Subject/Course

MACHINE LEARNING

By Prof. Shraddha Dalvi

Syllabus

Syllabus structure :CSC 604

- 06 Modules(Topics)
- 03 lectures/week
- 02 hours practical/week
- Course outline-

Course Outlines

ML

Introduction to Machine Learning

Unit No.6
Unit No.5
Unit No.4
Unit No.3
Unit No.2

Introduction to Machine Learning

- Introduction to Machine Learning
- Issues in Machine Learning
- Application of Machine Learning
- Steps in developing a Machine Learning Application
- Supervised & Unsupervised Learning
- Concepts of Classification
- Clustering and prediction
- Training
- Testing and validation dataset
- Cross validation
- Overfitting and underfitting of model

Unit No.1

Unit No.6
Unit No.5
Unit No.4
Unit No.3
Unit No.2

Introduction to Machine Learning

- Performance Measures
- Measuring Quality of model
- Confusion Matrix
- Accuracy
- Recall
- Precision
- Specificity
- F1 Score
- RMSE

Unit No.1

Mathematical Foundation for ML

Unit No.6
Unit No.5
Unit No.4
Unit No.3

- System of Linear equations
- Norms
- Inner products

- Length of Vector
- Distance between vectors
- Orthogonal vectors

Unit No.2
Unit No.1

Mathematical Foundation for ML

Unit No.6
Unit No.5
Unit No.4
Unit No.3

- Symmetric Positive Definite Matrices
- Determinant
- Trace

- Eigenvalues and vectors
- Orthogonal Projections
- Diagonalization
- SVD and its applications

Unit No.2
Unit No.1

Linear Models

Unit No.6
Unit No.5
Unit No.4

- The least-squares method
- Multivariate Linear Regression
- Regularized Regression
- Using Least-Squares Regression for classification
- Support Vector Machines

Unit No.3
Unit No.2
Unit No.1

Clustering

Unit No.6

Unit No.5

- Hebbian Learning rule
- Expectation - Maximization algorithm for clustering

Unit No.4

Unit No.3

Unit No.2

Unit No.1

Classification models

Unit No.6

- Introduction, Fundamental concept
- Evolution of Neural Networks
- Biological Neuron
- Artificial Neural Networks
- NN architecture
- McCulloch-Pitts Model

Unit No.5
Unit No.4
Unit No.3
Unit No.2
Unit No.1

Classification models

Unit No.6

- Designing a simple network
- Non-separable patterns
- Perceptron model with Bias
- Activation functions, Binary, Bipolar, continuous, Ramp
- Limitations of Perceptron

Unit No.5
Unit No.4
Unit No.3
Unit No.2
Unit No.1

Classification models

Unit No.6

- Perceptron Learning Rule
- Delta Learning Rule (LMS-Widrow Hoff)
- Multi-layer perceptron network
- Adjusting weights of hidden layers
- Error back propagation algorithm
- Logistic regression

Unit No.5
Unit No.4
Unit No.3
Unit No.2
Unit No.1

Dimensionality Reduction

- Curse of Dimensionality
- Principal Component Analysis
- Feature Selection and Feature Extraction
- Dimensionality Reduction Techniques

Subject Introduction

Module wise weightage for Examination scheme

- Module No.1 : Introduction to Machine Learning [15%]
- Module No.2 : Learning with Regression and Trees [13%]
- Module No.3 : Ensemble Learning [18%]
- Module No.4 : Learning with Classification [10%]
- Module No.5 : Learning with Clustering [25%]
- Module No.6 : Dimensionality Reduction [18%]

Subject Introduction

Examination scheme and T/W marking scheme

- Total Marks=150
- Internal assessment: Average Test Marks=20
  [Test:1(20 Marks)+Test:2(20 Marks)]/2
- End semester theory exam=80
- External Practical & Oral exam=25
- Term Work=25 [5(Attendance)+05(Assignment)+15(Lab Work/Performance)]

Subject Introduction

Teaching scheme and course credits

- Theory lectures: 3 Hours/week
- Practical: 2 Hours/week par batch

- No tutorial work
- Course Credits=4 (3 Theory+ 1 Practical)

Subject Introduction

End semester theory Examination

- Question paper will comprise a total of six questions
- All questions carries equal marks
- Questions will be mixed in nature (e.g. suppose Q2 has part(a) from module 3 then part(b) will be from any other module than 3

Subject Introduction

Books

Text Books:

1. Nathalie Japkowicz & Mohak Shah, —”Evaluating Learning Algorithms: A Classification Perspective”, Cambridge
2. Marc Peter Deisenroth, Aldo Faisal, Cheng Soon Ong, —”Mathematics for machine learning”
3. Samir Roy and Chakraborty, —Introduction to soft computing‖, Pearson Edition
4. Ethem Alpaydın, —Introduction to Machine Learning‖, MIT Press McGraw-Hill Higher Education
5. Peter Flach, —Machine Learning‖, Cambridge University Press

Subject Introduction

Books

Reference Books:

1. Tom M. Mitchell, ―Machine Learning‖, McGraw Hil
2. Kevin P. Murphy, ―Machine Learning ― "A Probabilistic Perspective", MIT Press
3. PearsonStephen Marsland, ―"Machine Learning an Algorithmic Perspective", CRC Press
4. Shai Shalev-Shwartz, Shai Ben-David, ―Understanding Machine Learning‖, Cambridge University Press
5. Peter Harrington, ―Machine Learning in Action‖, DreamTech Press

Subject Introduction

Tentative University Question Paper Pattern

- Question No.1 is a compulsory question. From Question 2 to Question 6 students can attempt any three question.
- Question No.1 should consists of four questions, each carry 5 marks.
- Question No.6 should include short notes type of questions
- Question No.6 may include 3 short note questions with choice. (any two from three)

