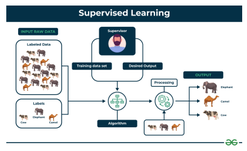
Supervised and Unsupervised Learning

Supervised Learning

- Trained on set of labeled data
- Input data is paired with the desired output data
- Learns to the output for new input data



- Applications
  - Spam filtering — To identify and classify spam emails based on their content
  - Image classification — Classify images into different categories, such as animals, objects, or scenes
  - Medical diagnosis — Analyze patient data, such as medical images, test results, and patient history, to identify patterns that suggest specific diseases
  - Fraud detection — Analyze financial transactions and identify patterns that indicate fraudulent activity
  - Natural language processing — Understand and process human language effectively Eg: Sentiment Analysis
- Commonly used algorithms — Linear and Logistics regression, Random forest, Multi-class classification, Decision tree, Support Vector Machine, Neural Network
- Types
  - Regression
    - When the output variable is a real value, such as “rupees” or “height”
    - Common model evaluation metrics — Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (Coefficient of Determination)
  - Classification
    - When the output variable is a category, such as “Red” or “blue” , “disease” or “no disease”
    - Common model evaluation metrics — Accuracy, Precision, Recall, F1 score, Confusion matrix

Unsupervised Learning

- Trained on a set of unlabeled data
- Input data is not paired with the desired output
- Learns to find patterns and relationships in the data

- Applications
  - Anomaly detection — identify unusual patterns or deviations from normal behavior in data, enabling the detection of fraud, intrusion, or system failures
  - Scientific discovery — uncover hidden relationships and patterns in scientific data
  - Recommendation systems — identify patterns & similarities in user behavior & preferences; recommend products, movies, or music
  - Customer segmentation — identify groups of customers with similar characteristics; target marketing campaigns & improve customer service
  - Image analysis — can group images based on their content; eg: object dection
- Commonly used algorithms — K-Means clustering, Hierarchical clustering, KNN, Apriori algorithm
- Types
  - Clustering
    - Used to group similar data points together
    - For example, grouping customers based on their purchasing behavior
    - Hands-on (K-Means)

```
Initialize k means with random values
--> For a given number of iterations:

    --> Iterate through items:

        --> Find the mean closest to the item by calculating
        the euclidean distance of the item with each of the means

        --> Assign item to mean

        --> Update mean by shifting it to the average of the items in that cluster
```

  - Dimensionality Reduction
    - Reducing the number of features in the data while retaining as much information as possible
    - For example, reducing the number of variables in a dataset to make it easier to visualize