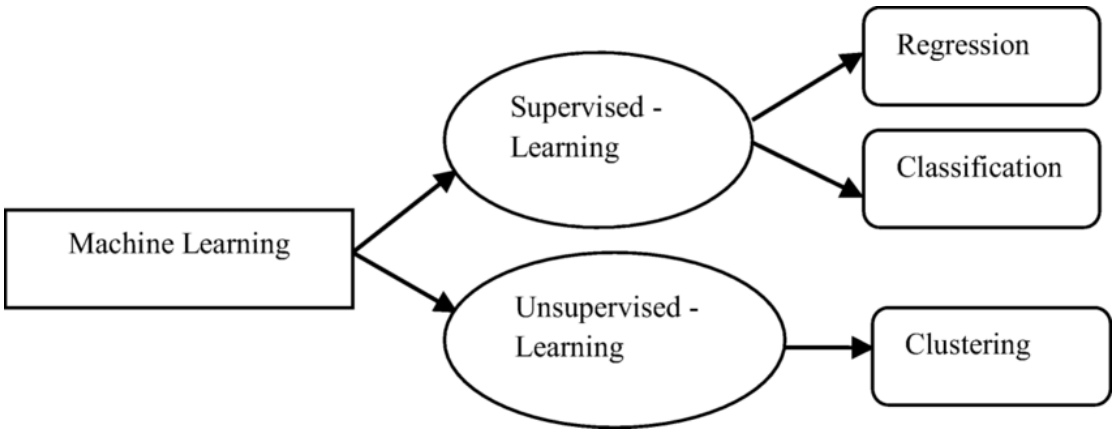
2. **Data Preprocessing**: In this stage, the collected data is cleaned and transformed to ensure it is suitable for analysis. This may involve handling missing values, normalizing data, and splitting the dataset into training and testing sets.
3. **Feature Extraction**: This step involves selecting important features or attributes from the data that will be used for model training. Techniques like encoding and dimensionality reduction may be applied here.
4. **Model Selection**: Based on the problem at hand (e.g., classification, regression), an appropriate machine learning algorithm is chosen. Common algorithms include decision trees, support vector machines, and neural networks.
5. **Model Training**: The selected model is trained using the preprocessed data. During this phase, the algorithm learns patterns and relationships in the data by adjusting its parameters.
6. **Model Evaluation**: After training, the model's performance is assessed using a separate test dataset. Various metrics such as accuracy, precision, and recall are used to evaluate its effectiveness.
7. **Model Deployment**: Once satisfied with the model's performance, it is deployed in a production environment where it can make predictions on new data.

Classification of Machine Learning:



Supervised learning:

Supervised learning is a type of machine learning where a model is trained on labeled data. This means that each training example consists of an input and the corresponding correct output (label). The goal is for the model to learn the relationship between the inputs and outputs so it can make accurate predictions on new, unseen data.

Key Features of Supervised Learning:

- **Labeled Data**: The training dataset contains input-output pairs. For example, in a spam detection system, the inputs might be email features (like subject line and sender), and the outputs are labels indicating whether the email is spam or not.
- **Training Process**: The model is trained using algorithms that adjust its parameters to minimize the difference between its predictions and the actual labels. This process is often done using techniques like gradient descent.

Types of Problems:

- **Classification**: The output is a category or class. For instance, predicting whether an email is spam or not (binary classification) or recognizing handwritten digits (multi-class classification).
- **Regression**: The output is a continuous value. For example, predicting house prices based on features like location, size, and number of bedrooms.
- **Evaluation**: After training, the model's performance is assessed using a separate test dataset. Common metrics include accuracy, precision, recall, and mean squared error, depending on whether the task is classification or regression.

Applications:

Supervised learning is widely used in various applications, including:

- Email filtering
- Credit scoring
- Medical diagnosis
- Image and speech recognition

Overall, supervised learning is a powerful approach that enables machines to learn from past data and make informed predictions.

Unsupervised learning:

Unsupervised learning is a type of machine learning where a model is trained on data that does not have labeled outputs. Instead of learning from specific input-output pairs, the model identifies patterns and structures in the data on its own. The goal is to uncover hidden patterns or groupings within the data.

Key Features of Unsupervised Learning:

- **Unlabeled Data**: The training dataset consists solely of input data without corresponding labels or outcomes. For example, a dataset of customer purchase histories without any classification or category labels.
- **Training Process**: The model explores the data to find natural groupings or relationships. It does not require predefined categories, which allows for more flexibility in data analysis.

Types of Problems: