

Unsupervised Learning and How does it Work

- An Overview of Machine Learning
- What is Unsupervised Learning?
- Why is it important?
- Types of Unsupervised Learning
- Applications of Unsupervised Learning
- Supervised Learning vs. Unsupervised Learning
- Disadvantages of Unsupervised Learning

So take a deep dive and know everything there is to about Unsupervised Machine Learning. Let's get started! :)

An Overview of Machine Learning



Machine Learning, in the simplest of terms, is **teaching your machine about something**. You *collect and clean data, create algorithms, teach the algorithm essential patterns* from the data and then *expect the algorithm to give you a helpful answer*. If the algorithm lives up to your expectations, you have successfully taught your algorithm. If not, just scrap everything and start from scratch. That is how it works here. And if you are looking for a formal definition, Machine Learning is the process of creating models that can perform a certain task without the need for a human **explicitly programming** it to do something.

There are 3 types of Machine Learning which are based on the way the algorithms are created. They are:

- **Supervised Learning** – You supervise the learning process, meaning the data that you have collected here is labelled and so you know what input needs to be mapped to what output. This helps you correct your algorithm if it makes a mistake in giving you the answer.
- **Unsupervised Learning** – The data collected here has no labels and you are unsure about the outputs. So you model your algorithm such that it can understand patterns from the data and output the required answer. You do not interfere when the algorithm learns.

- **Reinforcement Learning** – There is no data in this kind of learning, nor do you teach the algorithm anything. You model the algorithm such that it interacts with the environment and if the algorithm does a good job, you reward it, else you punish the algorithm. With continuous interactions and learning, it goes from being bad to being the best that it can for the problem assigned to it.

Now that we know what is Machine Learning and the different types of Machine Learning, let us dwell into the actual topic for discussion here and answer What is Unsupervised Learning? Where is Unsupervised Learning used? Unsupervised Learning Algorithms and much more.

What is Unsupervised Learning?



Unsupervised Learning, as discussed earlier, can be thought of as **self-learning** where the algorithm can find previously **unknown patterns** in datasets that do not have any sort of labels. It helps in modelling probability density functions, finding anomalies in the data, and much more. To give you a simple example, think of a student who has textbooks and all the required material to study but has no teacher to guide. Ultimately, the student will have to learn by himself or herself to pass the exams. This sort of self-learning is what we have scaled into Unsupervised Learning for machines.

Let me give you a real-life example of where Unsupervised Learning may have been used you to learn about something.

Example of Unsupervised Learning

Suppose you have never watched a cricket match in your entire life and you have been invited by your friends to hang out at their house for a match between India and Australia. You have no idea about what cricket is but just for your friends, you say yes and head over with them. The match starts and you just sit there, blank. Your friends are enjoying the way Virat Kohli plays and want to join in the fun.

Here is when you start learning about the game. You analyse the screen and come up with certain conclusions that you can use to understand the game better.



- There are 2 teams with jerseys of colour Blue and Yellow. Since Virat Kohli belongs to India and you see the score of India on the screen, you conclude that India has the jersey of Blue which makes Australia have yellow Jersey.
- There are different types of players on the field. 2 which belong to India have bats in their hand meaning that they are batting. There is someone who runs up and bowls the ball, making him a bowler. There are around 9 players around the field who try to stop the ball from reaching the boundary of the stadium. There is someone behind the wickets and 2 umpires to manage the match.
- If the ball hits the wickets or if the ball is caught by the fielders, the batsman is out and has to walk back.
- Virat Kohli has the number 18 and his name on the back of his jersey and if this player scores a 4 or a 6, you need to cheer.

You make these observations one-by-one and now know when to cheer or boo when the wickets fall. From knowing nothing to knowing the basics of cricket, you can now enjoy the match with your friends.

What happened here? You had every material that you needed to learn about the basics of cricket. The TV, when and who your friends cheer for. This made you learn about cricket by yourself without someone guiding you about anything. This is the principle that unsupervised learning follows. So having understood what Unsupervised Learning is, let us move over and understand what makes it so important in the field of Machine Learning.

Why is it important?

So what does Unsupervised Learning help us obtain? Let me tell you all about it.

- Unsupervised Learning algorithms work on datasets that are unlabelled and find patterns which would previously not be known to us.
- These patterns obtained are helpful if we need to categorize the elements or find an association between them.

