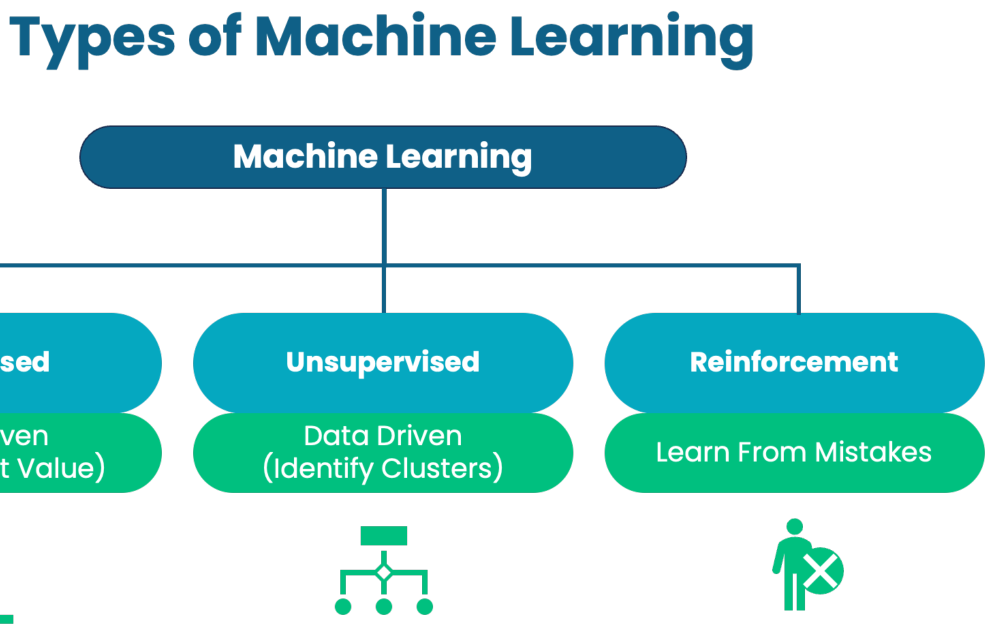
Types of Machine Learning and Applications

There are three main types of machine learning:



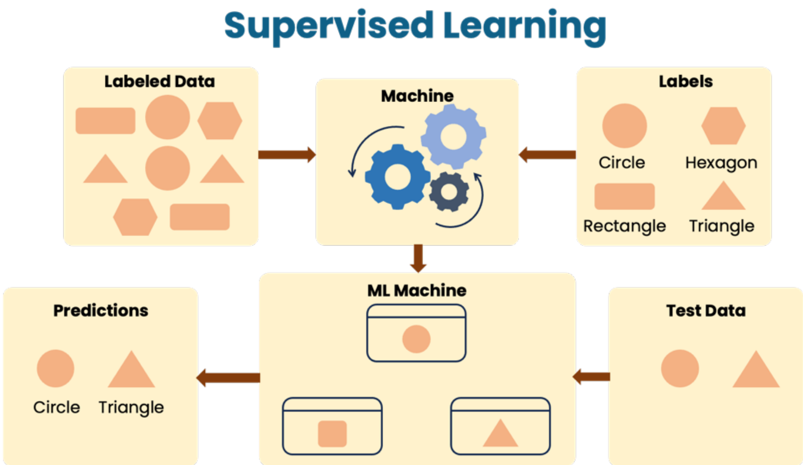
Alt text: Types of Machine Learning

a. Supervised learning: In supervised learning, the algorithm is trained on a **labelled** dataset, where the correct output is provided for each input. The goal is to learn a **mapping** between inputs and outputs, so that the algorithm can make accurate predictions on new, unseen data.

Supervised learning algorithms involve machine learning techniques that rely on supervision, utilizing labelled data for training to predict output. Labelled data entails the machine being aware of input data and its corresponding output during training, enabling it to predict output in the testing process. Well-trained models should offer accurate predictions close to real-world outputs for new input data sets.

To illustrate the functioning of supervised machine learning algorithms, consider the following example. A machine is provided with a training dataset containing images of bicycles and cars. Through the labelled training data, the machine comprehends the features of these images.

