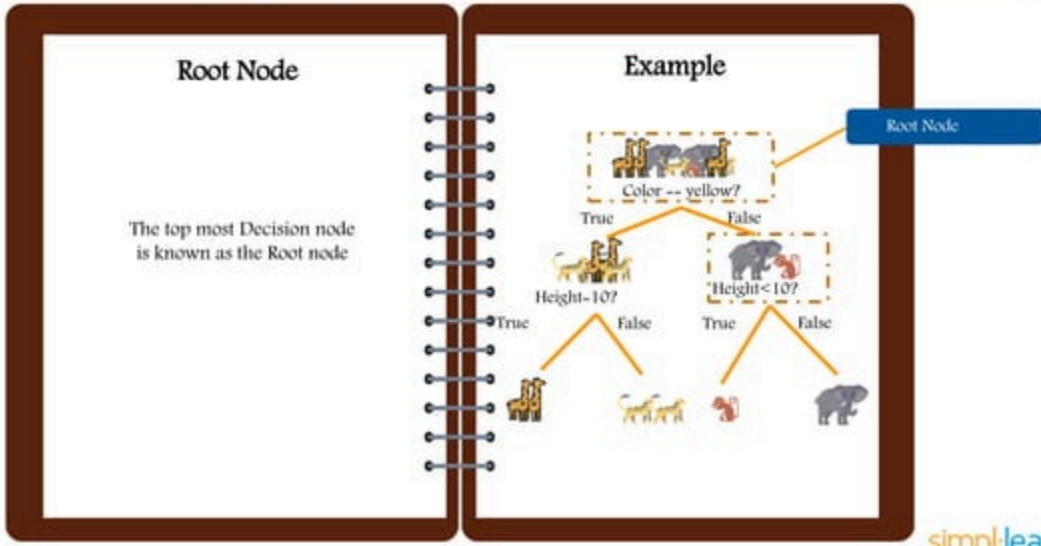


simplilearn

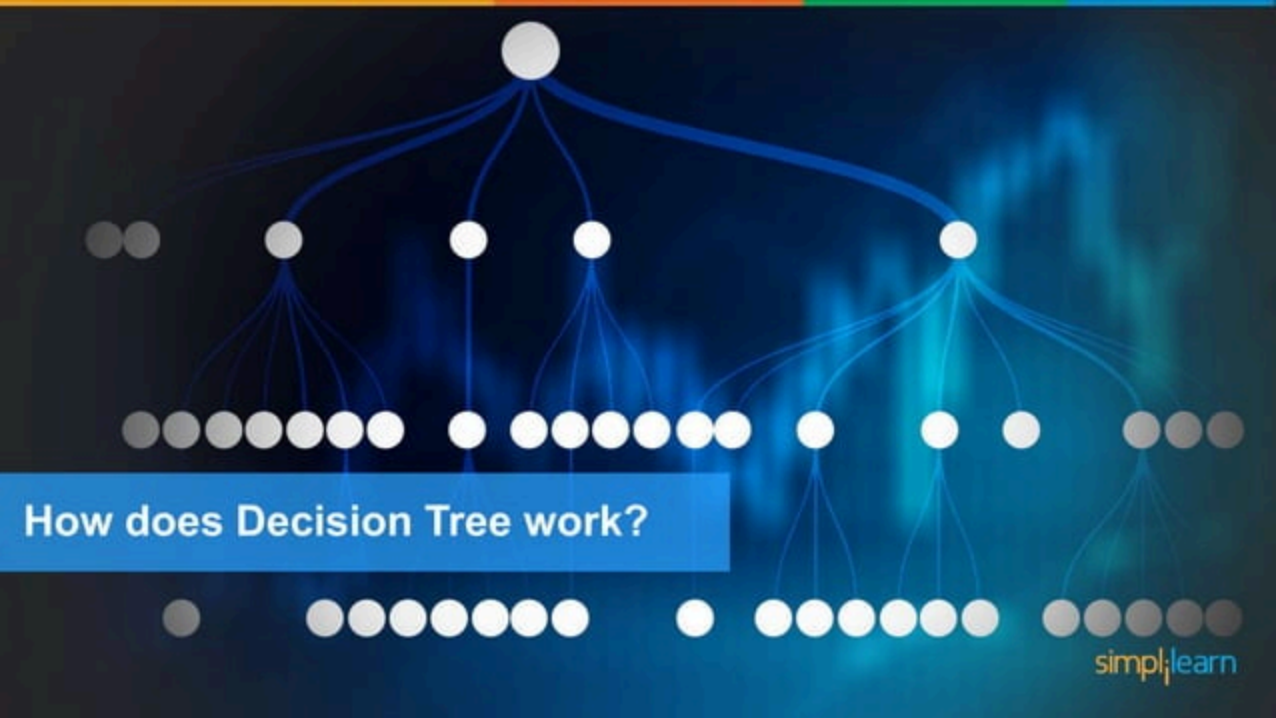
Decision Tree – Important Terms



Decision Tree – Important Terms



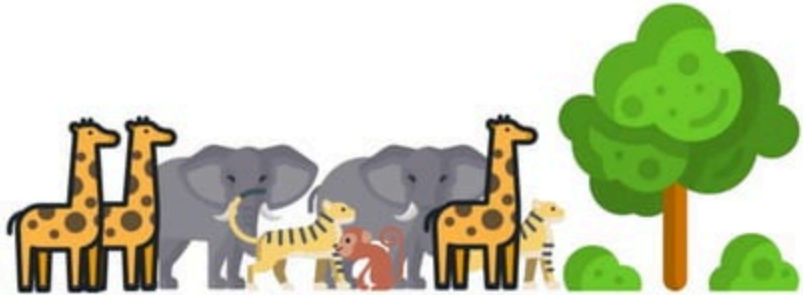
simplilearn


How does Decision Tree work?
simplilearn

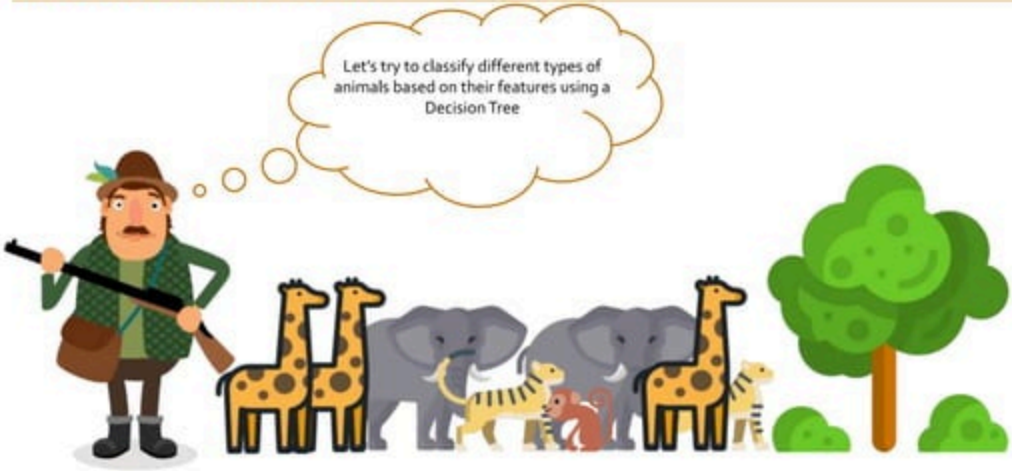
How does a Decision Tree work?

simplilearn

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

simplilearn

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

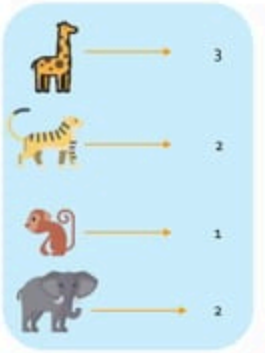
How does a Decision Tree work?

Formula for entropy

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

Let's try to calculate the entropy for the current dataset

How does a Decision Tree work?



simplilearn

How does a Decision Tree work?

Let's use the formula

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

How does a Decision Tree work?

