Machine learning Session

Ayush Jain

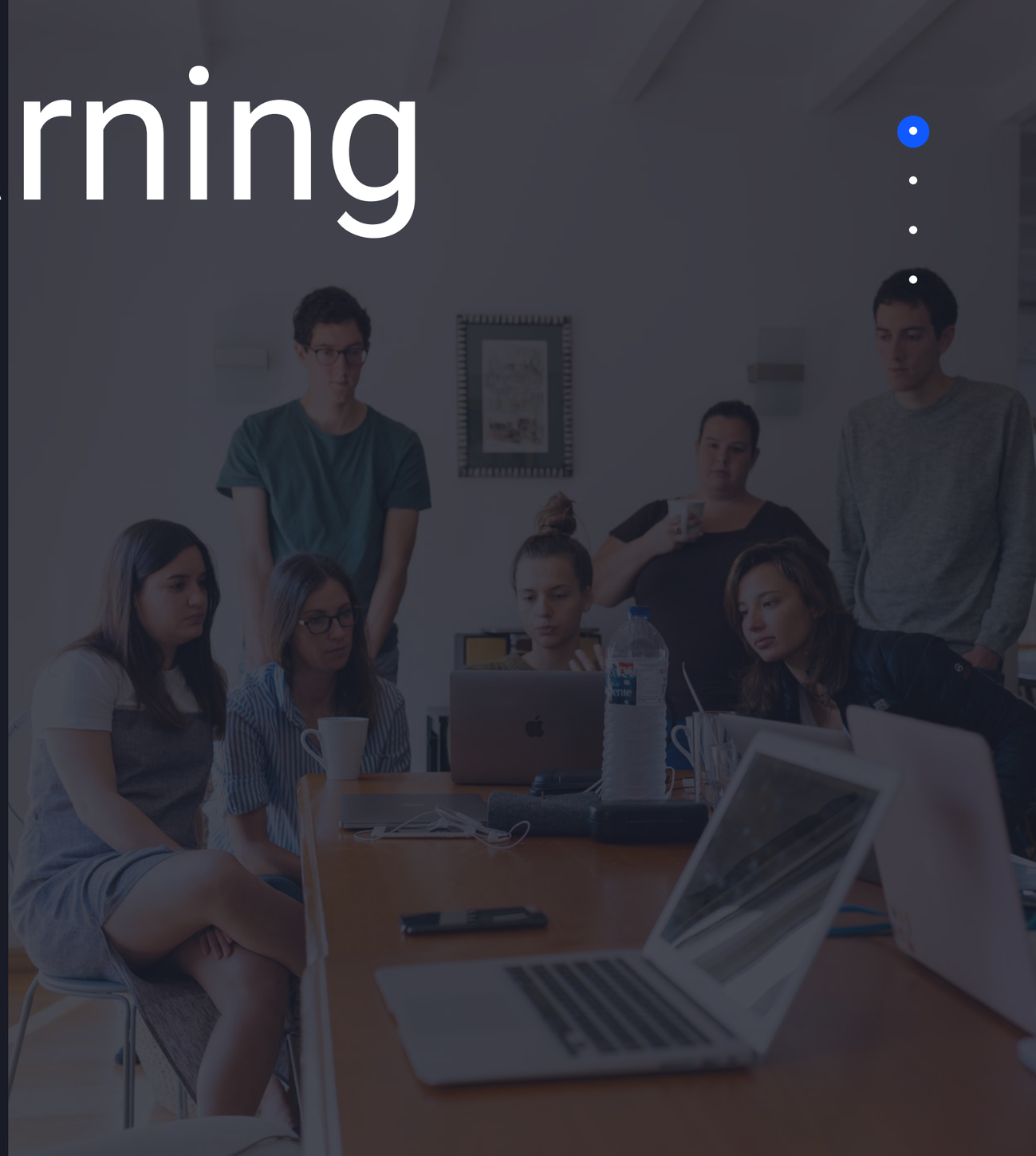


Table of Contents

SUPERVISED MACHINE LEARNING

Supervised learning is one of the most basic types of machine learning. In this type, the machine learning algorithm is trained on labelled data. Even though the data needs to be labelled accurately for this method to work, supervised learning is extremely powerful when used in the right circumstances.

NEXT

In supervised learning, the ML algorithm is given a small training dataset to work with. This training dataset is a smaller part of the bigger dataset and serves to give the algorithm a basic idea of the problem, solution, and data points to be dealt with. The training dataset is also very similar to the final dataset in its characteristics and provides the algorithm with the labelled parameters required for the problem.

The algorithm then finds relationships between the parameters given, essentially establishing a cause and effect relationship between the variables in the dataset. At the end of the training, the algorithm has an idea of how the data works and the relationship between the input and the output.

