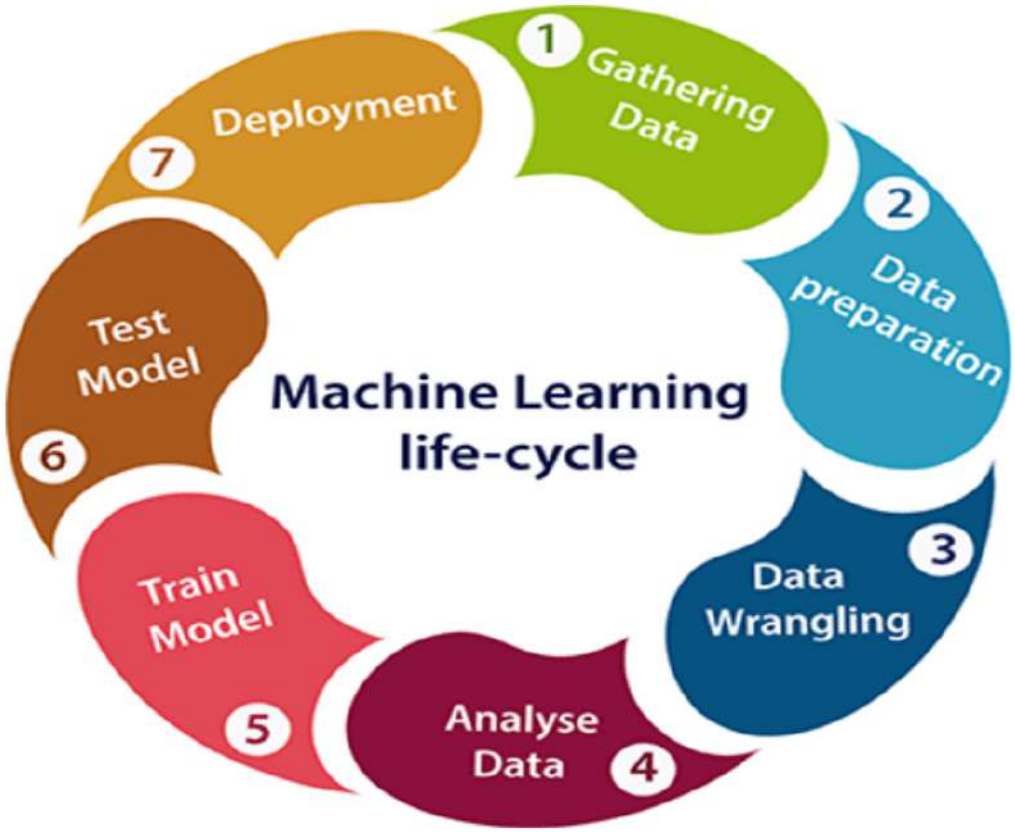
Lifecycle of a ML Project


Machine Learning life-cycle
1 Gathering Data
2 Data preparation
3 Data Wrangling
4 Analyse Data
5 Train Model
6 Test Model
7 Deployment

Gathering Data

- **Description:** The first step in any machine learning project is to gather relevant data.Data can be of any form Structured (or) Unstructured
- **Objective:** Collect data that is representative of the problem you're trying to solve.
- **Methods:** Data can be collected from various sources such as databases, web scraping, APIs, and more.

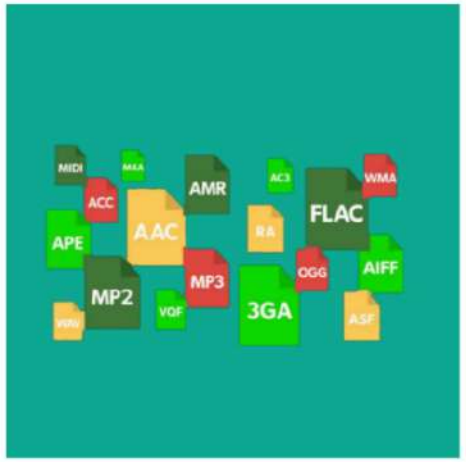
Few Examples of Forms of Data

company	division	sector	tryint
00nil_Combined_Company	00nil_Combined_Division	00nil_Combined_Sector	14625
apple	00nil_Combined_Division	00nil_Combined_Sector	10125
apple	hardware	00nil_Combined_Sector	4500
apple	hardware	business	1350
apple	hardware	consumer	3150
apple	software	00nil_Combined_Sector	5625
apple	software	business	4950
apple	software	consumer	675
microsoft	00nil_Combined_Division	00nil_Combined_Sector	4500
microsoft	hardware	00nil_Combined_Sector	1890
microsoft	hardware	business	855
microsoft	hardware	consumer	1035
microsoft	software	00nil_Combined_Sector	2610
microsoft	software	business	1215
microsoft	software	consumer	1395

Table (Structured data)