Let's use the formula

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

Entropy = $(\frac{1}{8})\ log_2(\frac{1}{8}) + (\frac{2}{8})\ log_2(\frac{2}{8}) + (\frac{1}{8})\ log_2(\frac{1}{8}) + (\frac{2}{8})\ log_2(\frac{2}{8})$

Entropy=0.571

How does a Decision Tree work?

Let's use the formula

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

$$Entropy = \left(\frac{2}{6}\right)\log_2\left(\frac{2}{6}\right) + \left(\frac{3}{6}\right)\log_2\left(\frac{3}{6}\right) + \left(\frac{1}{6}\right)\log_2\left(\frac{1}{6}\right) + \left(\frac{5}{6}\right)\log_2\left(\frac{5}{6}\right)$$

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?

We will do that by splitting the data using each condition and checking the gain that we get out them.

Now we will try to choose a condition that gives us the highest gain

How does a Decision Tree work?

We will do that by splitting the data using each condition and checking the gain that we get out them.

The condition that gives us the highest gain will be used to make the first split

How does a Decision Tree work?

Conditions

- Color== Yellow?
- Height>=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

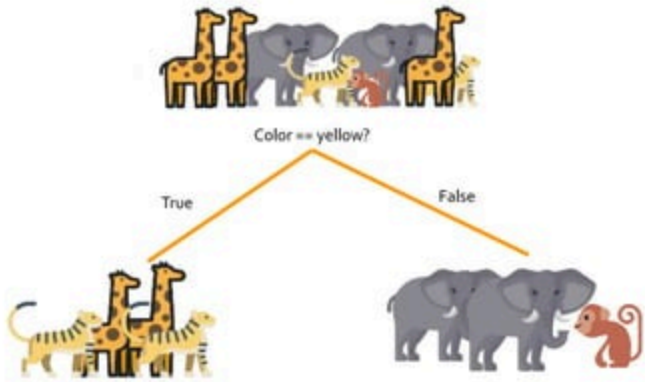
Let's say this condition gives us the maximum gain

Training Dataset

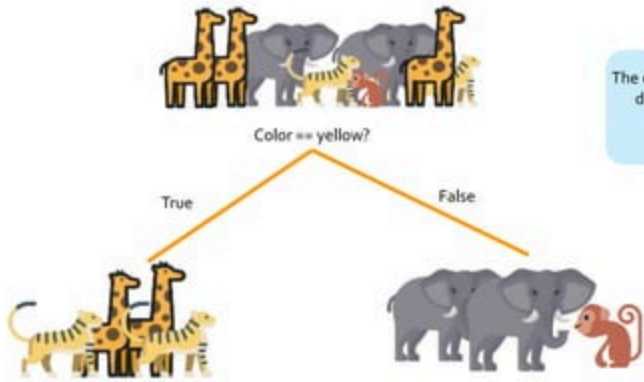
Color	Height	Label
grey	10	elephant
Yellow	10	giraffe
brown	3	Monkey
grey	10	elephant
Yellow	4	Tiger

How does a Decision Tree work?

We split the data



How does a Decision Tree work?



The entropy after splitting has decreased considerably

simplilearn

How does a Decision Tree work?

Color == yellow?

