

<https://www.udemy.com/course/artificial-intelligence-machine-learning-comprehensive-guide-2024/learn/lecture/40485716#overview>

## 15. Machine Learning Glossary – Essential Terminology Explained

=> what is data set, features, model.

Data set – collection data from examples from past.

Features – important data points like salary, culture, how from home the office etc. they column of data sets.

Model – Mathematical representation of solutions. From Past data you create formula.

What are steps for machine learning?

=> 1. Understand/define the problem.

2. Convert problem into data problem

3. Data collection

4. Data Preparation

5. Developing model (EDA – Exploratory Data Analysis)

6. Model Training – Identifying & adjusting parameter.

7. Model Deployment

8. Monitor the performance.

What is model ?

=> Mathematical representation of the solutions.

Data Set 70% for train & 30% data set for test.

## 1. what is Machine Learning ?

=> give intelligence to the machine, is done thr predection from past/history data.

=> 1. In basic terms, ML is the process of training a piece of software, called a model, to make useful predictions or generate content (like text, images, audio, or video) from data.

2. ML subdomain of computer science , It focuses on algorithms which help a computer to learn from data without explicit programming .

## Pillar of the Machine learning ?

1. Domain/Business expertise

2. Mathematical/Statistical

3. Computer Science programming .

4.

# Types of ML Systems(Models)

(types of approach)

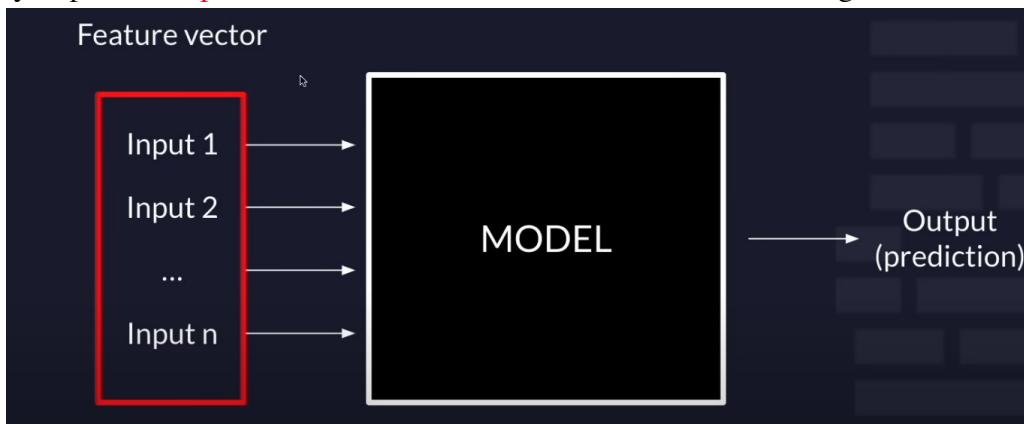
- Supervised learning – used labels inputs
- Unsupervised learning – unlabeled , form structure as pattern to learn.
- Reinforcement learning – no right or wrong , it iterative learning.
- Generative AI

## Supervised learning Model(inputs and correct answers)

=>It make predication after **seeing lots of data with correct answers**. After that it see connection between element which produce correct answers.  
It is paired with input and correct output .

Ex. supervised learning is like teaching a child with examples:

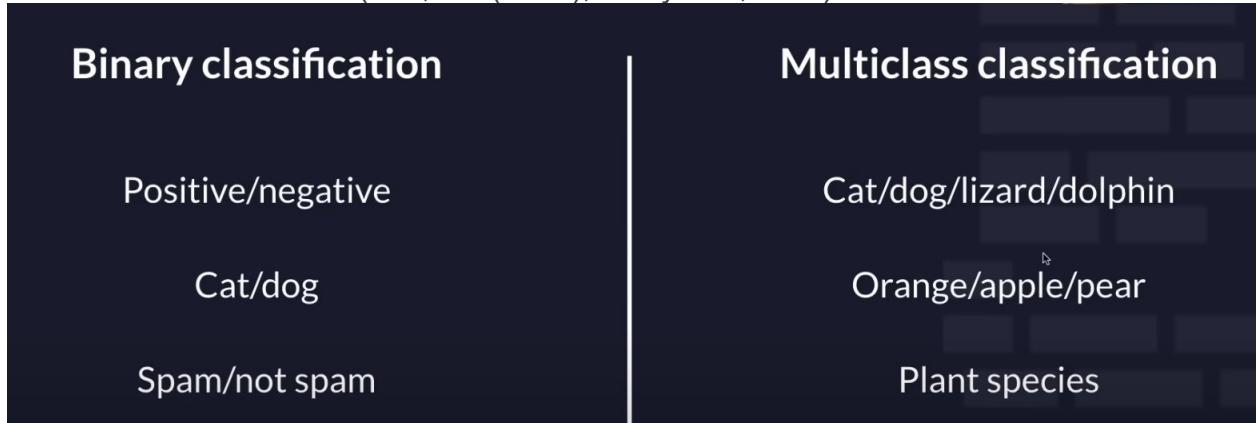
you provide **inputs and correct answers**, and the model learns to generalize.



## Types of Supervised Learning:

### 1. Classification

- Binary classification (spam vs. not spam).(rain or no rain), Two possibility Yes or No.resign or not resign.
- Multi-class classification (rain, **hail(Gara)**,heavy rain,snow)



Algorithm for Supervised learning :

- i) classification : Naive Bayes , Decision tree, Support vector Machine, Random forest,K Nearest Neighbor(KNN)

## 2. Regression

- Predicts a numerical value.(house price , stock market )
- It continue variable ex. Reached home (may take 30 min, 45 min, 60 min but finally u reached home)
- Algorithm for Regression: Linear regression, Neural network regression. Deep learning, Support vector, Decision tree regression. Lasso regression & Ridge regression(better result than normal regression)

## Unsupervised

1. It work on unlabeled data. So no label, no training data set.
2. So it find similarities to form a cluster.
3. We solve two type of Problems i) clustering ii) association(recommend )

### What is Clustering ?

=> finding similar data & group together to form as logical cluster

Ex, Anomaly (fraud) detection

Diff stages

1. Unsupervised
2. Convert cluster to class
3. Convert supervised learning

Two of unsupervised

1. Clustering : K mean clustering, Hierarchical clustering , DBScan, Mean shift clustering, Agglomerative hierarchical clustering,Gaussian Mixture
2. Associate : A Priori,

## Reinforcement learning

Algorithm for Reinforcement : Q Learning, R Learning, TD Learning,

**Q 1. If you wanted to use an ML model to predict energy usage for commercial buildings, what type of model would you use?**

Energy usage is measured in kilowatthours (kWh), which is a number, so you'd want to use a regression model.

•

What is bias?

=>

## Step By step Recipe for Machine Learning ?

=> 1. Do you need Machine Learning or not.

Complexity is very high,problem scale/volume large,unstructure data.

2. Can you formulated/explain problem clearly.

    What i/p provide to get o/p that clarity

3. Data Collection

    Do you have sufficient data for analysis

4. Do you have Regular Pattern in data

**What is Deep Learning ?**

=> Deep learning involved creating artificial neurons called as perceptron. inspire by nervous system. Here different network layered created, each layer take information & process it , transfer to next layer.

There are diff framework

1. Ann (artificial neural network)
2. CNN (Convolution neural network) – images processing
3. RNN (recurrent neural network) – text proceesing

Different types of data?

1. Structure – any database in rows-column, excel sheet.
2. Unstructure –text, images ,video
3. Semi structure – like JSON, XML

**Deep Learning give better result than Machine Learning.**

**Drawback of Deep Learning**

- => 1. Its blackbox , inside algorithms not explainable  
2. Required huge data.

**Natural Languages processing (NLP)**

-> Linguistic +Deep Learning+ programming(Python)

**What is generative AI?**

=> Generative AI open-ended AI model which help create new & original content. which help to genarate new text, images, video, music.