Image

Text



Audio

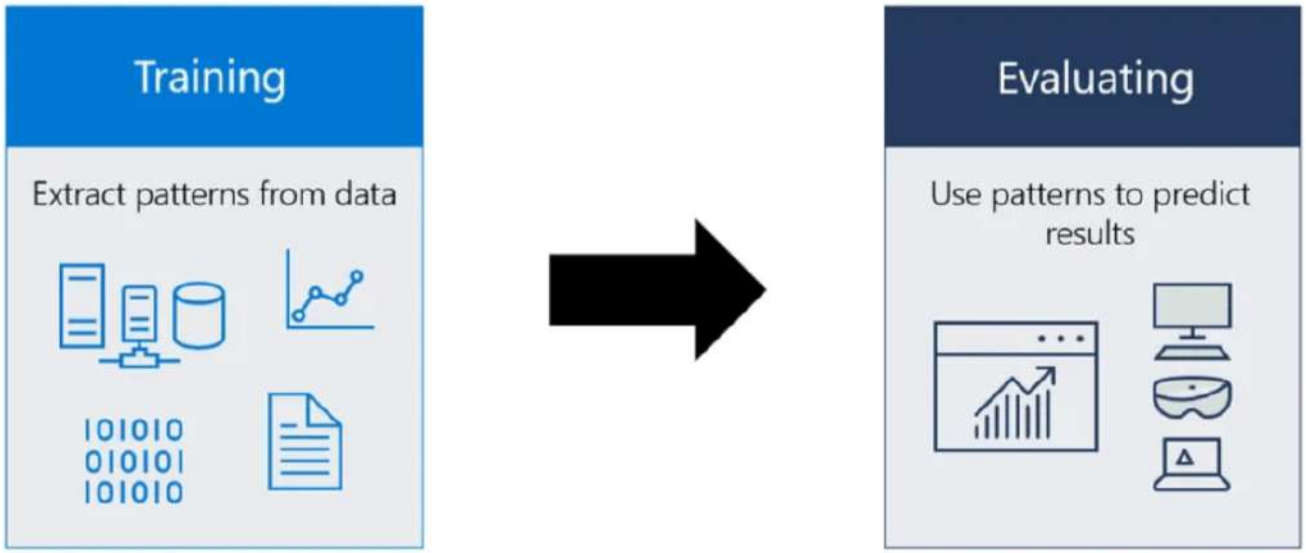
Data Pipeline

- **Data Preparation:** Clean and organize data into a usable format by removing inaccuracies, filling missing values, and converting data appropriately.
- **Data Wrangling:** Transform and map raw data into a structured format that can be easily analyzed by merging, filtering, and aggregating data.
- **Data Preparation Techniques:** Data cleaning, normalization, and transformation.
- **Data Wrangling Techniques**: Merging data from different sources, filtering data, and aggregating data.

Train Model

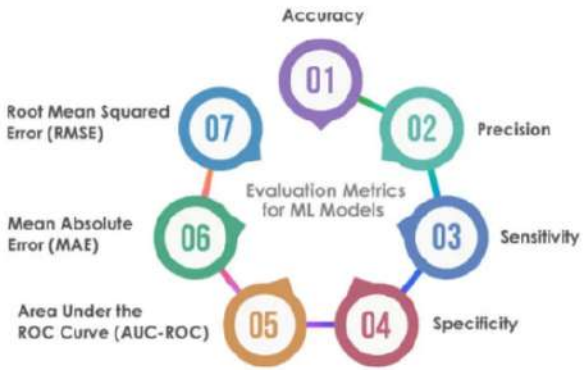
- **Description:** Training the model involves feeding the data into a machine learning algorithm to learn patterns.
- **Objective:** Develop a model that can generalise well to new, unseen data.
- **Methods:** Techniques include supervised learning, unsupervised learning, and reinforcement learning.

First Train, then Evaluate



Test Model

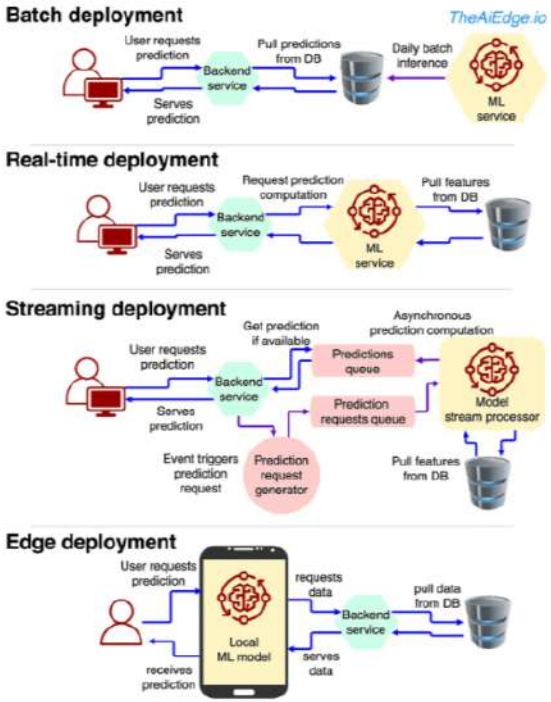
- **Description:** Testing the model involves evaluating its performance on a separate test dataset.
- **Objective:** Ensure the model's accuracy and ability to generalize to new data.
- **Methods:** Use metrics like precision, recall, F1-score, and confusion matrix.



Deployment

- **Description:** Deployment is the process of integrating the model into a production environment.
- **Objective:** Make the model available for practical use in real-world applications.
- **Methods:** Techniques include using APIs, cloud services, and continuous integration/continuous deployment (CI/CD) pipelines.

Forms of Deployment



An example End to End ML Project