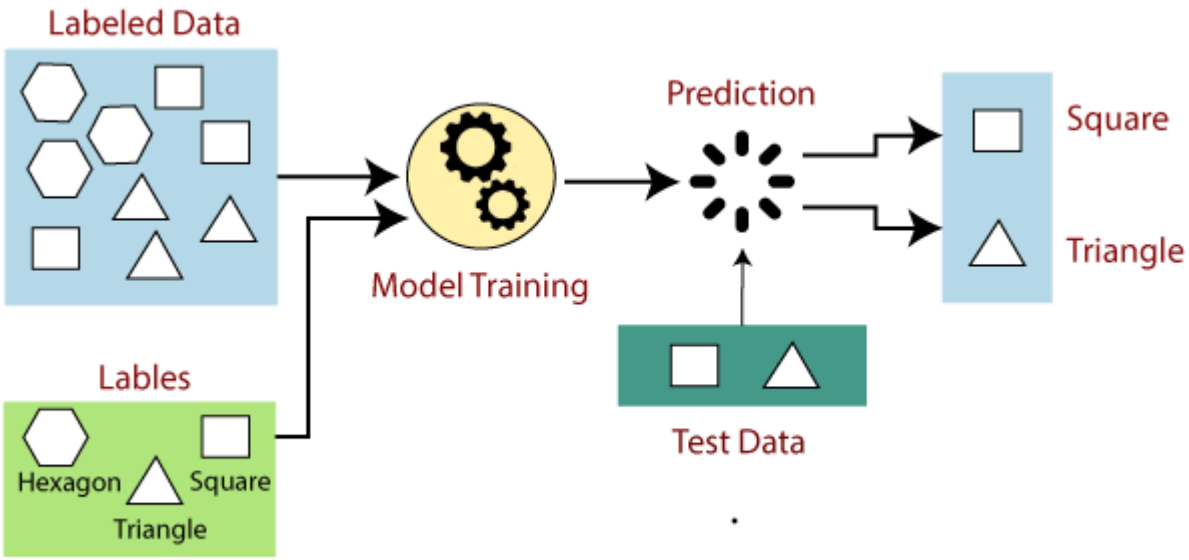
# CSE 3207 – Artificial Intelligence

**Artificial Intelligence**: Artificial Intelligence (AI) refers to systems or machines that mimic human intelligence to perform tasks and can iteratively improve themselves based on the information they collect.