ChatGpt- create a text

Dali2 – create image

Google Music LM – create Music.

**What is prompt Engineering ?**

=> A simple text OR question asked to AI model.

**Types of AI?**

- => 1. Reactive AI – Present data from current situation.  
2. Limited Memory AI – improve decision based on past data ex. Self Driven car  
3. Theory of Mind AI - emotions  
4. Self aware AI – Self decision making

**Q 2. What is the main advantage of using ensemble learning?**

**Q 3. What is the primary challenge of training deep neural networks?**

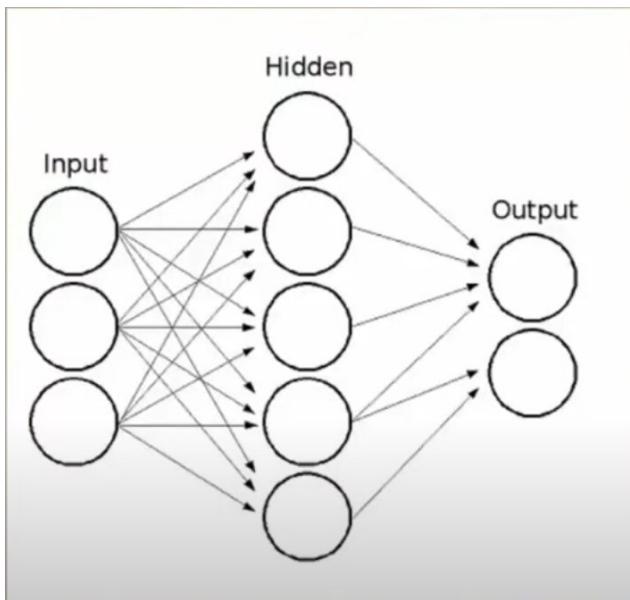
**Q 4. What is a limitation of using decision trees for classification tasks?**

**Q 5. What is a limitation of using the RBF kernel in support vector machines (SVMs)?**

**Q 6. What is the primary advantage of using deep learning models over traditional machine learning algorithms?**

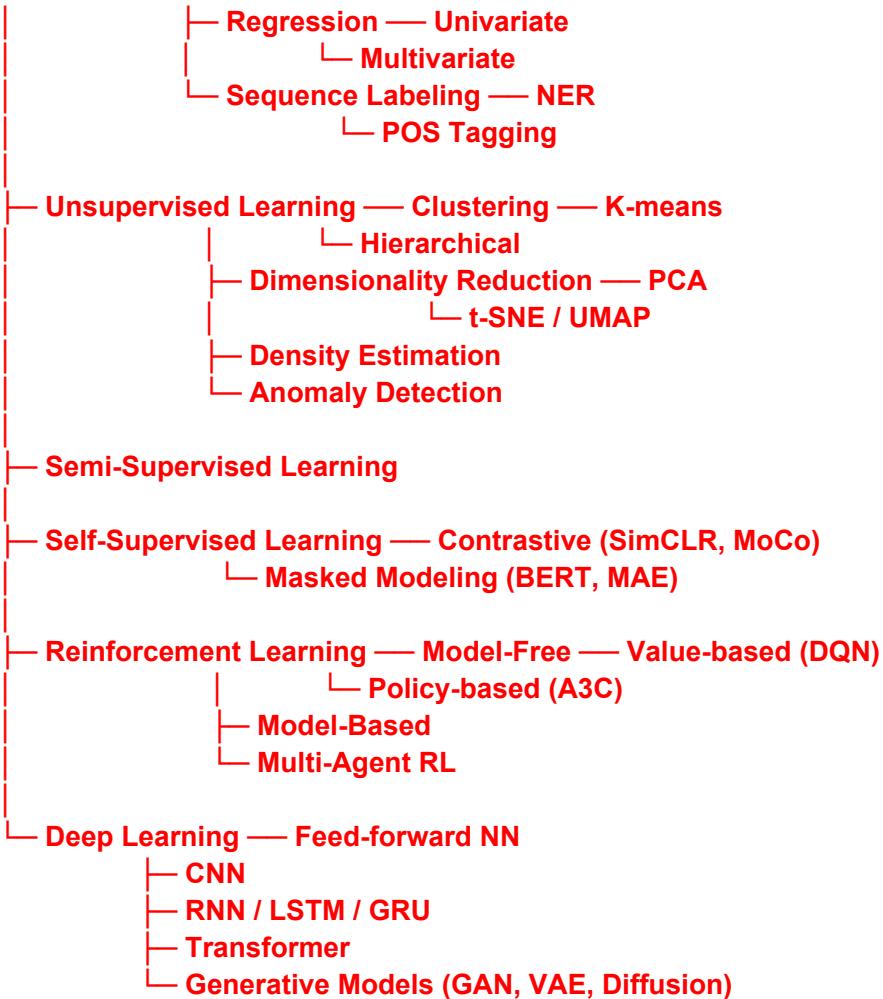
- Strong knowledge of C/C++, software design, programming techniques, and AI algorithms.
- Firsthand work experience with parallel programming, ideally CUDA C/C++.
- Strong communication and organization skills, with a logical approach to problem solving, good time management, and task prioritization skills.
- Some travel is required for conferences and for on-site visits with developers

