Machine learning space encompasses 2 significant tasks which aids in data analysis and output prediction. They are called **Supervised Machine Learning** and **Unsupervised Machine Learning**.

Supervised Machine Learning works on labeled data, which is fed as the input, to generate a training model that is capable of making future predictions.

Unsupervised Machine Learning on the other hand, works on unlabeled data to identify the patterns/relationships/ hidden features in the given dataset.

Labeled data - When each input has a corresponding output, it is termed as labeled data.
Unlabeled data - When there is no predefined output for each input, it is termed as unclassified or unlabeled data.

Listed below are the major differences between **Supervised Machine Learning** and **Unsupervised Machine Learning**.-

	Supervised Machine Learning	**Unsupervised Machine Learning**
Training Dataset	Labeled data is used to build a training model. i.e. both input and output labels are present in the dataset	Unlabeled data is used for training the algorithm. .i.e. no predefined output for each input
Usage	It is used to predict future end results. Continuous values for regression analysis and class labels in classification tasks.	It is used for analyzing the data. Identify hidden patterns and categorize them.
Aim	To approximate the relationship, so that when a new input data point is fed, the model can predict the output variable.	To model the distribution of the data in order to gain more insights from the dataset.
Problem domain	Suitable for problems with a clear, measurable and well-defined end goals.	Suitable for problems with no specific goal, but to traverse through the dataset to generate insights, patterns/relationships.
Nature of Results	Results are accurate and reliable.	Results are creative in nature because it generates patterns/clusters/groups.
Number of classes	Known.	Not known.
Input parameter	Both input and output data (labeled) are furnished to develop the training model for the purpose of predicting an outcome.	Only input data is furnished to the model for pattern generation.
Supervision	Needs supervision for training model. Output is present in the dataset. There is clear distinction between the input and the output variable.	Does not need supervision for training model. No output is present. Needs to identify the classes as part of the model prediction.
Categories	Can be categorized into Classification and Regression.	Dimensionality reduction, Clustering and Autoencoders.

Feedback	Accepts direct feedback to check if the output prediction is right or wrong.	The model does not take any feedback.

Categories of Supervised Learning-

1. **Classification**- It is used to specify data into specific class. It predicts discrete labels or categories for the data points (Output is always categorical). It can be either binary classification or of multi-class classification.
2. **Regression**- It predicts continuous numerical values (Output is always a numerical value).

Categories of Unsupervised Learning-

1. **Dimensionality Reduction**- In this method, data is represented using lesser columns/features while preserving data integrity.
2. **Clustering**: Groups together unlabeled data according to "distance", i.e based on their similarities or differences.
3. **Autoencoders**: It encodes input distribution into a common pattern (representations) for all samples and then decode the representations back into input space. Here we use neural network for representation.

Use Cases -

Supervised Learning -

1. Spam email detection
2. Time series forecasting (stock price prediction, weather forecasting)
3. Object Recognition, Image Tagging. (OCR- optical character recognition)

Unsupervised Learning -

1. Customer segmentation
2. in the fields of Bioinformatics and Genetic engineering (DNA pattern clustering)
3. Recommendation system (news feed generation, relevant ads generation)

Pros and Cons-

Supervised Learning-

Pros:

1. We can use this model to predict future outcome based on some prior experiences.
2. This model gives us the accurate idea on the object classes.
3. Helps us address different real time problems such as spam detection, fraud identification, etc

Cons:

1. Building the training models takes long computation time.
2. It cannot generate the right result if the data to be tested is different from the training model.
3. Not best suited for dealing with complex tasks.

Unsupervised Learning -

Pros:

1. It is comparatively easier to get unlabeled data.
2. Yields underlying, hidden pattern which is previous unknown.

Cons:

1. It is complex as compared to Supervised Learning.
2. Since the input dataset is not labeled/unclassified and we have no prior idea of the result to be generated, therefore the output generated by this algorithm may not be very accurate.

Note: It is important to pre-process the data in order to get rid of any anomalies (noise/outliers, null values, etc.)