- They can also help detect anomalies and defects in the data which can be taken care of by us.

Lastly and most importantly, data which we collect is usually unlabelled which makes work easier for us when we use these algorithms.

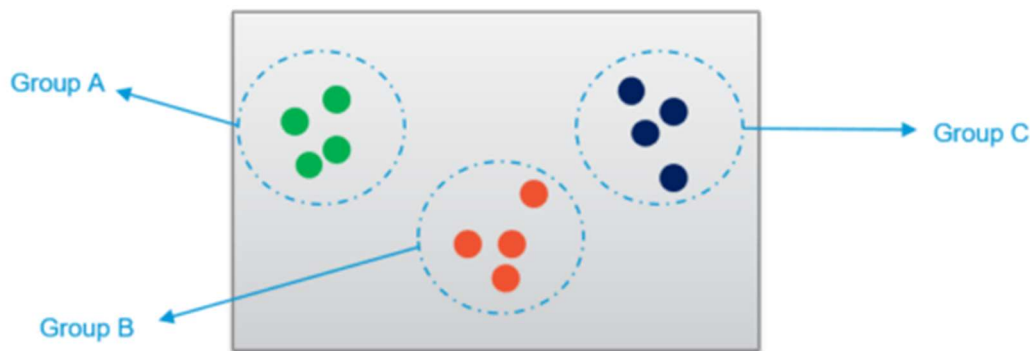
Now that we know the importance, let us move ahead and understand the different types of Unsupervised Learning.

Types of Unsupervised Learning

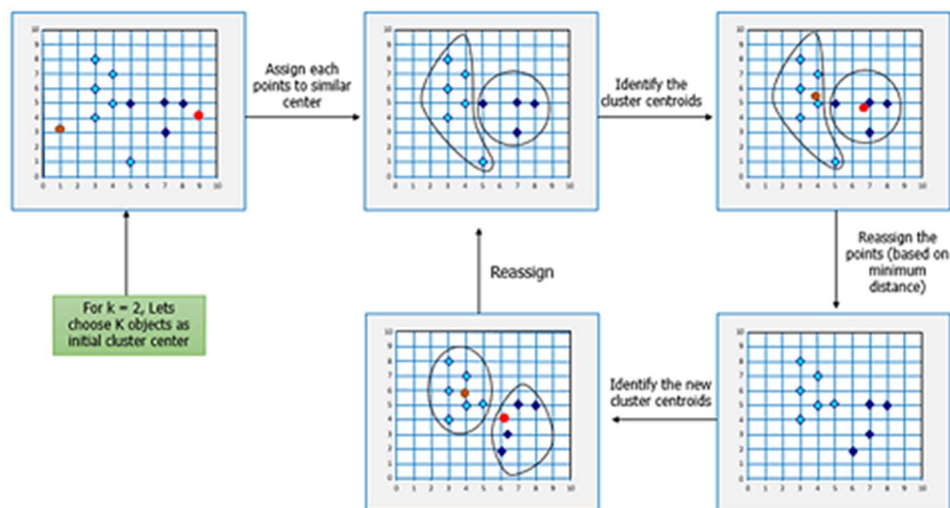
Unsupervised Learning has been split up majorly into 2 types:

- **Clustering**
- **Association**

Clustering is the type of Unsupervised Learning where you find patterns in the data that you are working on. It may be the shape, size, colour etc. which can be used to group data items or create clusters. Some popular algorithms in Clustering are discussed below:



- **Hierarchical Clustering** – This algorithm builds clusters based on the similarity between different data points in the dataset. It goes over the various features of the data points and looks for the similarity between them. If the data points are found to be similar, they are grouped together. This continues until the dataset has been grouped which creates a hierarchy for each of these clusters.
- **K-Means Clustering** – This algorithm works step-by-step where the main goal is to achieve clusters which have labels to identify them. The algorithm creates clusters of different data points which are as homogenous as possible by calculating the centroid of the cluster and making sure that the distance between this centroid and the new data point is as less as possible. The smallest distance between the data point and the centroid determines which cluster it belongs to while making sure the clusters do not interlay with each other. The centroid acts like the heart of the cluster. This ultimately gives us the cluster which can be labelled as needed.