Technology – c++, Python , AWS.

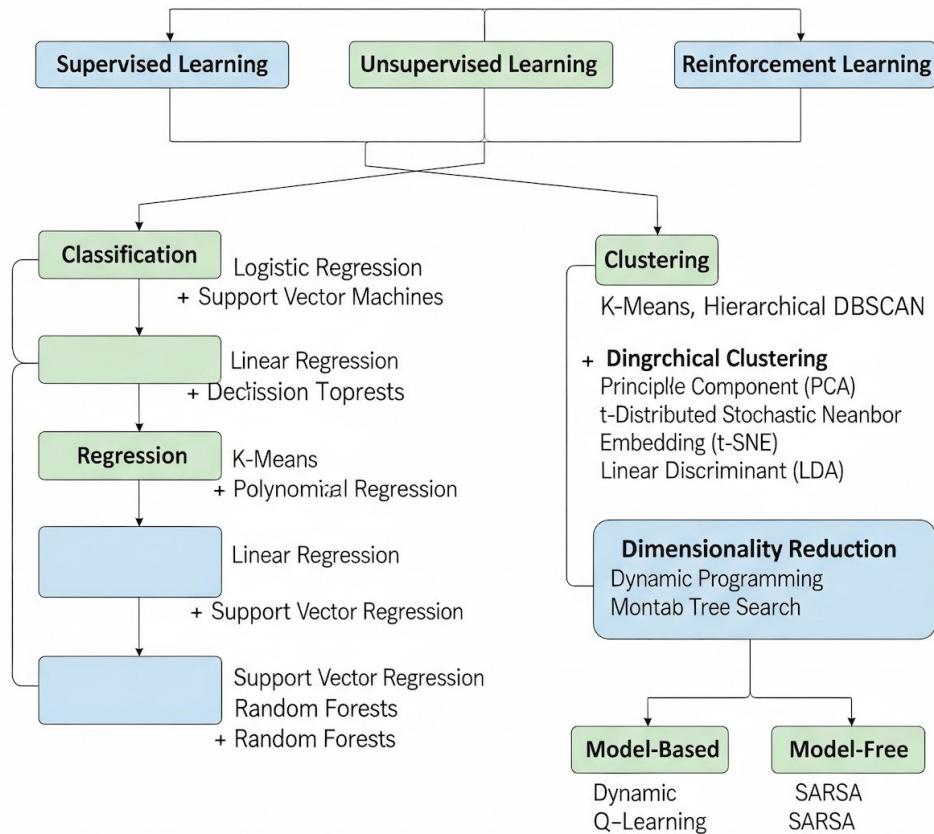


#### Types of Machine Learning

```
└── Supervised Learning └── Classification └── Binary  
    └── Multi-class
```



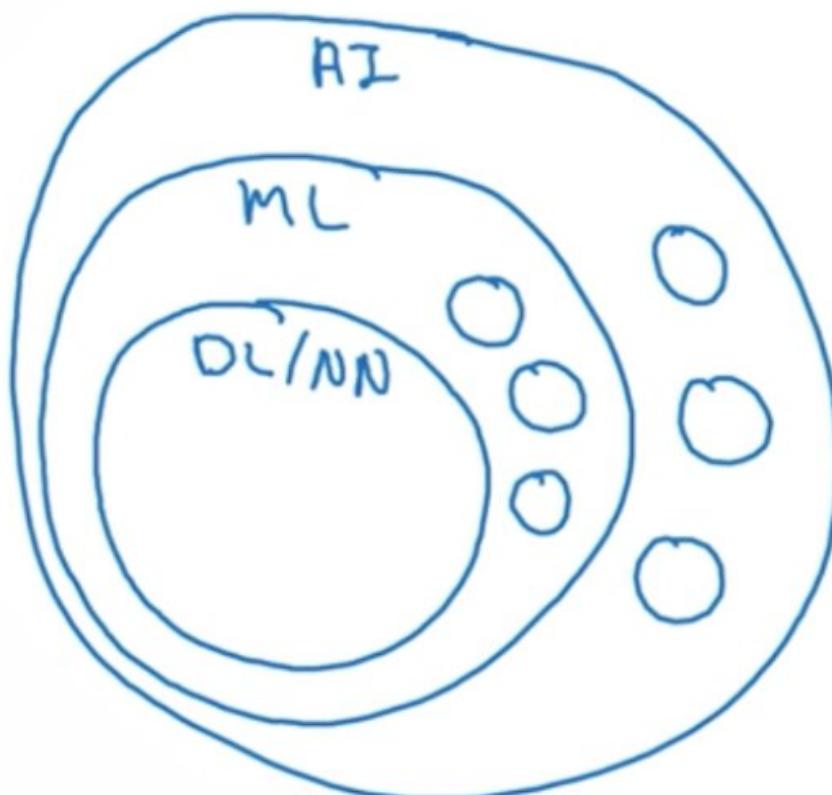
# Types of Machine Learning



## Types of Machine Learning

- 1. Supervised Learning
  - Classification
    - Binary (e.g., spam vs. not-spam)
    - Multi-class (e.g., digit recognition 0–9)
  - Regression
    - Univariate (predict one continuous value)
    - Multivariate (predict several continuous values)
  - Sequence Labeling
    - Named-entity recognition
    - Part-of-speech tagging
- 2. Unsupervised Learning
  - Clustering

- | | └ K-means
- | | └ Hierarchical
- | └ Dimensionality Reduction
  - | └ PCA (Principal Component Analysis)
  - | └ t-SNE / UMAP
- | └ Density Estimation
- | └ Anomaly / Outlier Detection
- └ 3. Semi-Supervised Learning
  - └ (Mix of labeled & unlabeled data)
- └ 4. Self-Supervised Learning
  - └ Contrastive methods (SimCLR, MoCo)
  - └ Masked-language / masked-image modeling (BERT, MAE)
- └ 5. Reinforcement Learning
  - └ Model-Free
    - | └ Value-based (Q-learning, DQN)
    - | └ Policy-based (REINFORCE, A3C)
  - └ Model-Based
  - └ Multi-Agent RL
- └ 6. Deep Learning (can overlap with any branch above)
  - └ Feed-forward neural networks
  - └ Convolutional Neural Networks (CNNs)
  - └ Recurrent Neural Networks (RNNs, LSTMs, GRUs)
  - └ Transformers
  - └ Generative Models (GANs, VAEs, Diffusion)