Unit Name	Weightage	Q1	Q2	Q3	Q4	Q5	Q6	Module total
Module 1	10%	x					x	10
Module 2	25%	x		x	x		x	30
Module 3	15%	x		x			x	20
Module 4	20%	x			x	x	x	30
Module 5	20%	x	x			x	x	30
Module 6	15%	x	x				x	20
	100%	20 marks	20 marks	20 marks	20 marks	20 marks	20 marks	

Experiment List

Sr. No.	Name of the Experiment
1	Introduction to platforms such as Anaconda, COLAB
2	Study of Machine Learning Libraries and tools (Python library, tensorflow, keras,...)
	Implementation of following algorithms for a given example data set-
3	Linear Regression.
4	Logistic Regression.
5	Support Vector Machines
6	Hebbian Learning
7	Expectation -Maximization algorithm
8	McCulloch Pitts Model.
9	Single Layer Perceptron Learning algorithm
10	Error Backpropagation Perceptron Training Algorithm
11	Principal Component Analysis
12	Applications of above algorithms as a case study (E.g. Hand Writing Recognition using MNIST data set, classification using IRIS data set, etc)

Case study details

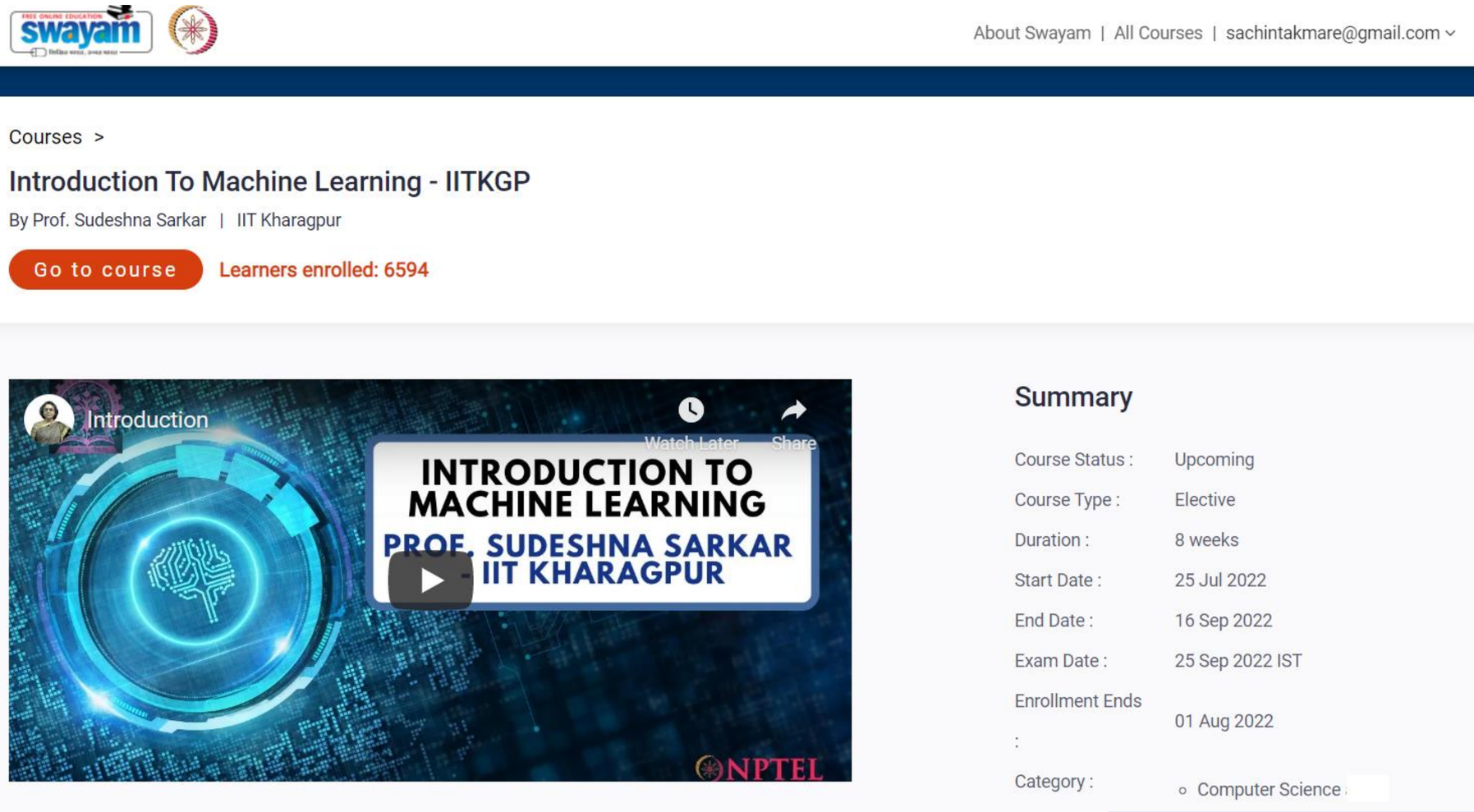
Topic list

1. Retail store sales prediction
2. Restaurant sales prediction
3. credit card fraud detection
4. Inventory prediction
5. Diabetes prediction
6. Caterpillar tube assembly pricing
7. Breast cancer prediction
8. Coal production estimation
9. Heart diseases prediction
10. whether prediction
11. players salary prediction
1. Uber
2. Spotify
3. Search Engine Optimization
4. Recommendations for Online Shopping
5. Translation
6. Netflix
7. Google Maps
8. Traffic Alerts
9. PayPal
10. Yelp’s Photo Classifier
11. Stock Market Signals Using Machine Learning
12. Supply chain optimization

By Shraddha Dalvi

NPTEL Course

Course url: https://onlinecourses.nptel.ac.in/noc22_cs97/preview



By Shraddha Dalvi

Pre-Requisite

Before proceeding with this course.