- **K-NN Clustering** – This is probably the most simple of the Machine Learning algorithms as the algorithm does not really learn but rather classifies the new data point based on the datasets that have been stored by it. This algorithm is also called as a lazy learner because it learns only when the algorithm is given a new data point. It works well with smaller datasets as huge datasets take time to learn.

Association is the kind of Unsupervised Learning where you find the dependencies of one data item to another data item and map them such that they help you profit better. Some popular algorithms in Association Rule Mining are discussed below:



- **Apriori algorithm** – The Apriori Algorithm is a breadth-first search based which calculates the support between items. This support basically maps the dependency of one data item with another which can help us understand what data item influences the possibility of something happening to the other data item. For example, bread influences the buyer to buy milk and eggs. So that mapping helps increase profits for the store. That sort of mapping can be learnt using this algorithm which yields rules as for its output.
- **FP-Growth Algorithm** – The Frequency Pattern (FP) algorithm finds the count of the pattern that has been repeated, adds that to a table and then finds the most plausible item and sets that as the root of the tree. Other data items are then added into the tree and the support is calculated. If that particular branch fails to

meet the threshold of the support, it is pruned. Once all the iterations are completed, a tree with the root to the item will be created which are then used to make the rules of the association. This algorithm is faster than Apriori as the support is calculated and checked for increasing iterations rather than creating a rule and checking the support from the dataset.

-

Now that you have a clear understanding between the two kinds of Unsupervised Learning, let us now learn about some of the applications of Unsupervised Learning.

Applications of Unsupervised Learning

Unsupervised Learning helps in a variety of ways which can be used to solve various real-world problems.

- They help us in understanding patterns which can be used to cluster the data points based on various features.
- Understanding various defects in the dataset which we would not be able to detect initially.
- They help in mapping the various items based on the dependencies of each other.
- Cleansing the datasets by removing features which are not really required for the machine to learn from.

This ultimately leads to applications which are helpful to us. Certain examples of where Unsupervised Learning algorithms are used are discussed below:

- **AirBnB** – This is a great application which helps host stays and experiences connecting people all over the world. This application uses Unsupervised Learning where the user queries his or her requirements and Airbnb learns these patterns and recommends stays and experiences which fall under the same group or cluster.
- **Amazon** – Amazon also uses unsupervised learning to learn the customer's purchase and recommend the products which are most frequently bought together which is an example of association rule mining.
- **Credit-Card Fraud Detection** – Unsupervised Learning algorithms learn about various patterns of the user and their usage of the credit card. If the card is used in parts that do not match the behaviour, an alarm is generated which could possibly be marked fraud and calls are given to you to confirm whether it was you using the card or not.

Those were some of the applications where Unsupervised Learning algorithms have shined and shown their grit. Now that we have finished the applications of Unsupervised Learning, let's move ahead to the differences between Supervised

and

Unsupervised

Learning.

Supervised Learning vs. Unsupervised Learning

Parameter	Supervised Learning	Unsupervised Learning
Dataset	Labelled Dataset	Unlabelled Dataset
Method of Learning	Guided learning	The algorithm learns by itself using dataset
Complexity	Simpler method	Computationally complex
Accuracy	More Accurate	Less Accurate

Disadvantages of Unsupervised Learning

Even though Unsupervised Learning is used in many well-known applications and works brilliantly, there are still many disadvantages to it.

- There is no way of obtaining the way or method the data is sorted as the dataset is unlabelled.
- They may be less accurate as the input data is not known and labelled by the humans making the machine do it.
- The information obtained by the algorithm may not always correspond to the output class that we required.
- The user has to understand and map the output obtained with the corresponding labels.

Those are basically the major disadvantages that you may face when you work with Unsupervised Learning algorithms. So now, let us move ahead and summarize everything that you have learned in the article.

We had an overview of what Machine Learning is and its various types. We then understood in depth of what unsupervised learning is, why is it so important. Later, we went through the various types of Unsupervised Learning which are Clustering and Association Mining. After that, we discussed the various algorithms, the applications of Unsupervised Learning, differences between Supervised and Unsupervised Learning and the disadvantages that you may face when you work with Unsupervised Learning Algorithms.

That brings us to the end of the article. I hope it has helped you understand what Unsupervised Learning is in a clear and precise manner. Till next time, Happy Learning!