



COURSE 1

Machine Learning for Data Science

UNIT 1	
INTRODUCTION TO MACHINE LEARNING	1
1. Machine Learning	1
2. Types of Machine Learning	1
2.1. Supervised	1
2.1.1. Types of Supervised	2
2.1.1.1. Classification	2
2.1.1.2. Regression	2
2.1.2. Supervised Learning Algorithms	2
2.2. Unsupervised/Clustering	2
2.2.1. Unsupervised Learning Algorithms	3
2.3. Semisupervised	4
2.4. Reinforcement	4
2.4.1. Reinforcement Learning Algorithms	4
UNIT 2	
ML MODE BUILDING TO DEPLOYMENT STEPS A-Z	5
1. Define the Problem	5
2. Data Collection (Primary) and Data Gathering (Secondary—online available)	5
3. Data Preprocess	5
4. Choose a Model	5
5. Split/Splitting the Data	5
6. Model Evaluation	6
7. Hyperparameter Tuning	7
8. Cross Validation	7
9. Model Finalization	8
10. Deploy the Model	8
11. Retest, Update, Version Controls and Monitor the Model (MLOps)	9
UNIT 3	

ALGORITHMS, TRAINING & TESTING DATA FEATURES MODELS, IMPORTANT	10
PYTHON LIBRARIES FOR MACHINE LEARNING AND INSTALLATION STEPS	
1. Algorithms	10
2. Training Data, Testing Data, Features and Model	10
3. Overfitting vs Underfitting	11
4. Important Python libraries for Machine Learning	11
5. Installation Steps for Machine Learning Environment	12
UNIT 4	
DATA PRE-PROCESSING BEFORE ML MODEL	
A. Data	14
B. Data Preprocessing	14
C. Data Preprocessing Key Steps and Techniques in Machine Learning	15
1. Data Cleaning	15
2. Data Integration	16
3. Data Transformation	16
4. Data Reduction	18
5. Data Discretization	19
UNIT 5	
PREPROCESSING DATA CLEANING → MISSING VALUES	
1. Data Cleaning	20
1.1. Exploring/Identifying Missing Values in the Dataset	21
1.2. Other Names of Missing Values	23
1.3. How to Identify Missing Values?	24
1.4. Why Is Handling Missing Values So Important?	24
1.5. Deciding Whether to Remove a Column with Missing Values	25
1.6. Detailed Methods for Handling Missing Values	26
1.7. How to Perform Data Cleaning	27
1.8. Dealing/Impute Missing Values Using Pandas Library	27
1.9. Imputer age column missing values using forward fill method using pandas	31
1.10. Imputer age column missing values using backward fill method using pandas	32
1.11. Imputation of Missing Values Using scikit-learn Library	33

1.12. Conclusion	39
PREPROCESSING DATA CLEANING → DATA INCONSISTANCIES/ANOMALIES	40
1. Data Inconsistency	40
2. Data Inconsistency by Data Type / Format	40
PREPROCESSING DATA CLEANING → D	50
1. ---	50

INTRODUCTION TO MACHINE LEARNING

1. Machine Learning

- Machine Learning allows machine to learn and make decisions smartly.
- In Machine Learning, machines can learn from the data provided or their own experience. It depends upon the type of machine learning.



Figure 1.1 Machine Learning Work.

- $$\begin{aligned} 2 \times 1 &= 2 \\ 2 \times 2 &= 4 \\ 2 \times 3 &= 6 \\ 2 \times 4 &= ? \quad \text{i.e., data patterns} \end{aligned}$$
- Machine learning analyzes patterns in data and predicts unknown values based on learned patterns.
 - $2x + 2 = ?$ (i.e., called ML Model)
 - Machine learning is completed in four steps:

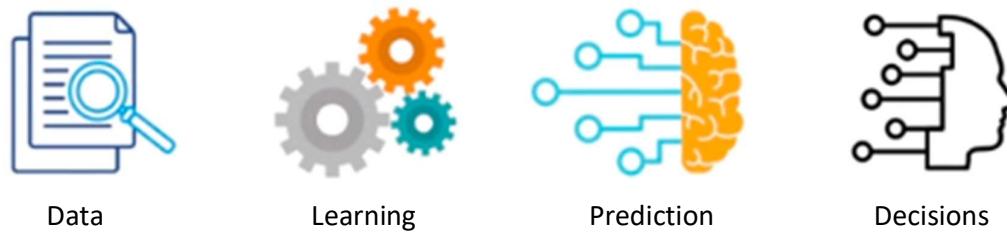


Figure 1.2 Machine Learning Steps.