The basic knowledge of Computer Science is mandatory

To get started with Machine Learning you must be familiar with the following concepts

- Statistics
- Probability
- Linear Algebra
- Calculus
- Programming Languages (Python/R etc.)

By Shraddha Dalvi

Understanding Machine Learning with an Analogy

Task: As a human how will you choose the best mangoes?

- Let’s suppose one day you went for shopping mangoes. The vendor had a cart full of mangoes from where you could handpick the mangoes

By Shraddha Dalvi

Understanding Machine Learning with an Analogy

Given below is set of learning, human gains from his experience of shopping mangoes

- Learning 1: Bright yellow mangos are sweeter than pale yellow ones

  Experience 1:

  You were informed that bright and yellow mangoes are sweeter than pale and yellow ones.

  So you make a simple rule: pick only from the bright yellow mangoes. You check the colour of the mangoes, pick the bright yellow ones, pay up, and return home.

By Shraddha Dalvi

Understanding Machine Learning with an Analogy

Learning 2: The smaller and bright yellow mangos are sweet only half the time

Experience 2:

Now when you went home and tasted the mangoes, some of them were not sweet as you thought. You are worried as your wisdom was insufficient. You concluded that when it comes shopping mangoes, you have to look for more than just the colours.

- You concluded that the bigger and bright yellow mangoes are guaranteed to be sweet, while the smaller, bright yellow mangoes are sweet only half the time
- You will then update your rule about the mango shopping and from next time you will keep this in mind.

Understanding Machine Learning with an Analogy

Learning 3: Small pale yellow once are the sweetest of all

Experience 3:

- Tragedy: Next time at the market, you see that your favorite vendor has gone out of town. You decide to buy from a different vendor, who supplies mangoes grown from a different part of the country.
- Now, you realize that the rule which you had learnt (that big, bright yellow mangoes are the sweetest) is no longer applicable. You have to learn from scratch. You taste a mango of each kind from this vendor and realize that the small, pale yellow ones are in fact the sweetest of all.

Understanding Machine Learning with an Analogy

Learning 4: Soft mangos are juicer

Experience 4:

- One day your cousin visits you from another city. You decide to treat her with mangoes. But she is like “I don’t care about the sweetness of a mango, I only want the juiciest ones”.
- Now once again, you run your experiments, tasting all kinds of mangoes, and realizing that the softer ones are juicier.

Understanding Machine Learning with an Analogy

What if you have to write a code for it?

- As a Human Written Code: Now, imagine you were asked to write a computer program to choose your mangoes (or oranges). You might write the following rules/algorithm:

if it is bright yellow and size is big and sold by__: mango is sweet.
if (soft): mango is juicy

You would use these rules to choose the mangoes.

Understanding Machine Learning with an Analogy

Conclusion as a human:

- But every time you make a new observation from your experiments, you have to modify the list of rules manually.
- You have to understand the details of all the factors affecting the quality of mangoes. If the problem gets complicated enough, it might get difficult for you to make accurate rules by hand that covers all possible types of mangoes.
- Suppose you marry someone who hates mangoes but loves oranges instead. Now you go for shopping oranges instead of mangoes. Now, all your accumulated knowledge about mangoes is worthless.
- This will take a lot of research and effort and not everyone has this amount of time.

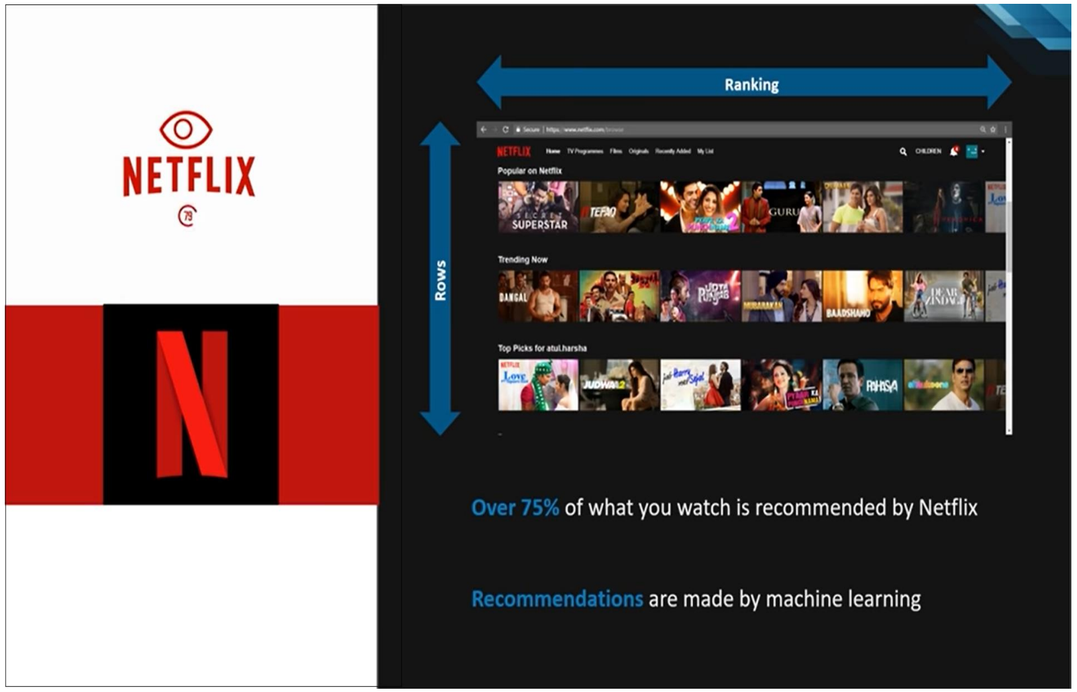
This is where Machine Learning comes into the picture

By Shraddha Dalvi

Why Machine Learning?