NEXT

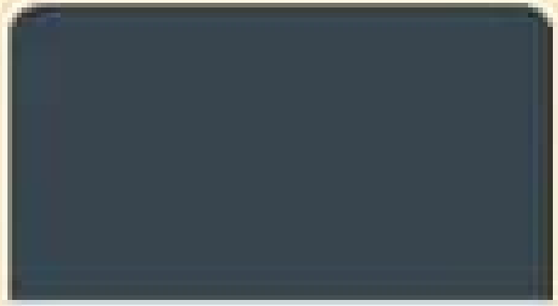
This solution is then deployed for use with the final dataset, which it learns from in the same way as the training dataset. This means that supervised machine learning algorithms will continue to improve even after being deployed, discovering new patterns and relationships as it trains itself on new data.

Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to find a mapping function to map the input variable(x) with the output variable(y).

In the real-world, supervised learning can be used for Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.

NEXT

Known Data



Response

Model

Known Response

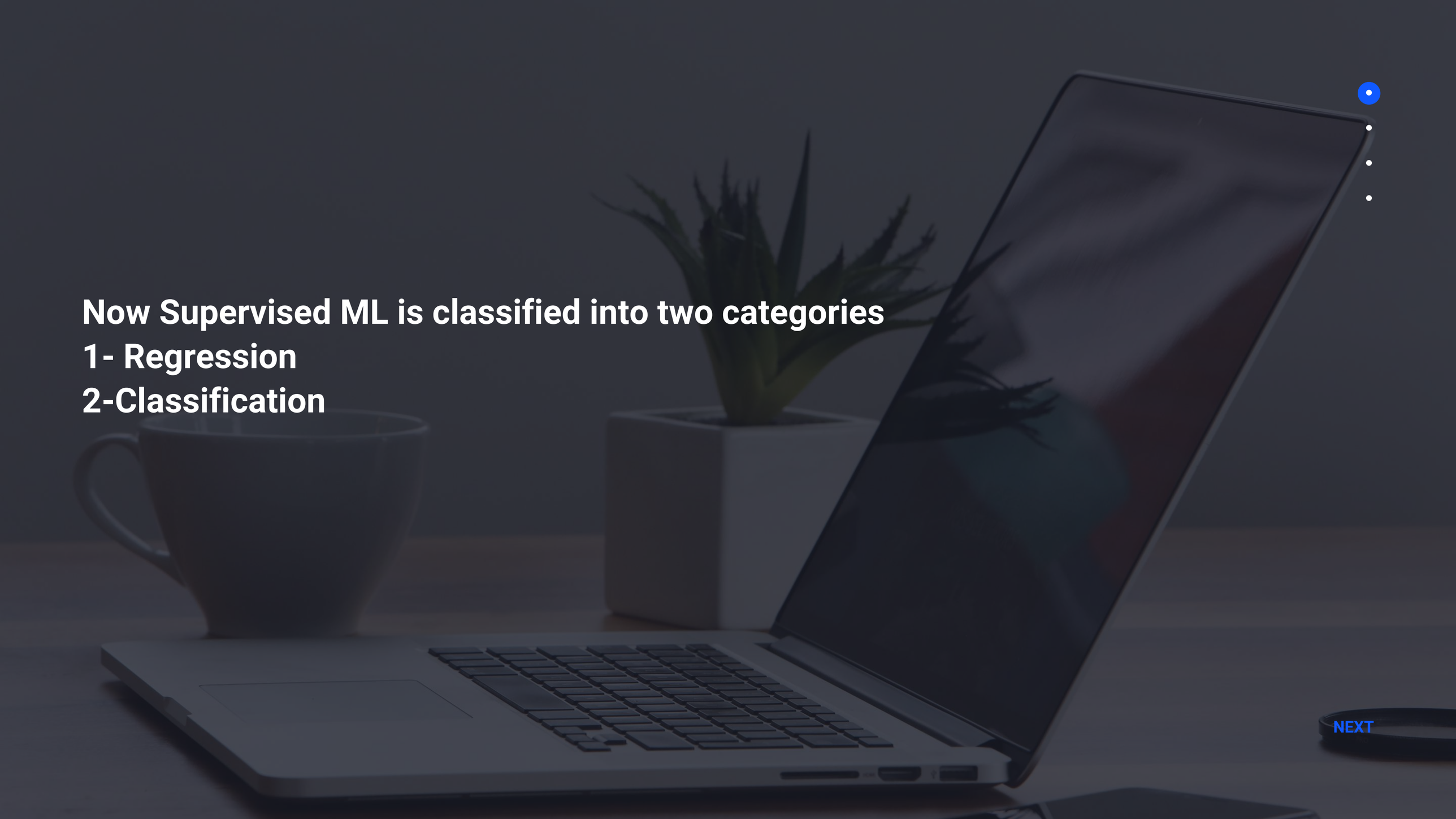
New Data

NEXT

Steps Involved in Supervised Learning:

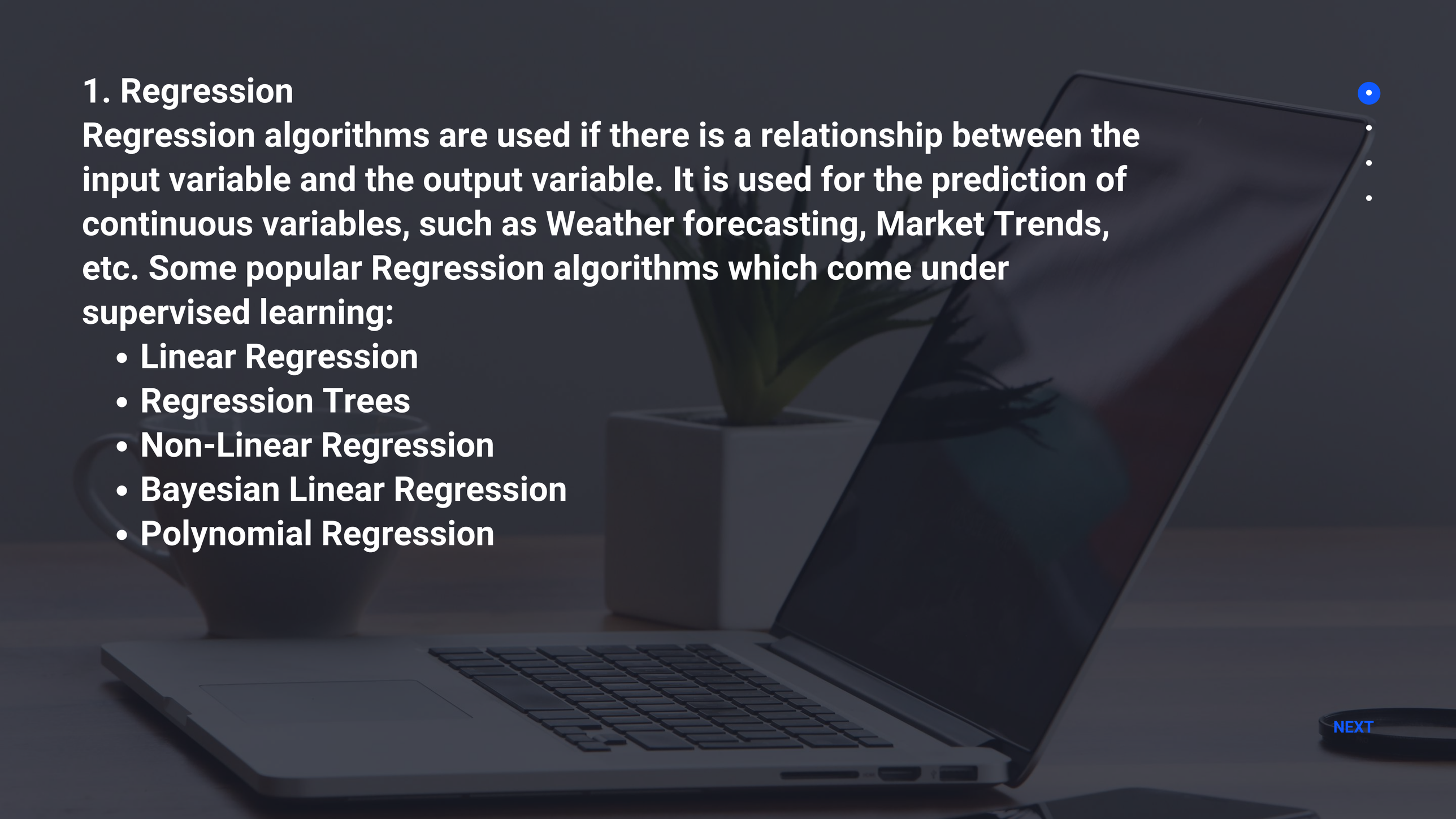
- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training dataset, test dataset, and validation dataset.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

NEXT



Now Supervised ML is classified into two categories
1- Regression
2-Classification