2. Types of Machine Learning

2.1. Supervised

- Work under supervision (here, supervisor means labeled data)
- Teacher teaches

- Labeled data means the machine or algorithm learns from labelled data with a predefined output or target variable.
- Prediction
- Outcome

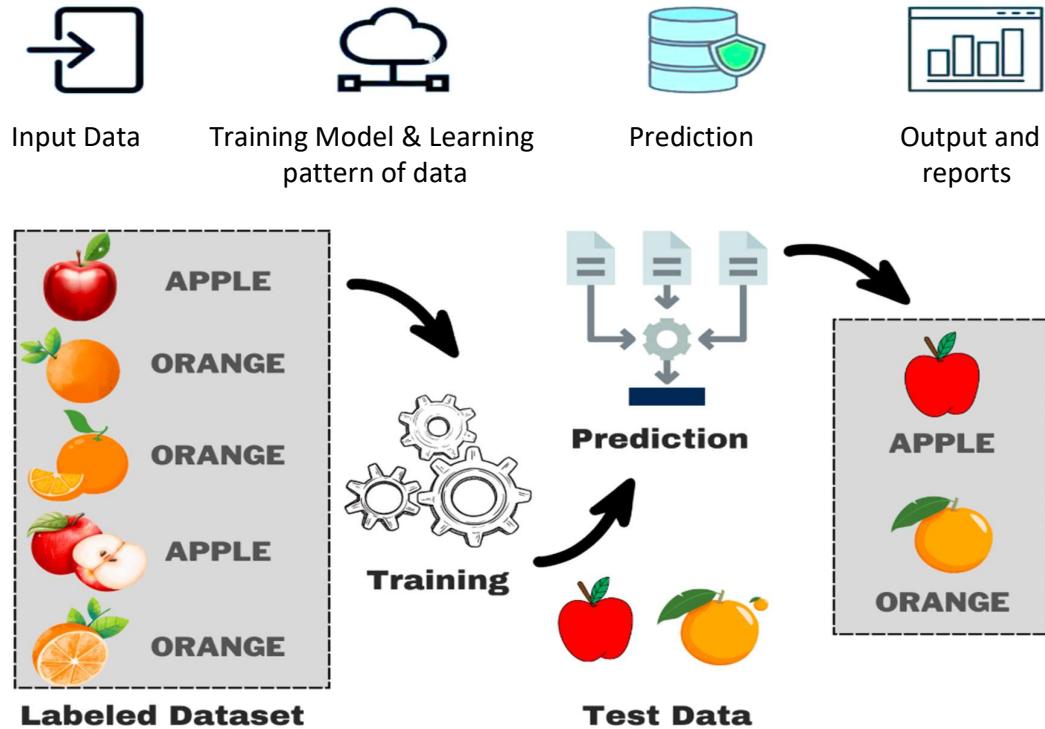


Figure 2.1.1 Supervised Learning.

2.1.1. Types of Supervised

2.1.1.1 Classification

- If the unknown value/quantity to be predicted is a **categorical variable**, the process is called classification.

2.1.1.2 Regression

- If the unknown value/quantity to be predicted is a **numeric variable**, the process is called regression.

2.1.2. Supervised Learning Algorithms

1. Logistic Regression
2. K-Nearest Neighbors (K-NN)
3. Support Vector Machine (SVM)
4. Kernel SVM

5. Naive Bayes
6. Decision Tree Classification
7. Random Forest Classification
8. Many more...

2.2. Unsupervised

- No supervision
- No teacher
- Self-learning
- Find pattern by itself
- No labelling of data means machine/algorith learn from unlabeled data without any predefine output or target variable.