Netflix is using ML to recommend movies to user

- Netflix uses different algorithms that can predict and recommend the user for the new content based on the users previous watch history.
- e.g.
  - I have watched movies A, B, C, D and E
  - Netflix finds movies A, B and F have been watched along with G by millions of people
  - So Netflix recommend movies F and G for me



By Shraddha Dalvi

Why Machine Learning?

Facebook Tags

- FB is using ML to tag and post an images , it automatically recognize the face in the image and give suggestion for who is the person in the image

- Machines are defining the person in the pictures

By Shraddha Dalvi

Why Machine Learning?

Amazon Alexa

- Alexa is amazon’s virtual personal assistance

- Alexa is using ML algorithms starting from speech recognition to wave detection for answering the questions and even to the knowledge extraction and synthesis of spoken languages

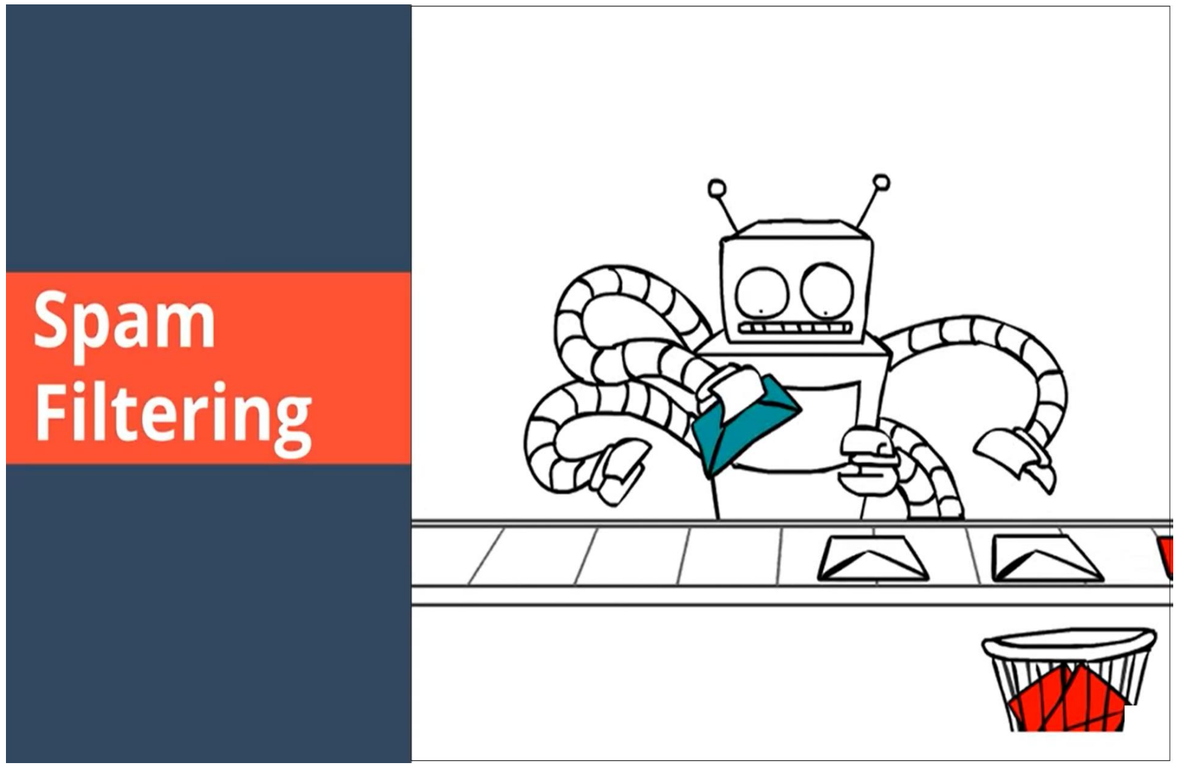
By Shraddha Dalvi

Why Machine Learning?

Spam Filtering

- Do you know how your email is getting classified and filtered as
  - Primary
  - Social
  - Promotional
  - Spam etc.

It is also done using ML algorithms.



By Shraddha Dalvi

What is Machine Learning?

Definition: 1

- Machine Learning is a concept which allows the machine to learn from examples and experience, and that too without being explicitly programmed. So instead of you writing the code, what you do is you feed data to the generic algorithm, and the algorithm/machine builds the logic based on the given data.

Definition: 2

- The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data.

By Shraddha Dalvi

What is Machine Learning?

Definition: 3

- Machine learning is an application of AI that enables systems to learn and improve from experience without being explicitly programmed. Machine learning focuses on developing computer programs that can access data and use it to learn for themselves.

Definition: 4

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.

By Shraddha Dalvi

What is Machine Learning?

Example: Training of students during exam.

- While preparing for the exams students don't actually cram the subject but try to learn it with complete understanding.
- Before the examination, they feed their machine(brain) with a good amount of high-quality data (questions and answers from different books or teachers notes or online video lectures).
- Actually, they are training their brain with input as well as output i.e. what kind of approach or logic do they have to solve a different kind of questions.
- Each time they solve practice test papers and find the performance (accuracy /score) by comparing answers with answer key given

What is Machine Learning?

Example: Training of students during exam.

- Gradually, the performance keeps on increasing, gaining more confidence with the adopted approach.
- That's how actually models are built, train machine with data (both inputs and outputs are given to model) and when the time comes test on data (with input only) and achieves our model scores by comparing its answer with the actual output which has not been fed while training.
- Researchers are working hard to improve algorithms, techniques so that these models perform even much better

What is Machine Learning?