NEXT

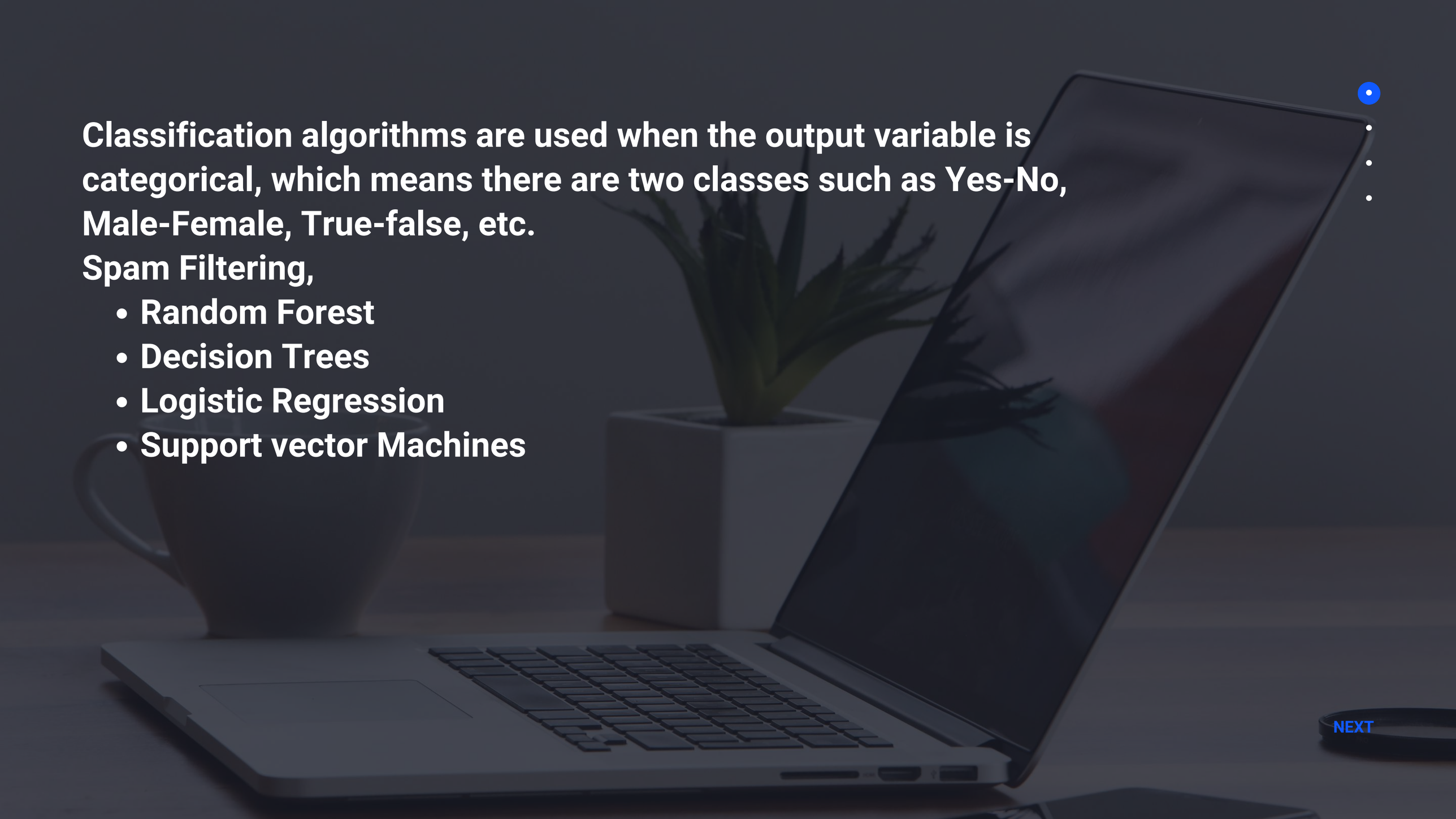


1. Regression

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Some popular Regression algorithms which come under supervised learning:

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

NEXT



Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.
Spam Filtering,

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines

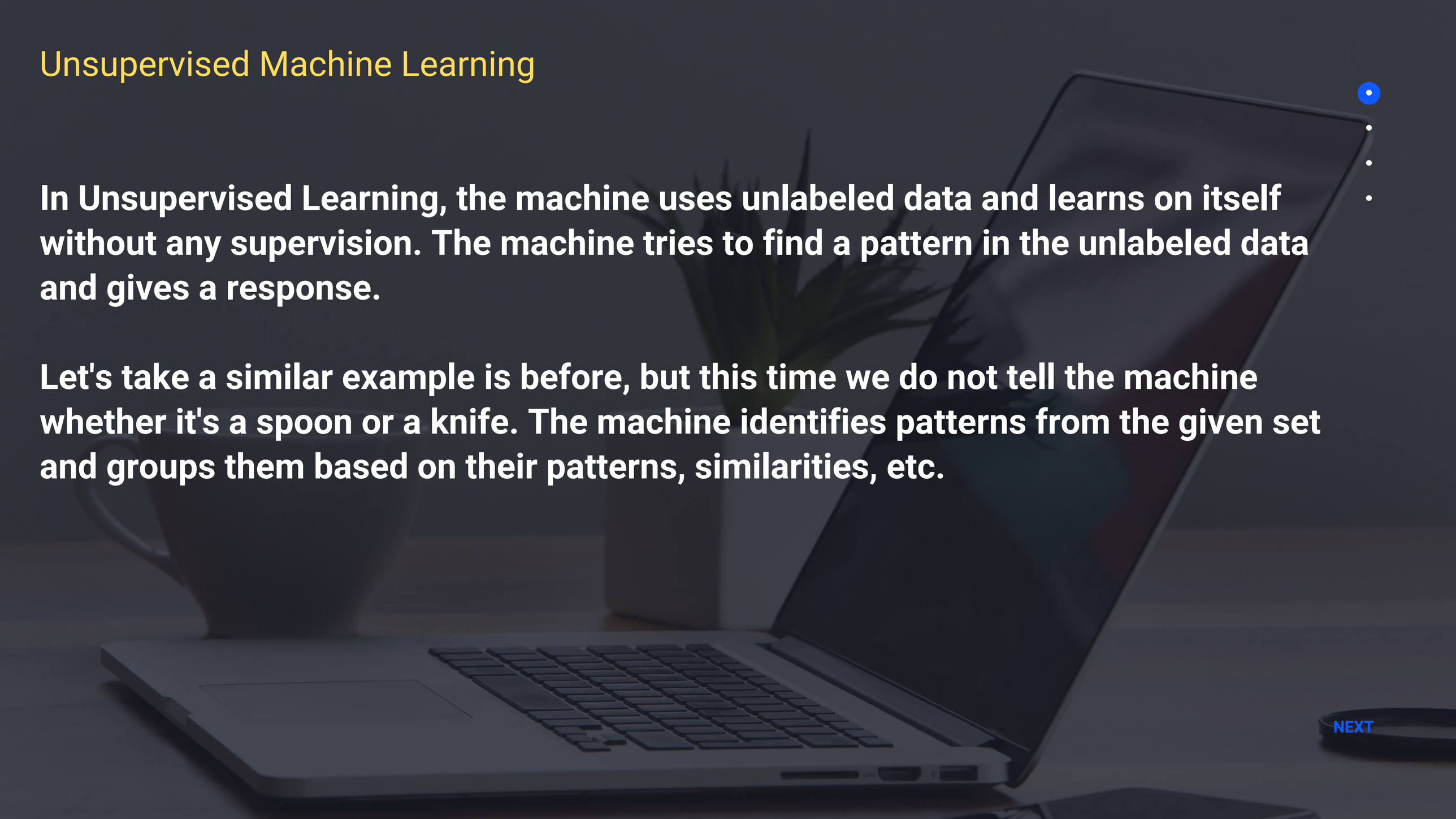
NEXT

- Linear regression: Linear regression is used to identify the relationship between a dependent variable and one or more independent variables and is typically leveraged to make predictions about future outcomes. When there is only one independent variable and one dependent variable, it is known as simple linear regression. As the number of independent variables increases, it is referred to as multiple linear regression. For each type of linear regression, it seeks to plot a line of best fit, which is calculated through the method of least squares. However, unlike other regression models, this line is straight when plotted on a graph.
- Logistic regression: While linear regression is leveraged when dependent variables are continuous, logistic regression is selected when the dependent variable is categorical, meaning they have binary outputs, such as "true" and "false" or "yes" and "no." While both regression models seek to understand relationships between data inputs, logistic regression is mainly used to solve binary classification problems, such as spam identification.

- Real-Life Applications of Supervised Learning
- **Risk Assessment**
- Supervised learning is used to assess the risk in financial services or insurance domains in order to minimize the risk portfolio of the companies.
- **Image Classification**
- Image classification is one of the key use cases of demonstrating supervised machine learning. For example, Facebook can recognize your friend in a picture from an album of tagged photos.
- **Fraud Detection**
- To identify whether the transactions made by the user are authentic or not.
- **Visual Recognition**
- The ability of a machine learning model to identify objects, places, people, actions, and images.
-

NEXT

Unsupervised Machine Learning



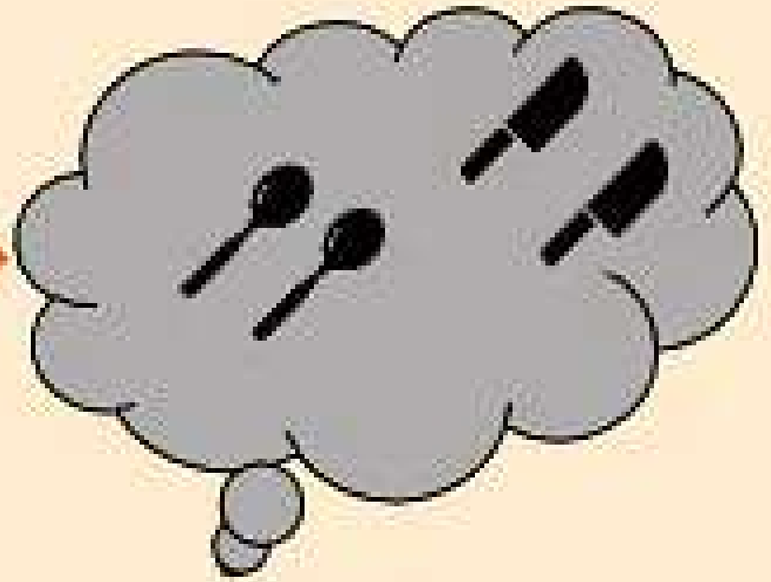
In Unsupervised Learning, the machine uses unlabeled data and learns on itself without any supervision. The machine tries to find a pattern in the unlabeled data and gives a response.

Let's take a similar example is before, but this time we do not tell the machine whether it's a spoon or a knife. The machine identifies patterns from the given set and groups them based on their patterns, similarities, etc.

NEXT

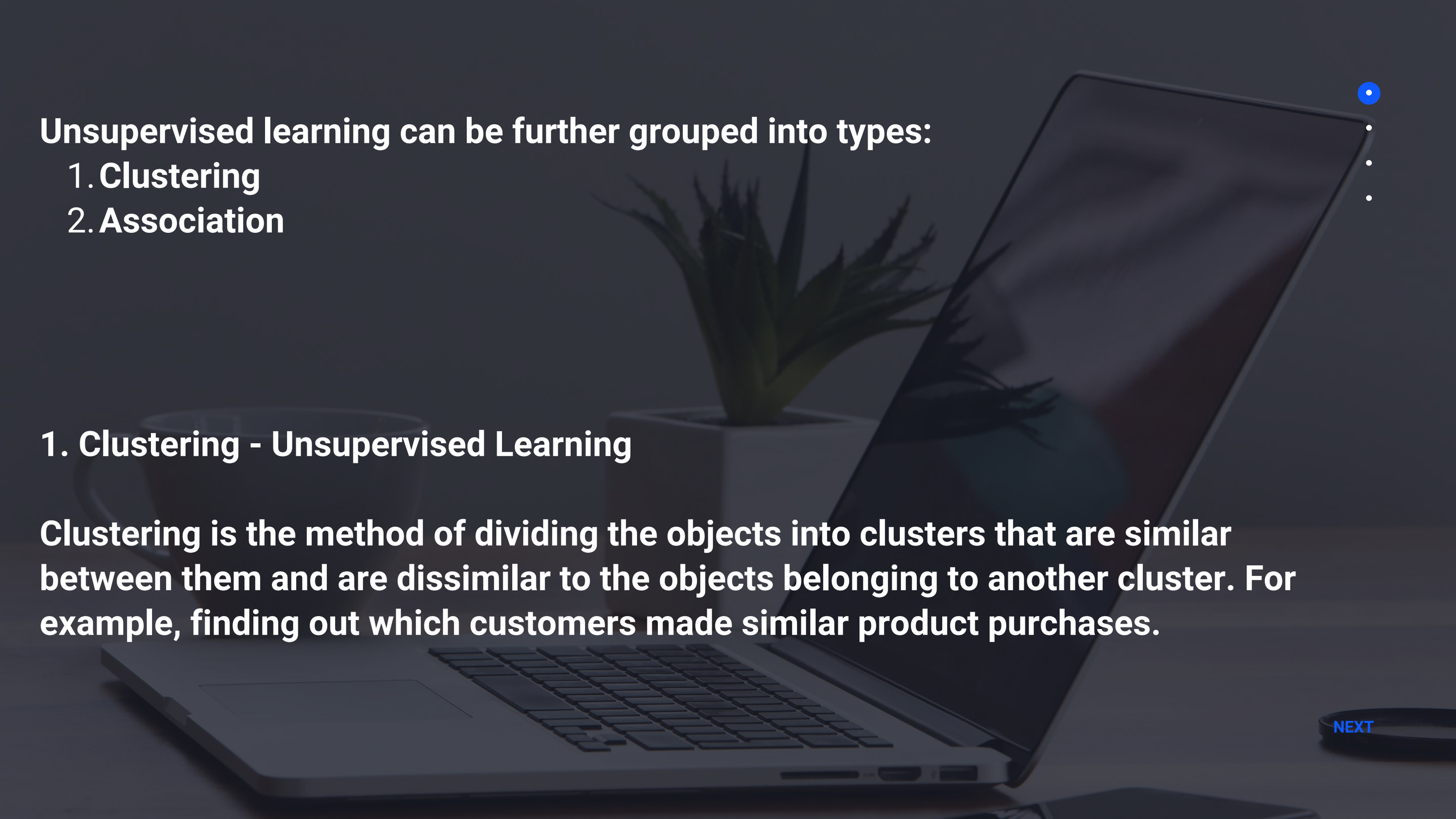
Known Data

Model



Response

NEXT



Unsupervised learning can be further grouped into types:

1. Clustering
2. Association

1. Clustering - Unsupervised Learning

Clustering is the method of dividing the objects into clusters that are similar between them and are dissimilar to the objects belonging to another cluster. For example, finding out which customers made similar product purchases.

NEXT

2. Association - Unsupervised Learning

Association is a rule-based machine learning to discover the probability of the co-occurrence of items in a collection. For example, finding out which products were purchased together.

NEXT

Real-Life Applications of Unsupervised Learning

- **Market Basket Analysis**
- It is a machine learning model based on the algorithm that if you buy a certain group of items, you are less or more likely to buy another group of items.
- **Semantic Clustering**
- Semantically similar words share a similar context. People post their queries on websites in their own ways.