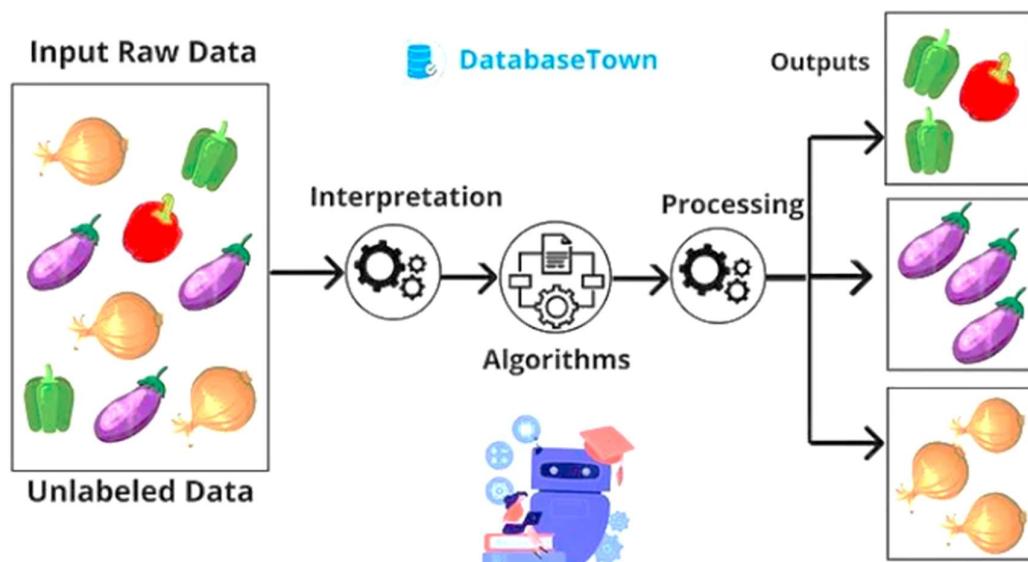
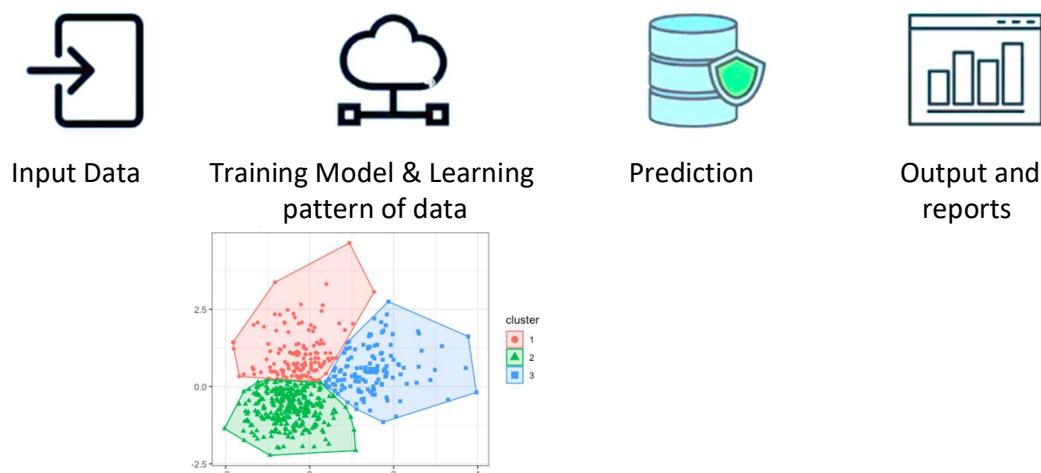


Figure 2.2.1 Unsupervised Learning.

2.2.1. Unsupervised Learning Algorithms

1. K-Means Clustering
2. Hierarchical Clustering
3. Probabilistic Clustering (key keywords of Probabilistic Clustering are Probability, likelihood, mixture model, latent variables, soft clustering, membership probability, expectation–maximization (EM), Gaussian Mixture Model (GMM), density estimation, uncertainty.)
4. Many more...

2.3. Semisupervised

- Mixture of supervised and unsupervised.
- Some data is labelled and most isn't.



Figure 2.3.1 Semisupervised Learning.

2.4. Reinforcement

- Hit and Trail learning
- Algorithms learn from mistakes
- Reward and Punishment rule (Drive rules and make policies)
- Prediction based on rewards and punishment (i.e., autonomous driving cars, reinforcement learning mostly used in gaming industry, 3d animation, etc.)
- Depends on feedback



Figure 2.4.1 Reinforcement Learning.

2.1.3. Reinforcement Learning Algorithms

1. Model-Free Reinforcement Learning
 - Policy Optimization

- Q-Learning
2. Model-Based Reinforcement Learning
 - Learn the Model
 - Given the Model
 3. Many more...