Subsequently, when presented with an input image of a car during the testing phase, a successful training should result in the machine correctly categorizing the image as a car.

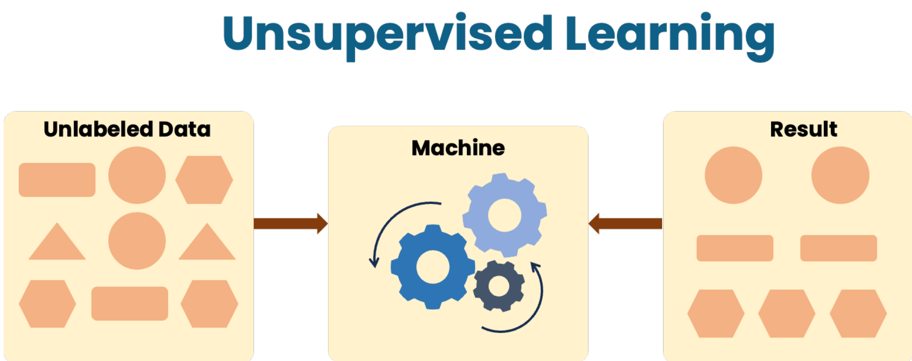


Alt text: Supervised Learning

b. **Unsupervised learning**: In unsupervised learning, the algorithm is trained on an unlabeled dataset, where no output is provided. The goal is to identify patterns or structure in the data, without any prior knowledge of what the output should be.

Unlike supervised learning, unsupervised machine learning doesn't require supervision during training. It utilizes unlabeled data for training, and the output prediction is unsupervised. Unlabeled data lacks prior identification in the machine. In unsupervised algorithms, the machine learns from unlabeled data to independently recognize and classify objects.

For example, when presented with random, uncategorized toy images, the machine autonomously analyzes and categorizes each toy based on its features.

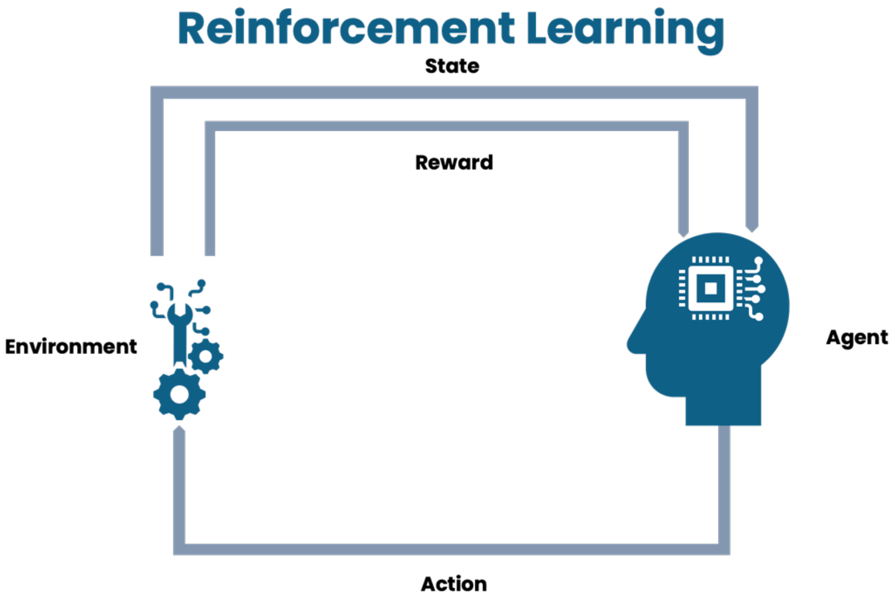


Alt text: Unsupervised Learning

c. **Reinforcement learning**: In reinforcement learning, the algorithm learns by interacting with an **environment** and receiving **feedback** in the form of rewards or penalties. The goal is to learn a policy that maximizes the **cumulative reward** over time.

Reinforcement learning, akin to human learning through trial and error, is a machine learning algorithm. In deep reinforcement learning, the agent learns by interacting with its environment, receiving rewards (positive or negative) in return.