NEXT

Difference Between Supervised and Unsupervised Learning

Supervised Learning	Unsupervised Learning
It uses known and labeled data as input	It uses unlabeled data as input
It has a feedback mechanism	It has no feedback mechanism
The most commonly used supervised learning algorithms are: • Decision tree • Logistic regression • Support vector machine	The most commonly used unsupervised learning algorithms are: • K-means clustering • Hierarchical clustering • Apriori algorithm

NEXT

Reinforcement Learning

Reinforcement learning directly takes inspiration from how human beings learn from data in their lives. It features an algorithm that improves upon itself and learns from new situations using a trial-and-error method. Favorable outputs are encouraged or ‘reinforced’, and non favorable outputs are discouraged or ‘punished’.

Reinforcement learning works by putting the algorithm in a work environment with an interpreter and a reward system. In every iteration of the algorithm, the output result is given to the interpreter, which decides whether the outcome is favorable or not

Reinforcement Learning

In case of the program finding the correct solution, the interpreter reinforces the solution by providing a reward to the algorithm. If the outcome is not favorable, the algorithm is forced to reiterate until it finds a better result. In most cases, the reward system is directly tied to the effectiveness of the result.

Reinforcement Learning

Reinforcement learning is a sub-branch of Machine Learning that trains a model to return an optimum solution for a problem by taking a sequence of decisions by itself.

We model an environment after the problem statement. The model interacts with this environment and comes up with solutions all on its own, without human interference. To push it in the right direction, we simply give it a positive reward if it performs an action that brings it closer to its goal or a negative reward if it goes away from its goal.

Supervised Learning	Unsupervised Learning	Reinforcement Learning
Data provided is labeled data, with output values specified	Data provided is unlabeled data, the outputs are not specified, machine makes its own prediction	The machine learns from its environment using rewards and errors
Used to solve Regression and classification problems	Used to solve Association and clustering problems	Used to solve Reward based problems
Labeled data is used	Unlabeled data is used	No predefined data is used
External Supervision	No supervision	No supervision
Solves problems by mapping labeled input to known output	Solves problems by understanding patterns and discovering output	Follows Trail and Error problem solving approach

Simple Linear Regression

Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable. It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other. For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit. Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.

NEXT

Linear Regression is one of the most fundamental and widely known Machine Learning Algorithms which people start with. Building blocks of a Linear Regression Model are:

- Discreet/continuous independent variables
- A best-fit regression line
- Continuous dependent variable. i.e., A Linear Regression model predicts the dependent variable using a regression line based on the independent variables. The equation of the Linear Regression is:

$$Y=a+b*X + e$$

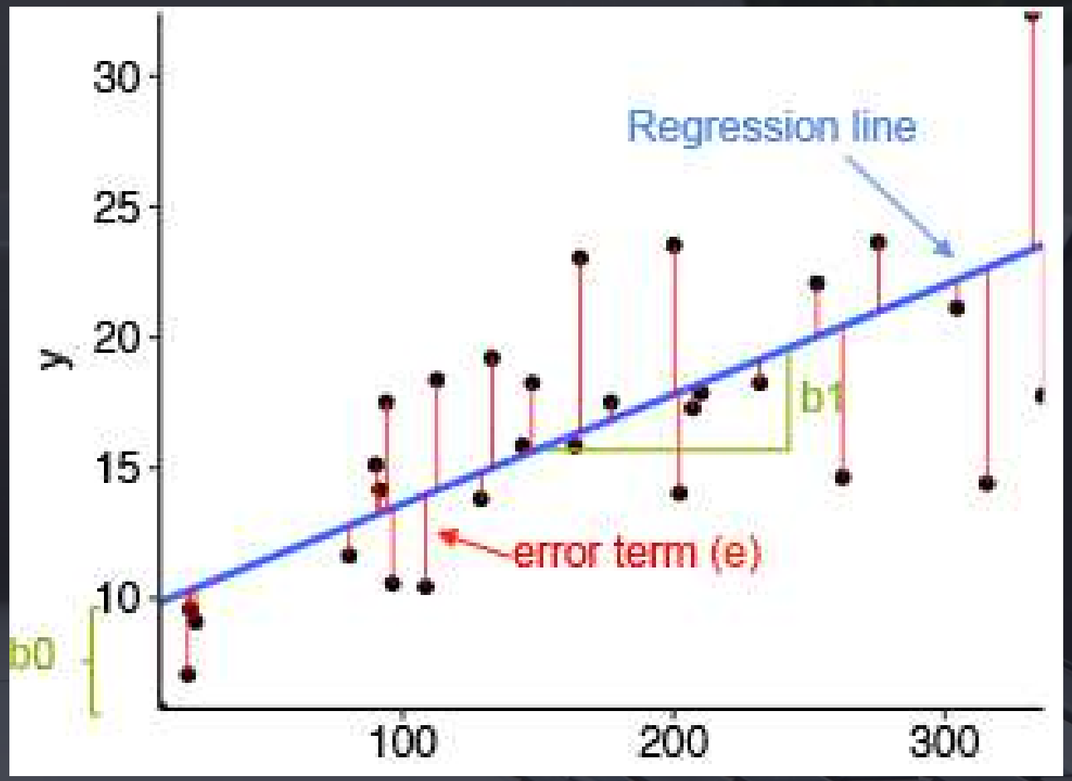
Where, a is the intercept, b is the slope of the line, and e is the error term. The equation above is used to predict the value of the target variable based on the given predictor variable(s).