**Machine Learning**: Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.

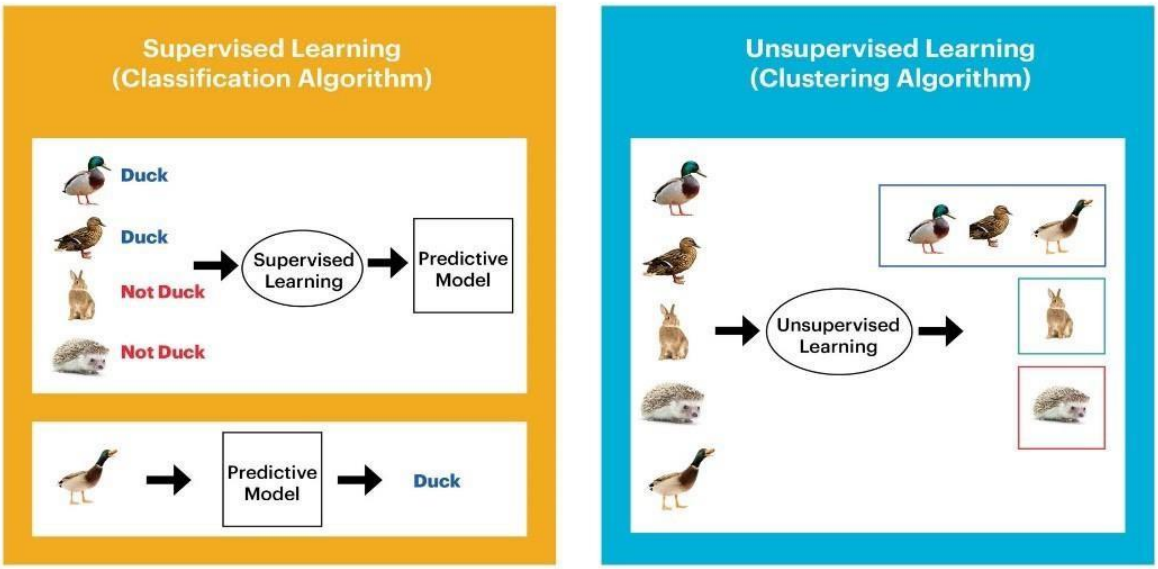
**How machine learning works?**




Machine Learning algorithm is trained using a training dataset to create a model. When new input data is introduced to the ML algorithm, it makes a prediction on the basis of the model. The prediction is evaluated for accuracy and if the accuracy is acceptable, the Machine Learning algorithm is deployed. If the accuracy is not acceptable, the Machine Learning algorithm is trained again and again with an augmented training data set. This is just a very high-level example as there are many factors and other steps involved.

**Types of Machine Learning**:
There are mainly three sub-categories of ML. They are

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning




**Supervised Learning**: In supervised learning, the machine is taught by example. The operator provides the machine learning algorithm with a known dataset that includes desired inputs and outputs, and the algorithm must find a method to determine how to arrive at those inputs and outputs.

Types of Supervised Learning:

- Classification
- Regression

The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age, etc. and Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.

**Unsupervised Learning**: Unsupervised learning works by analyzing the data without its labels for the hidden structures within it, and through determining the correlations, and for features that actually correlate two data items. It can be used for clustering.

**Reinforcement Learning**: Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error. Some of the autonomous driving tasks where reinforcement learning could be applied include trajectory optimization, motion planning, dynamic pathing, controller optimization, and scenario-based learning policies for highways. For example, parking can be achieved by learning automatic parking policies.