Copy  
Share

## Top AWS Services A Data Engineer Should Know

AWS Service	Purpose	Key Use Case
Amazon S3	Scalable object storage for data lakes and backups	Storing raw CSV files in a data lake for further processing
AWS Glue	Managed ETL service for data preparation and cataloging	Automating ETL jobs to transform raw data in S3 into structured format for Redshift
Amazon Redshift	Managed data warehouse for large-scale analytics	Running complex SQL queries for business intelligence reporting
Amazon Kinesis	Real-time data streaming and	Processing live user activity data from a

Kinesis	processing	website for real-time analytics
AWS Lambda	Serverless compute for event-driven processing	Triggering data transformation scripts on S3 file uploads
Amazon Athena	Serverless query service for analyzing S3 data with SQL	Running ad hoc SQL queries to validate data in a data lake
Amazon EMR	Managed big data processing with Hadoop, Spark, etc.	Running Spark jobs to process large datasets for machine learning
Amazon RDS	Managed relational database for structured data	Storing transactional data from an application for pipeline integration
Amazon DynamoDB	Managed NoSQL database for unstructured/semi-structured data	Storing real-time user interaction data for a mobile app
AWS Step Functions	Orchestrates complex workflows and dependencies	Orchestrating an ETL pipeline involving S3, Glue, and Redshift

□ **ML & Deep Learning Focused** 1 ml-interviews-book → Chip Huyen's ML interviews playbook – frameworks, signals, and practical advice. □ <https://lnkd.in/dsHu4256>

□ 2 deep-learning-interview-questions → 80+ solved DL Qs – loss functions, architectures, overfitting, tricks from real interviews. □ <https://lnkd.in/djgcMeFk>

□ 3 machine-learning-interview → Covers applied ML, real company questions, model design, trade-offs, and theory. □ <https://lnkd.in/ddpCihcN>

□ 4 data-science-interviews → 250+ questions: Python, SQL, stats, product ML, A/B testing – community-verified. □ <https://lnkd.in/dt4-cSQp>

□ 5 MLQuestions → A compact list of conceptual ML questions - bias/variance, metrics, model evaluation. □ <https://lnkd.in/d64WTkgF>

#### □ **GenAI, NLP & Transformers**

□ 6 awesome-generative-ai-guide → GenAI prep: frameworks, LangChain, LLM-based projects, multimodal updates. □ <https://lnkd.in/dtFTZsCs>

□ 7 stanford-cme-295-transformers-large-language-models → Illustrated cheat sheets on how transformers and LLMs work under the hood. □ <https://lnkd.in/dvxJPky7>

8 pytorch-interview-questions→ Everything PyTorch: autograd, model design, performance tuning, debugging.□ <https://lnkd.in/deJawrpQ>

□ **Core Coding Repos**

9 TheAlgorithms/Python→ Python implementations of all core algorithms & data structures – ideal for technical rounds.□ <https://lnkd.in/dZZVXsav>□

10 Tech-Interview-Handbook→ Covers coding rounds, behavioral questions, resume prep – all in one place.□ <https://lnkd.in/dFMBpf2v>

□ **System Design**

11 system-design-primer→ Design scalable systems, trade-offs, caching, queues, everything they ask in system design rounds.□ <https://lnkd.in/dEVmXwb5>

12 complete-system-design→ End-to-end prep for system design interviews with diagrams, FAQs, and practice prompts.□ <https://lnkd.in/dqqpHXRW>

Also:

Don't just prepare to answer interview questions. Use these to build mini projects and ship something useful.→ It'll stick better→ You'll learn deeper→ You'll stand out

I didn't know half of these till recently Shoutout to [Sairam Sundaresan](#) for curating so many gems  
□-- Have you checked out "The MOTHER of AI" Project? We released the first blog yesterday -  
<https://lnkd.in/dmEwJZtQ>

### Quick Tip:

- If you're in a hurry, focus on **ML first** (since AI is theoretical and DL is advanced ML).
- Hands-on coding (Python + scikit-learn) will help the most!