NEXT

A Linear Regression model's main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized.

Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.

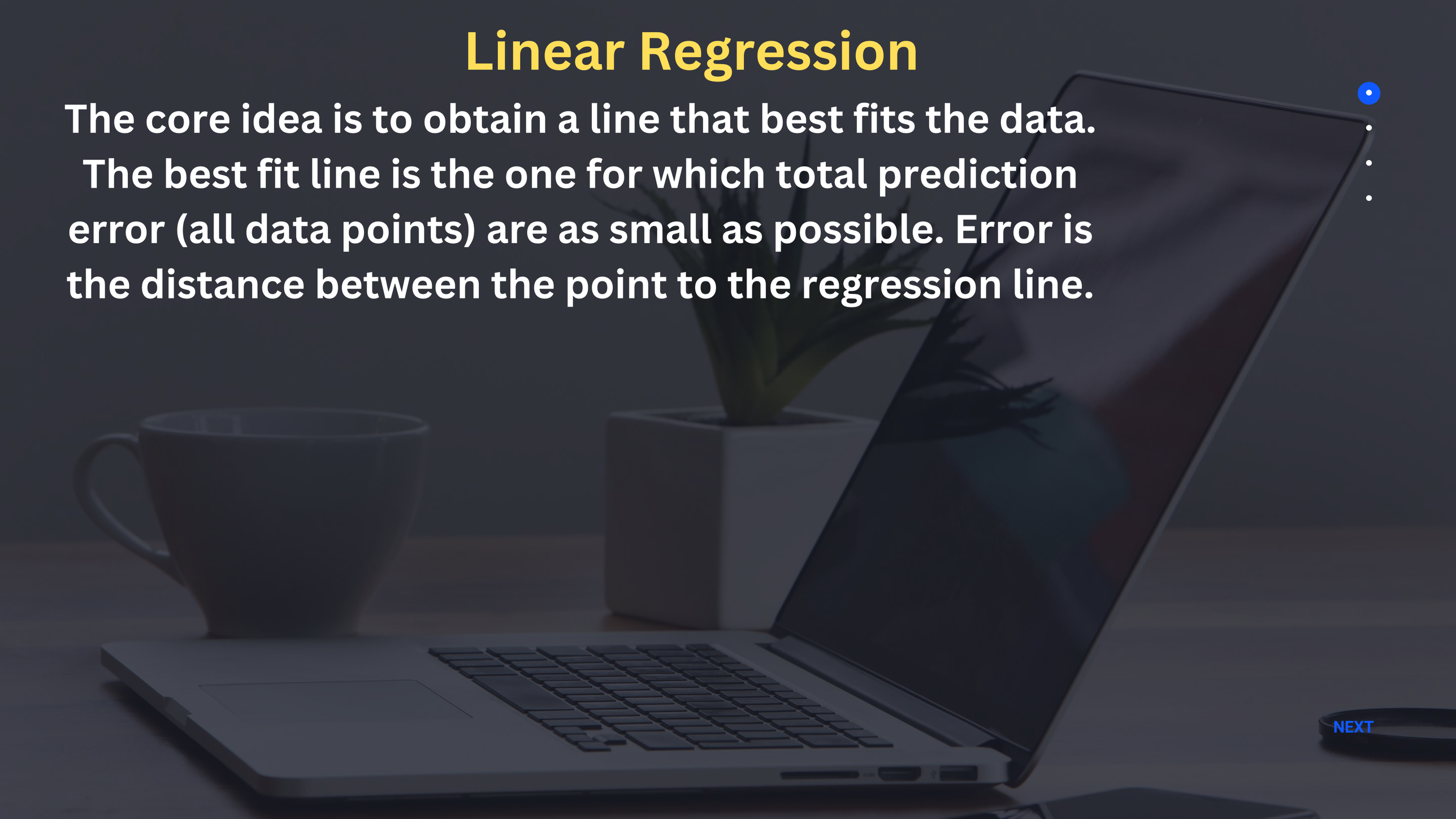
Let's understand this with the help of a diagram.



NEXT

Linear Regression

The core idea is to obtain a line that best fits the data. The best fit line is the one for which total prediction error (all data points) are as small as possible. Error is the distance between the point to the regression line.



NEXT


Thank you!