Example: Traditional Programming & Machine Learning

- Traditional Programming : We feed in DATA (Input) + PROGRAM (logic), run it on machine and get output. E.g. C Program to check the number is even or odd

```c
#include <stdio.h>
int main() {
    int num;
    printf("Enter an integer: ");   INPUT
    scanf("%d", &num);

    // true if num is perfectly divisible by 2
    if(num % 2 == 0)   LOGIC
        printf("%d is even.", num);
    else                                  OUTPUT
        printf("%d is odd.", num);

    return 0;
}
```

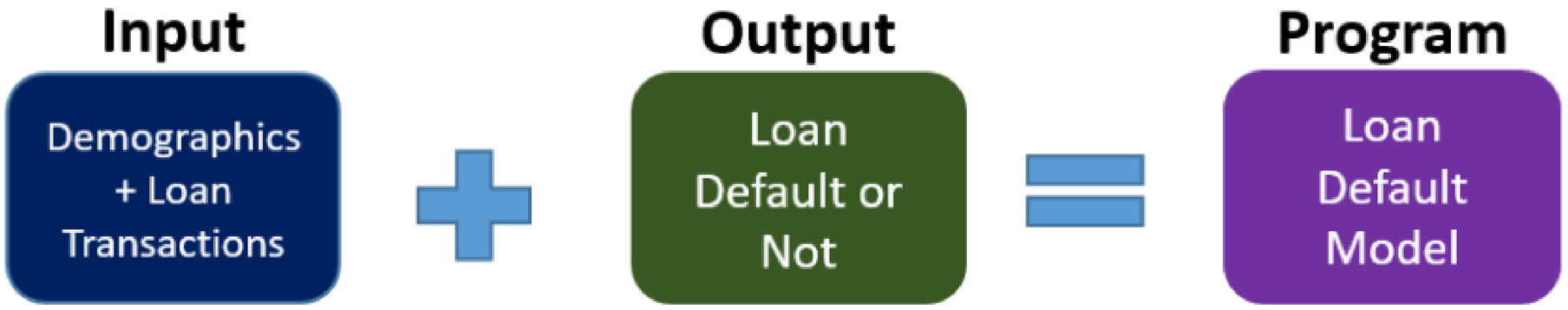
By Shraddha Dalvi

What is Machine Learning?

Example: Traditional Programming & Machine Learning

- Machine Learning : We feed in DATA(Input) + Output, run it on machine during training and the machine creates its own program(logic), which can be evaluated while testing.

  e.g. feed-in customer information/loan transactions (input) and how many defaulted on the loan (observed output), and it will create a model to predict who will default on the loan.



What is Machine Learning?

Example:

Basic Difference in ML and Traditional Programming?

- Traditional Programming : We feed in DATA (Input) + PROGRAM (logic), run it on machine and get output.
- Machine Learning : We feed in DATA(Input) + Output, run it on machine during training and the machine creates its own program(logic), which can be evaluated while testing.

Basics of Machine Learning

1. Dataset:

A dataset refers to a **collection of data** that is used for **training** and **evaluating** a Machine Learning model. It consists of input **features** (also called predictors or independent variables) and corresponding output **labels** (also called target or dependent variables).

Basics of Machine Learning

- 1. Dataset:

#	Age	Has_Job	Own_House	Credit_Rating	Class
1	Young	false	false	fair	No
2	Young	false	false	good	No
3	Young	true	false	good	Yes
4	Young	true	true	fair	Yes
5	Young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

Basics of Machine Learning

2. Model:

- A model represents a **mathematical** or **computational** representation of the relationships between the input features and output labels.
- It is built using Machine Learning **algorithms** and is trained on a dataset to learn patterns, make predictions, or classify data.

Basics of Machine Learning

3. Training Set:

- The training set is a **subset** of the dataset used to train the Machine Learning model.
- It contains **input-output pairs** that the model uses to learn the underlying patterns or relationships.

4. Testing Set:

- The testing set is a separate subset of the dataset that is used to evaluate the performance of the trained model.
- It contains data instances that were **not seen during training**, allowing us to assess how well the model generalizes to unseen examples.

Basics of Machine Learning

5. Validation:

- Validation is an **optional step** used during the training process to tune the hyperparameters of the model.
- Hyperparameters are parameters that are not learned from the data but affect the behavior of the model.
- By using a validation set, which is distinct from the training and testing sets, we can fine-tune the model's hyperparameters to optimize its performance.

Basics of Machine Learning

6. Accuracy:

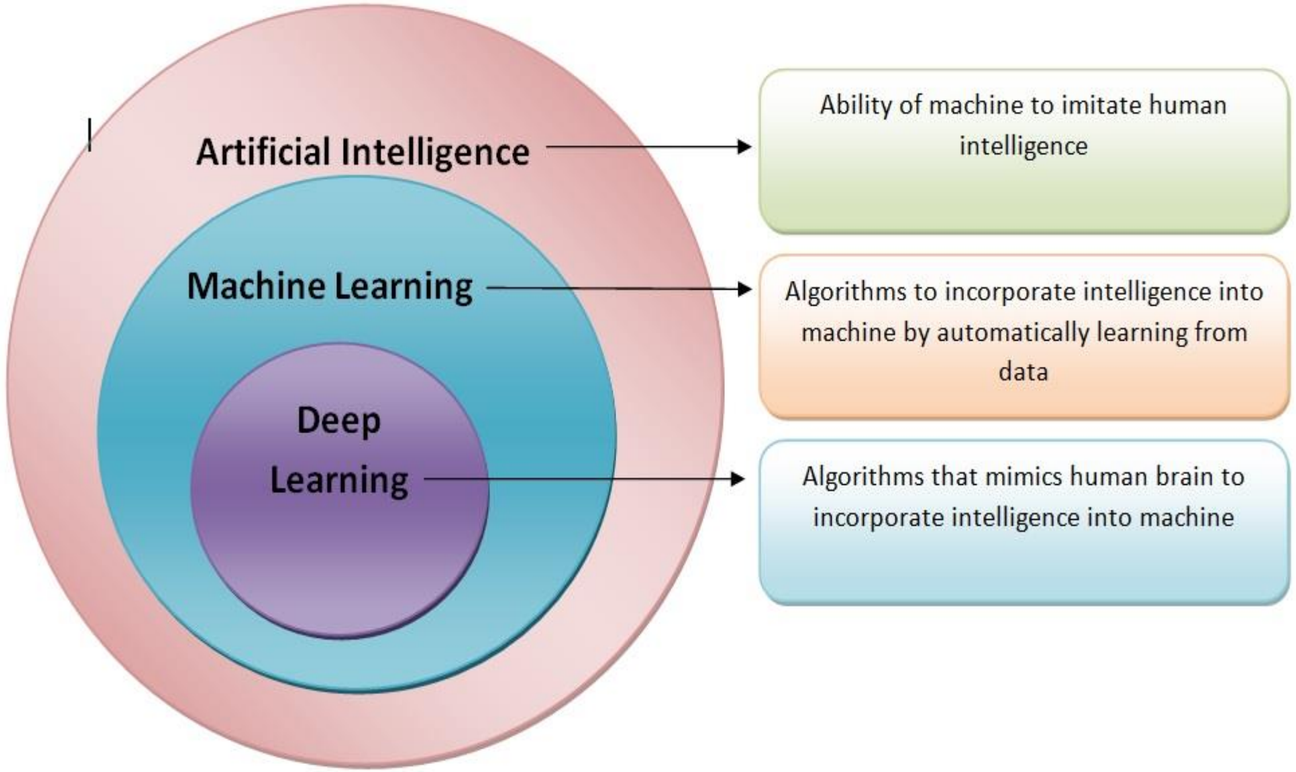
- Accuracy is a **performance metric** used to evaluate the **effectiveness** of a Machine Learning model.
- It measures the proportion **of correctly predicted instances** (or labels) in relation to the total number of instances.
- It is commonly used for classification tasks.

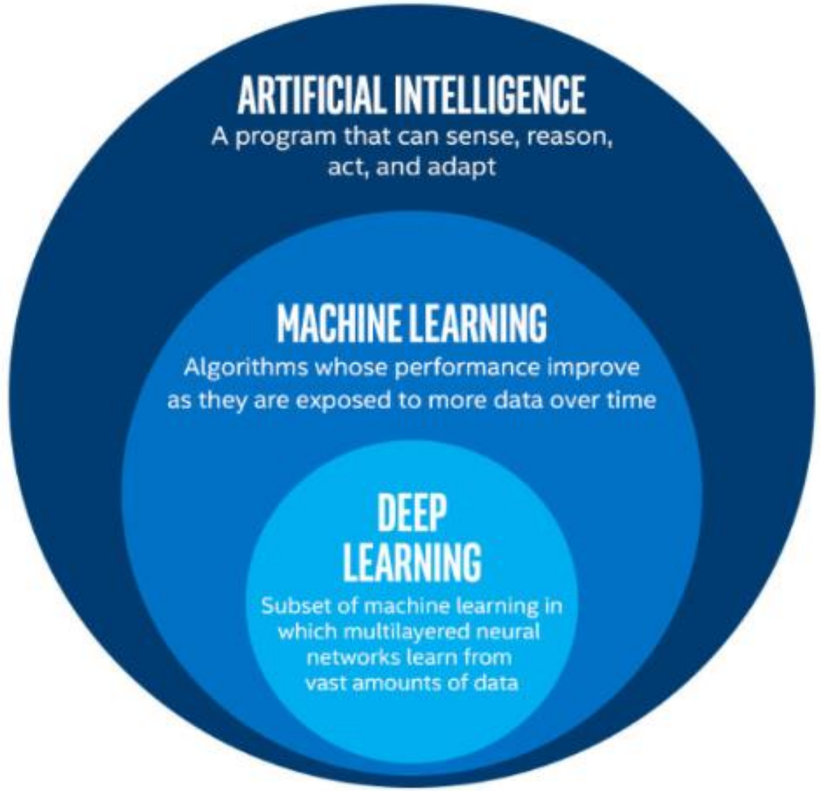
Basics of Machine Learning

- We split the dataset into a training set, a testing set, and a validation set.
- Generally the training set contains 70% of the data and is used to train the model.
- The validation set comprises 10% of the data and is utilized to fine-tune the model's hyperparameters.
- The remaining 20% of the data forms the testing set, which evaluates the model's performance on unseen examples.
- After training the model, we measure its accuracy using the testing set. For instance, if the model correctly classifies 90 out of 100 cases in the testing set, the accuracy would be 90%.

Biggest Confusion AI vs ML vs Deep Learning

- The term AI was coined by John McCarthy in 1956
- The term ML was coined in 1959 by Arthur Samuel
- The term Deep Learning was introduced by Rina Dechter in 1986, and ANN by Igor Aizenberg in 2000





Biggest Confusion AI vs ML vs Deep Learning

People think all three of them the AI, ML and the Deep Learning are same. But this is WRONG!,

- Artificial Intelligence

Artificial Intelligence is the broader concept of machines being able to carry out tasks in a smarter way. It covers anything which enables the computers to behave like humans.

Biggest Confusion AI vs ML vs Deep Learning

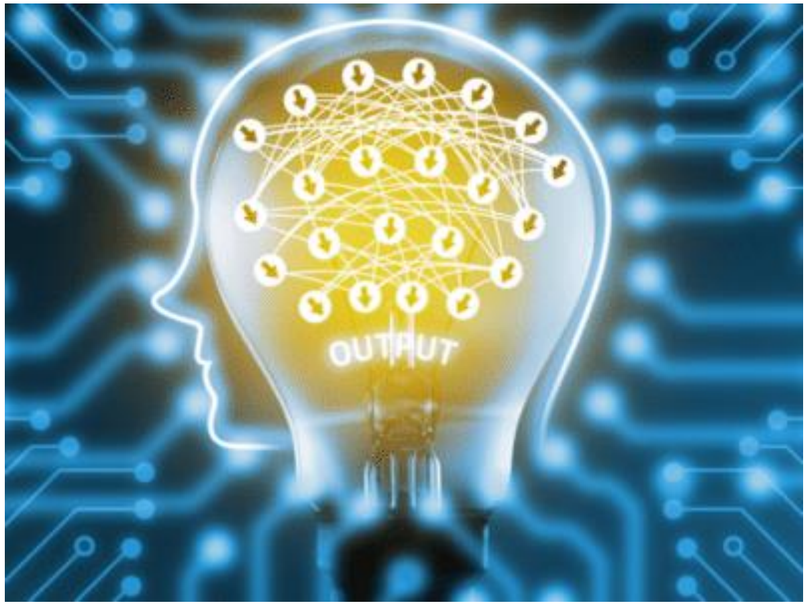
- Machine Learning

Machine Learning is a subset of AI and is based on the idea that machines should be given the access to data, and should be left to learn and explore for themselves. It deals with the extraction of patterns from a large data sets.

Biggest Confusion AI vs ML vs Deep Learning

- Deep Learning

Deep Learning is a subset of Machine Learning where similar Machine Learning Algorithms are used to train Deep Neural Networks so as to achieve better accuracy in those cases where former was not performing up to the mark.



Data Science Life Cycle

By Shraddha Dalvi

Data Science Life Cycle

1. Problem Definition:

- Identify the business problem or research question that the data science project aims to address.
- Understand the project objectives, requirements, and constraints to define the scope of the analysis.

Data Science Life Cycle

2. Data Collection:

- Gather relevant data from various sources, such as databases, APIs, spreadsheets, or web scraping.
- Clean and preprocess the data to handle missing values, outliers, and ensure data quality.

Data Science Life Cycle

3. Data Exploration and Analysis:

- Perform exploratory data analysis (EDA) to gain insights into the data's characteristics and patterns.
- Use descriptive statistics, data visualizations, and correlation analysis to understand the relationships between variables.

Data Science Life Cycle

4. Feature Engineering:

- Select or engineer relevant features from the data that can be used as input for the machine learning models.
- Transform and preprocess the data to make it suitable for modeling.

Data Science Life Cycle

5. Model Development:

- Choose appropriate machine learning algorithms based on the problem type (e.g., classification, regression, clustering).
- Split the data into training and testing sets for model evaluation.
- Train the selected models on the training data and fine-tune their hyperparameters.

Data Science Life Cycle

6. Model Evaluation:

- Evaluate the performance of the trained models using the testing set and relevant evaluation metrics (e.g., accuracy, precision, recall, F1-score).
- Compare different models and select the best-performing one.

Data Science Life Cycle

7. Model Deployment:

- Integrate the selected model into the production environment or the intended system to make predictions on new, unseen data.
- Monitor the model's performance in the real-world setting.

Data Science Life Cycle

8. Communication and Visualization:

- Present the findings, insights, and results of the data analysis in a clear and understandable manner.
- Use data visualizations, reports, and presentations to communicate the outcomes to stakeholders.

Data Science Life Cycle

9. Decision Making and Action:

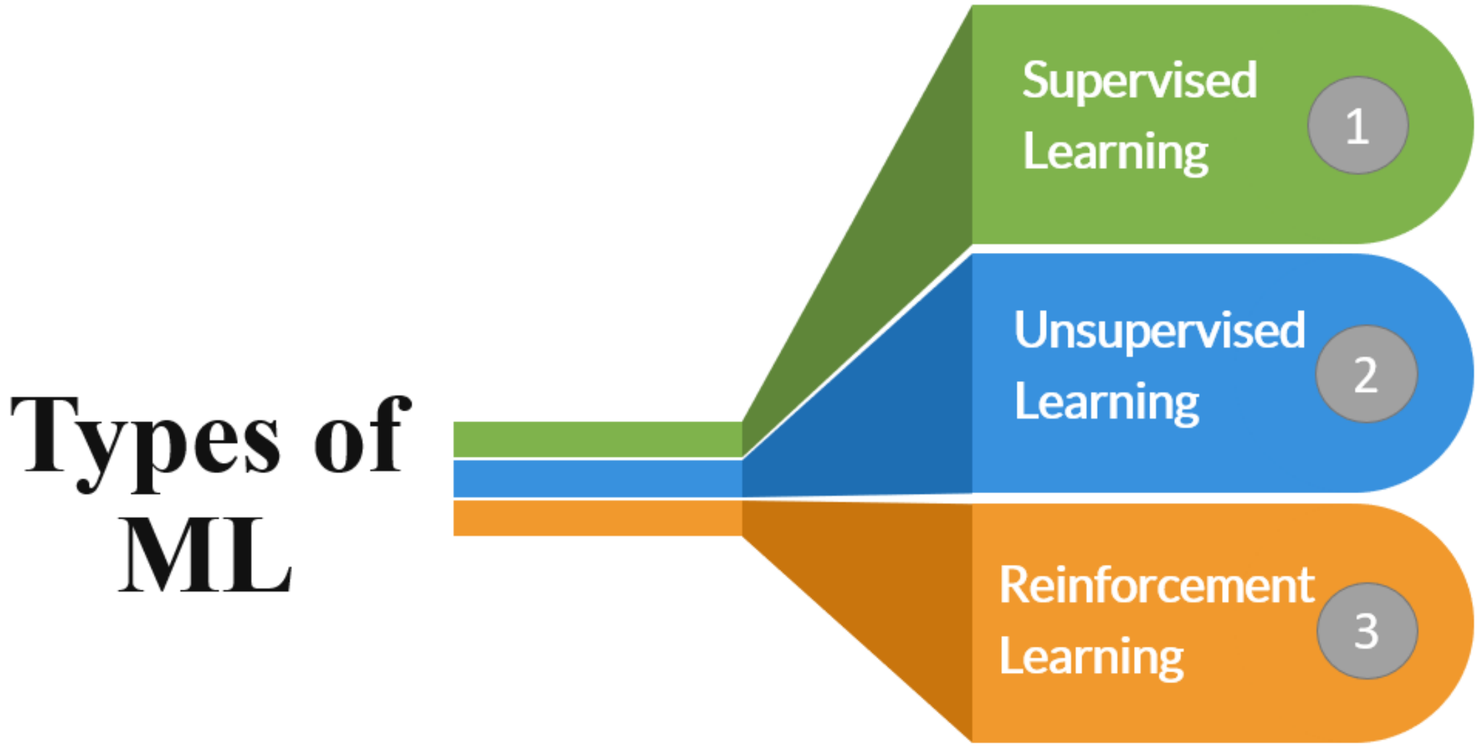
- Based on the insights and predictions from the model, make data-driven decisions or take appropriate actions to address the original problem.