Reinforcement learning involves three key components: the Agent (learner/decision-maker), the Environment (what the agent interacts with), and the **Actions** (what the agent does). Unlike algorithms with labelled data, reinforcement learning agents learn solely from interacting with their environment. This approach relies on feedback, as the agent aims to take actions that maximize positive rewards.



Alt text: Reinforcement Learning

Data in Machine Learning

Machine learning models utilize data to discern patterns and connections between input variables and target outputs, enabling subsequent analysis. The data is typically categorized into two main types:

1. **Labeled Data:** Labeled data comprises a dataset in which each instance is accompanied by labels or annotations, indicating the correct output or class.

2. **Unlabeled Data:** Unlabeled data lacks explicit annotations or class labels. Examples of unlabeled data encompass photo images, audio and video files, or news articles.

   Example: Suppose we have an image dataset with labeled data, and the images depict various types of fruits. Each image is tagged with the corresponding fruit name. The dataset includes images of apples, oranges, and bananas.

   An image tagged with "apple" or "orange" or "banana" represents a labeled example in the dataset.

In this scenario, the unlabeled data could consist of images that may include a mix of fruits, but without explicit annotations or class labels.

When we process raw data to transform it into a meaningful or useful form, it becomes information. At the beginning of each machine learning life cycle, copious amounts of data are generated. To manage this data effectively, it is categorized into three sections:

1. **Training Data:** Comprising input-output pairs, where each input is associated with a corresponding correct output or label. The algorithm employs these examples to grasp underlying patterns and relationships in the data.

2. **Testing Data:** Reserved for evaluating the performance of a trained machine learning model. This data is distinct from the training data and ensures that the model is assessed on examples it has not encountered during the training phase.

   During testing, the model predicts values without access to the actual output. The model's performance is then evaluated by comparing its predictions with the actual output present in the testing data, demonstrating how much the model has learned from the training data.

3. **Validation Data:** A subset of the dataset used to assess the model's performance during training, particularly when tuning hyperparameters or training the model. It provides an unbiased evaluation of the model fit.

   Example: Imagine we have a dataset containing 1,200 labeled images of flowers, comprising three types: roses, sunflowers, and tulips. The distribution of these labeled images is as follows:

   - Roses: 500 images
   - Sunflowers: 400 images
   - Tulips: 300 images

   To effectively train, validate, and test a machine learning model on this dataset, we could divide it as follows:

   - **Training Data (70%):**
     - Total images: 840 (70% of 1,200)
     - Roses: 350 images (70% of 500)
     - Sunflowers: 280 images (70% of 400)
     - Tulips: 210 images (70% of 300)
   - **Validation Data (15%):**
     - Total images: 180 (15% of 1,200)
     - Roses: 75 images (15% of 500)
     - Sunflowers: 60 images (15% of 400)
     - Tulips: 45 images (15% of 300)
   - **Testing Data (15%):**
     - Total images: 180 (15% of 1,200)
     - Roses: 75 images (15% of 500)
     - Sunflowers: 60 images (15% of 400)
     - Tulips: 45 images (15% of 300)

   This distribution ensures that the machine learning model is trained on a substantial portion of the data, validated on a separate set, and finally tested on a completely distinct set of images to assess its overall performance.

Applications of Machine Learning:

Machine learning entails creating algorithms and statistical models, allowing computer systems to learn from data without explicit programming. The necessity for machine learning arises due to the limitations of traditional rule-based programming, especially in solving complex problems with substantial data. These algorithms automatically detect patterns and relationships, enabling predictions or decisions. Machine learning has a wide range of applications in various fields, including:

- **Image and speech recognition**: Machine learning algorithms can be used to automatically identify objects in images or transcribe speech into text.

- **Natural language processing**: Machine learning algorithms can be used to automatically analyze and understand human language, and perform tasks such as sentiment analysis or language translation.
- **Recommendation systems**: Machine learning algorithms can be used to automatically recommend products or services to users based on their past behavior or preferences.
- **Fraud detection**: Machine learning algorithms can be used to automatically detect fraudulent transactions or activities.
- **Disease Diagnosis:** Machine learning can assist in diagnosing diseases by analyzing medical images, patient records, and genetic data.
- **Credit Scoring:** ML algorithms analyze credit history and other financial data to assess the creditworthiness of individuals or businesses.