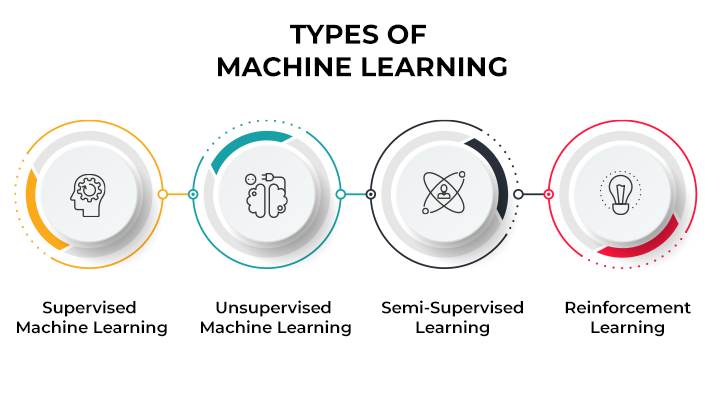
# Machine Learning

* Machine learning is a branch of artificial intelligence that teaches computers to learn and improve from data and experience without being explicitly programmed.
* It involves training algorithms on vast datasets to identify patterns, make predictions, and classify information, leading to systems that can enhance their own performance over time.

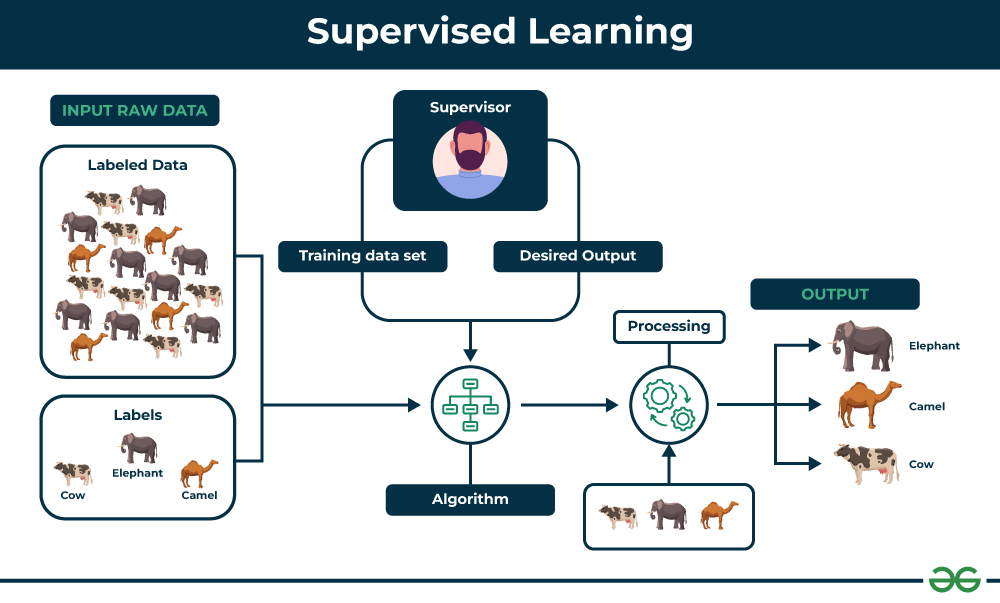
## Types of Machine Learning

1. Supervised Learning
2. Unsupervised Learning
3. Semi-Supervised Learning
4. Reinforcement Learning



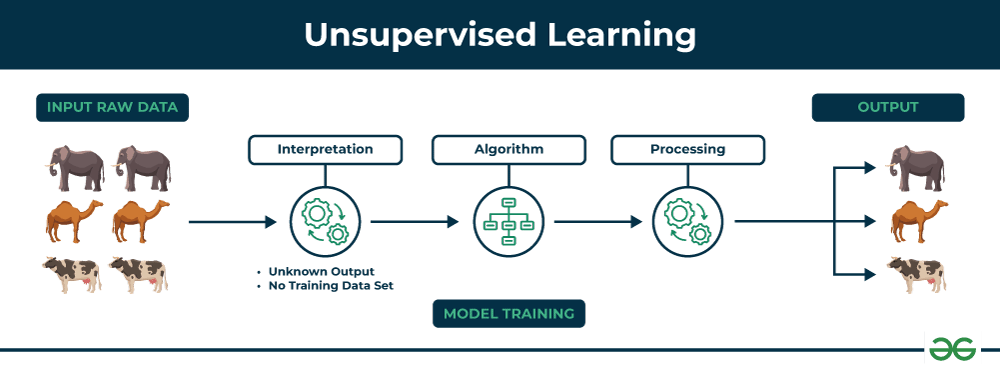
### Supervised Learning:

* Supervised Learning is defined as when a model gets trained on a "Labelled Dataset". Labelled datasets have both input and output parameters. In Supervised Learning algorithms learn to map points between inputs and correct outputs. It has both training and validation datasets labelled.
* Algorithms learn from labeled data with input-output pairs. Common algorithms include:
  + Classification: Logistic Regression, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Naive Bayes, Decision Trees, Random Forest, Gradient Boosting (e.g., XGBoost, LightGBM, CatBoost), Neural Networks.
  + Regression: Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression (SVR), Decision Trees Regression, Random Forest Regression, Gradient Boosting Regression, Neural Networks Regression.