True

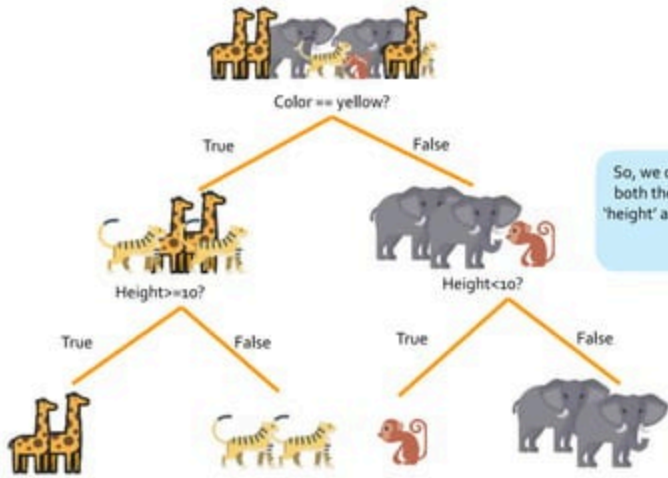
False

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

The entropy after splitting has decreased considerably

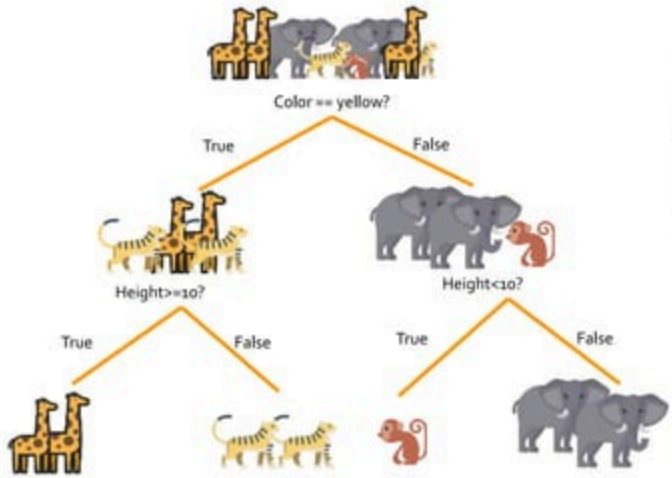
simplilearn

How does a Decision Tree work?



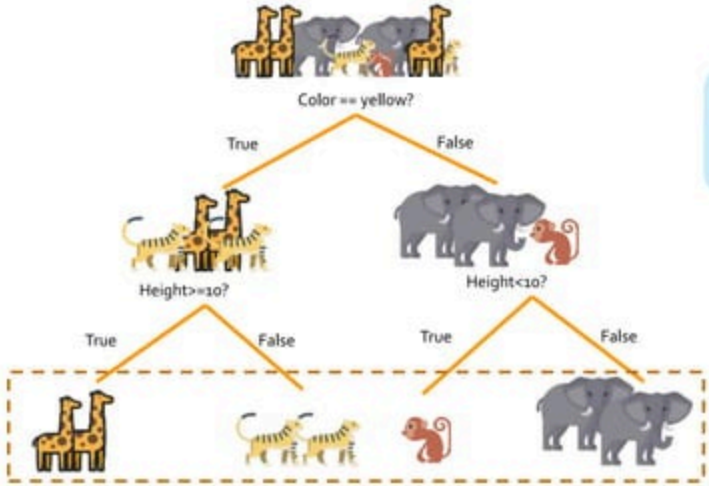
simplilearn

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

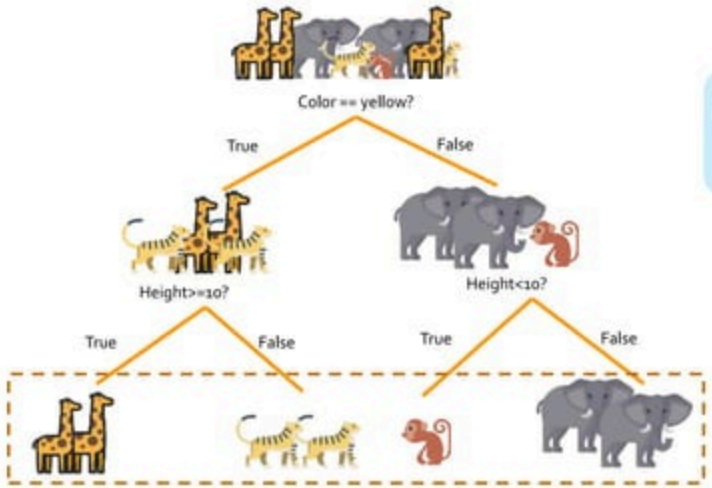
How does a Decision Tree work?