**Machine Learning Life cycle**

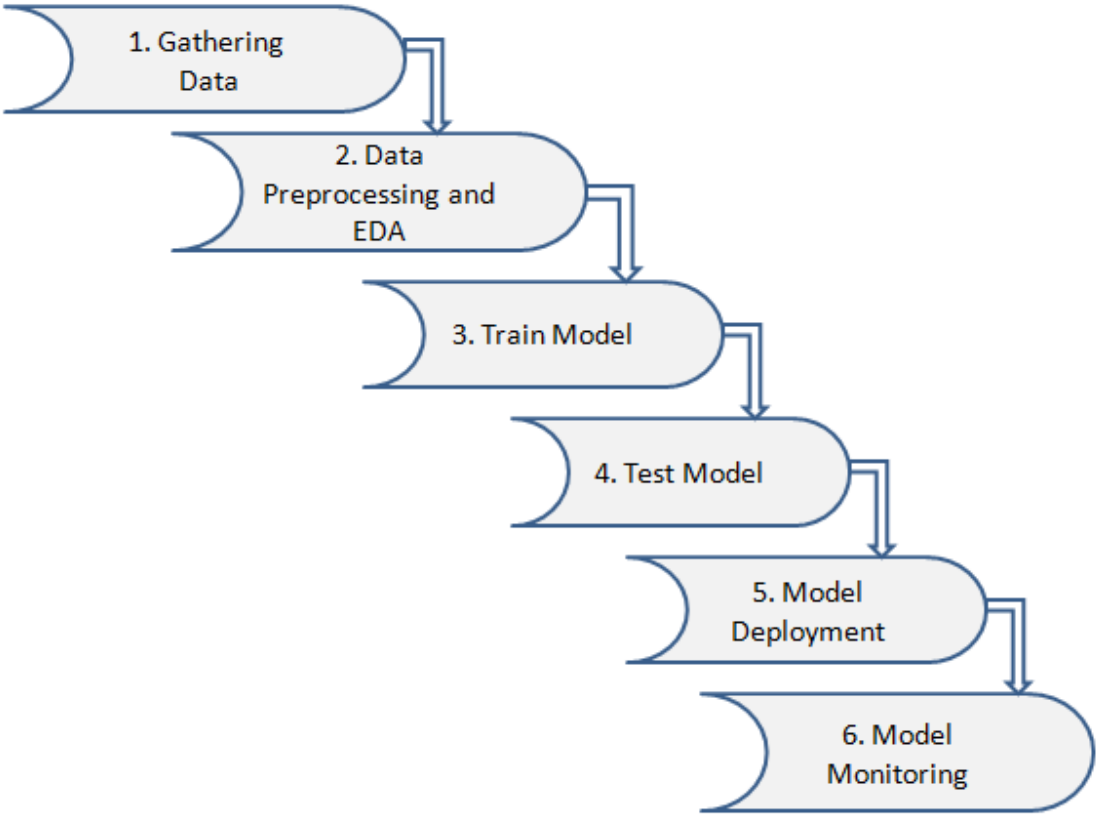



Figure: Machine Learning Life-cycle

**Datasets for Supervised and Unsupervised Learning:**

- Dataset for Regression (Supervised Learning)

| Features | | | | Target Label |
|---|---|---|---|---|
| RAM | Storage | Weight | Display | Price |
| 4 GB | 128 GB | 1.377 kg | 14 Inch | 25490 Tk |
| 8 GB | 256 GB | 1.39 kg | 14.1 Inch | 25700 Tk |
| 6 GB | 500 GB | 1.45 kg | 15.6 Inch | 26000 Tk |
| 16 GB | 1 TB | 1.8 kg | 15.6 Inch | 93000 Tk |
| 4 GB | 256 GB | 1.37 kg | 14 Inch | 29100 Tk |
| 16 GB | 512 GB | 1.4 kg | 14 Inch | 94000 Tk |
| 4 GB | 1 TB | 1.8 kg | 15.6 Inch | 29500 Tk |
| 16 GB | 512 GB | 1.61 kg | 14 Inch | 105500 Tk |

| | | | | |
|---|---|---|---|---|
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

- Dataset for Classification (Supervised Learning)

| Features | | | | Target Label |
|---|---|---|---|---|
| RAM | Storage | Weight | Display | Price Category |
| 4 GB | 128 GB | 1.377 kg | 14 Inch | Cheap |
| 8 GB | 256 GB | 1.39 kg | 14.1 Inch | Cheap |
| 6 GB | 500 GB | 1.45 kg | 15.6 Inch | Cheap |
| 16 GB | 1 TB | 1.8 kg | 15.6 Inch | Costly |
| 4 GB | 256 GB | 1.37 kg | 14 Inch | Cheap |
| 16 GB | 512 GB | 1.4 kg | 14 Inch | Costly |
| 4 GB | 1 TB | 1.8 kg | 15.6 Inch | Cheap |
| 16 GB | 512 GB | 1.61 kg | 14 Inch | Costly |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |

- Dataset for clustering (Unsupervised Learning)

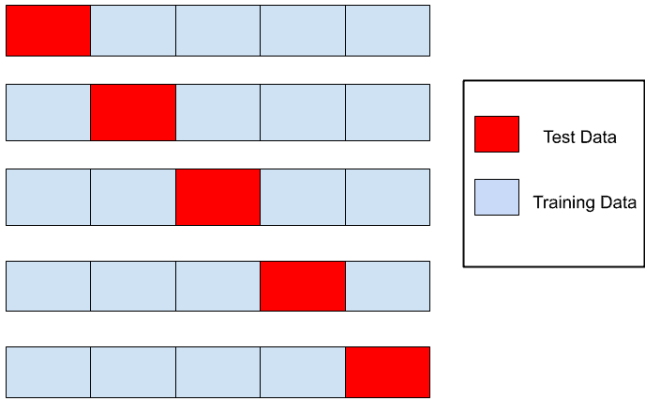
| Features | | | |
|---|---|---|---|
| RAM | Storage | Weight | Display |
| 4 GB | 128 GB | 1.377 kg | 14 Inch |
| 8 GB | 256 GB | 1.39 kg | 14.1 Inch |
| 6 GB | 500 GB | 1.45 kg | 15.6 Inch |
| 16 GB | 1 TB | 1.8 kg | 15.6 Inch |
| 4 GB | 256 GB | 1.37 kg | 14 Inch |
| 16 GB | 512 GB | 1.4 kg | 14 Inch |

| | | | |
|---|---|---|---|
| 4 GB | 1 TB | 1.8 kg | 15.6 Inch |
| 16 GB | 512 GB | 1.61 kg | 14 Inch |
| . . . . | . . . . | . . . . | . . . . |

**Train/Test**: Train/Test is a method to measure the accuracy of your model. It is called Train/Test because the data set is split into two sets: a training set and a testing set.