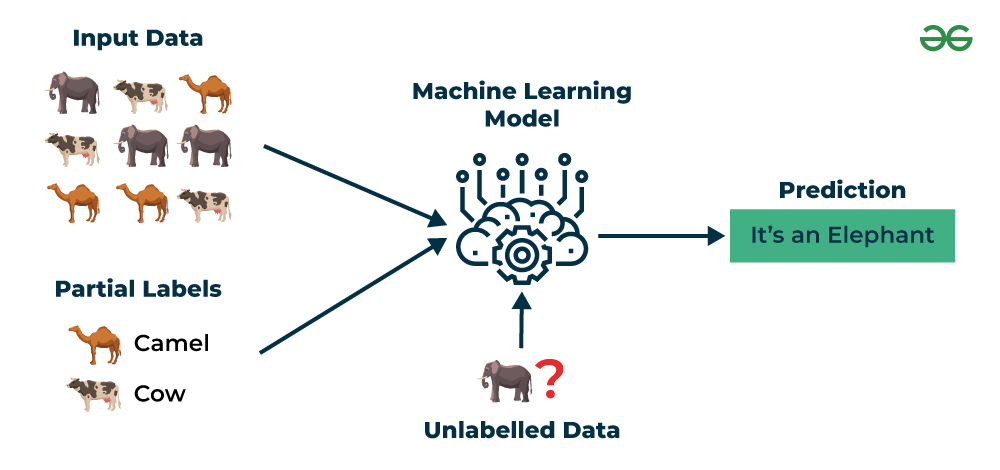
### Unsupervised Learning:

* Unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabeled data. Unlike supervised learning, unsupervised learning doesn't involve providing the algorithm with labeled target outputs. The primary goal of Unsupervised learning is often to discover hidden patterns, similarities, or clusters within the data, which can then be used for various purposes, such as data exploration, visualization, dimensionality reduction, and more.
* Algorithms work on unlabeled data to find patterns or structure. Types include:
  + Clustering: K-means, Hierarchical Clustering, DBSCAN, Gaussian Mixture Models (GMM).
  + Dimensionality Reduction: Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), UMAP.
  + Association: Apriori Algorithm, Eclat Algorithm.



### Semi - Supervised Learning:

* Semi-Supervised learning is a machine learning algorithm that works between the supervised and unsupervised learning so it uses both labelled and unlabelled data. It's particularly useful when obtaining labeled data is costly, time-consuming, or resource-intensive. This approach is useful when the dataset is expensive and time-consuming. Semi-supervised learning is chosen when labeled data requires skills and relevant resources in order to train or learn from it.
* We use these techniques when we are dealing with data that is a little bit labeled and the rest large portion of it is unlabeled. We can use the unsupervised techniques to predict labels and then feed these labels to supervised techniques. This technique is mostly applicable in the case of image data sets where usually all images are not labeled.
* Types of Semi-Supervised Learning Algorithms
  + Self-Training
    - This method starts by training a model on a small labeled dataset. The trained model then predicts labels (pseudo-labels) for the unlabeled data. The high-confidence predictions are added to the labeled dataset iteratively, refining the model over multiple rounds.
  + Co-Training
    - Co-Training involves training two classifiers on two different "views" or feature sets of the data. Each classifier labels the unlabeled dataset, and the most confident predictions from one classifier are used to improve the other one. This iterative process leverages complementary information from different data perspectives.
  + Graph-Based Methods
    - These algorithms represent data points as nodes in a graph, with edges reflecting similarities. Label information propagates through the graph to assign labels to unlabeled points based on their proximity to labeled nodes.
  + Generative Models
    - Models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) learn the underlying data distribution by using both labeled and unlabeled data. They help generate meaningful features or data points that improve classification tasks.



### Reinforcement Learning:

* Reinforcement machine learning algorithm is a learning method that interacts with the environment by producing actions and discovering errors.
* Trial, error, and delay are the most relevant characteristics of reinforcement learning. In this technique, the model keeps on increasing its performance using Reward Feedback to learn the behavior or pattern.
* These algorithms are specific to a particular problem e.g. Google Self Driving car, AlphaGo where a bot competes with humans and even itself to get better and better performers in Go Game. Each time we feed in data, they learn and add the data to their knowledge which is training data. So, the more it learns the better it gets trained and hence experienced.
* Algorithms learn to make sequences of decisions by receiving rewards or penalties.
  + Model-Free Methods: Q-Learning, Deep Q-Network (DQN), SARSA, Policy Gradient Methods.
  + Model-Based Methods: Deep Deterministic Policy Gradient (DDPG), Proximal Policy Optimization (PPO), Trust Region Policy Optimization (TRPO).
  + Value-Based Methods: Monte Carlo Methods, Temporal Difference (TD) Learning.