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

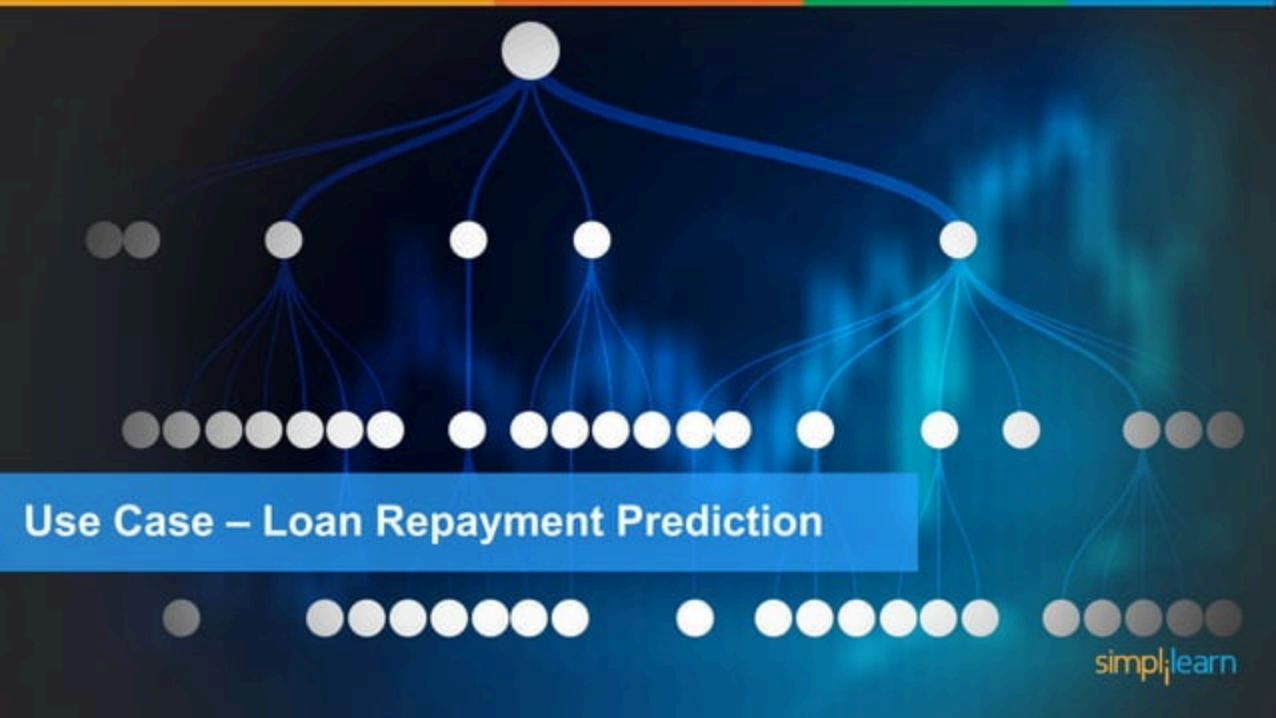
simplilearn

How does a Decision Tree work?



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

simplilearn


Use Case – Loan Repayment Prediction
simplilearn

Use Case – Loan Repayment prediction

simplilearn

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)
```

simplilearn

Use Case – Implementation

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

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

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

simplilearn

Use Case – Implementation

```
#Seperating 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')
```

simplilearn

Use Case – Implementation

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

```
Out[171]: array(['Yes', 'Yes', 'No', 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'No', 'No',
                 'No', 'No', 'Yes', 'No', 'No', 'Yes', 'Yes', 'No', 'Yes', 'No',
                 'No', 'Yes', 'No', 'Yes', 'Yes', 'No', 'No', 'Yes', 'No', 'No',
                 'No', 'Yes', 'Yes', 'Yes', 'Yes', 'No', 'No', 'No', 'Yes', 'No',
                 'Yes', 'Yes', 'Yes', 'No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'No',
                 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes',
                 'Yes', 'No', 'Yes', 'Yes', 'No', 'Yes', 'Yes', 'Yes', 'No', 'No',
                 'No', 'No', 'No', 'Yes', 'No', 'Yes', 'Yes', 'No', 'Yes', 'No',
```

simplilearn

Use Case – Implementation

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

```
          Accuracy is
Out[172]: (None, 94.666666666666671)
```

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

simplilearn


THANK YOU
For more information, visit
www.simplilearn.com
simplilearn