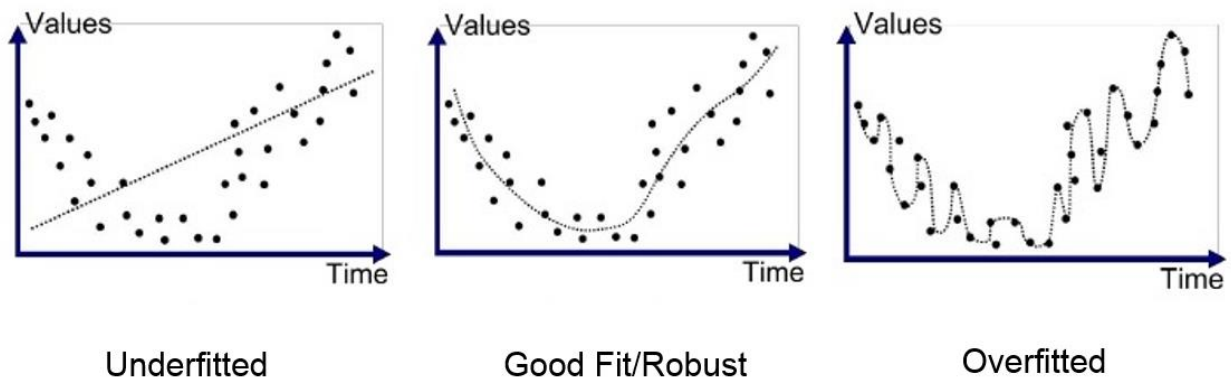
*train the model using the training set; test the model using the testing set*

**K Fold Cross Validation:**




In *k*-fold cross-validation, the original sample is randomly partitioned into *k* equal sized subsamples. Of the *k* subsamples, a single subsample is retained as the validation data for testing the model, and the remaining *k* − 1 subsamples are used as training data. The crossvalidation process is then repeated *k* times, with each of the *k* subsamples used exactly once as the validation data. The *k* results can then be averaged to produce a single estimation.

## Overfitting and Underfitting




**Overfitting:** In Machine Learning, overfitting is the case where the predictor model fits perfectly on the training examples, but does badly on the test examples. This often happens when the model is too complex and trivially fits the data (too many parameters), or when there is not enough data to accurately estimate the parameters. That is, when the ratio of model complexity to training set size is too high, overfitting will typically occur.

*Overfitting is like learning through memorization. Rather than understanding the concepts at play and making generalizations you simply recall what you've seen before and find the closest thing to what you already know. This means that while your model will perform very well on the training set by memorizing what it should do with each input, when it is faced with an input it has never seen before it won't have any general concepts to fall back on.*

**Underfitting:** Underfitting refers to a model that can neither model the training data nor generalize to new data. An underfit machine learning model is not a suitable model and will be obvious as it will have poor performance on the training data. Underfitting is often not discussed as it is easy to detect given a good performance metric. The remedy is to move on and try alternate machine learning algorithms.

### Some Real-World Applications of Machine Learning

1. Automatic Language Translation in Google Translate
2. Faster route selection in Google Map
3. Driverless/Self-driving car
4. Smartphone with face recognition
5. Speech Recognition
6. Ads Recommendation System

7. Netflix Recommendation System
8. Auto friend tagging suggestion in Facebook
9. Stock market trading
10. Fraud Detection
11. Weather Prediction
12. Medical Diagnosis
13. Chatbot
14. Machine Learning in Agriculture

**Benefits of machine learning**

1. Work Automation
2. Powerful predictive Ability
3. Increased in sales in the e-commerce market
4. ML benefits in the medical domain for enhancing medical diagnosis, drug development
5. Machine Learning is used in robotic medical surgery
6. ML in finance increases productivity enhances revenue and gives secure transactions
7. Modeling the data to make useful decisions