Data Science Life Cycle

10. Maintenance and Iteration:

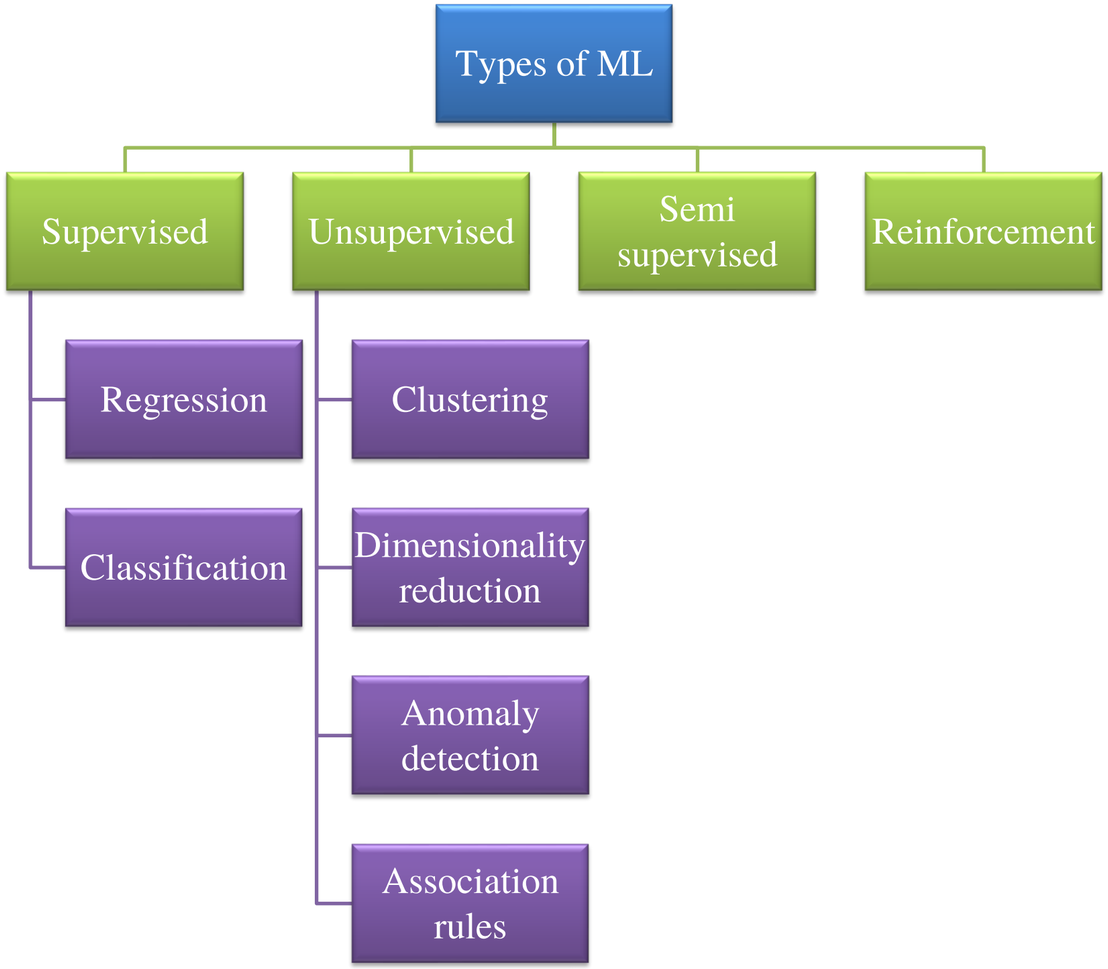
- Continuously monitor and maintain the deployed model, updating it when necessary to adapt to changing data or requirements.
- Iterate and improve the data science process based on feedback and new insights.

Types of Machine Learning



By Shraddha Dalvi

Types of Machine Learning



By Shraddha Dalvi

Types of Machine Learning

Supervised Learning

Unsupervised Learning

Reinforcement Learning

By Shraddha Dalvi

Types of Machine Learning

Supervised Learning

Definition:

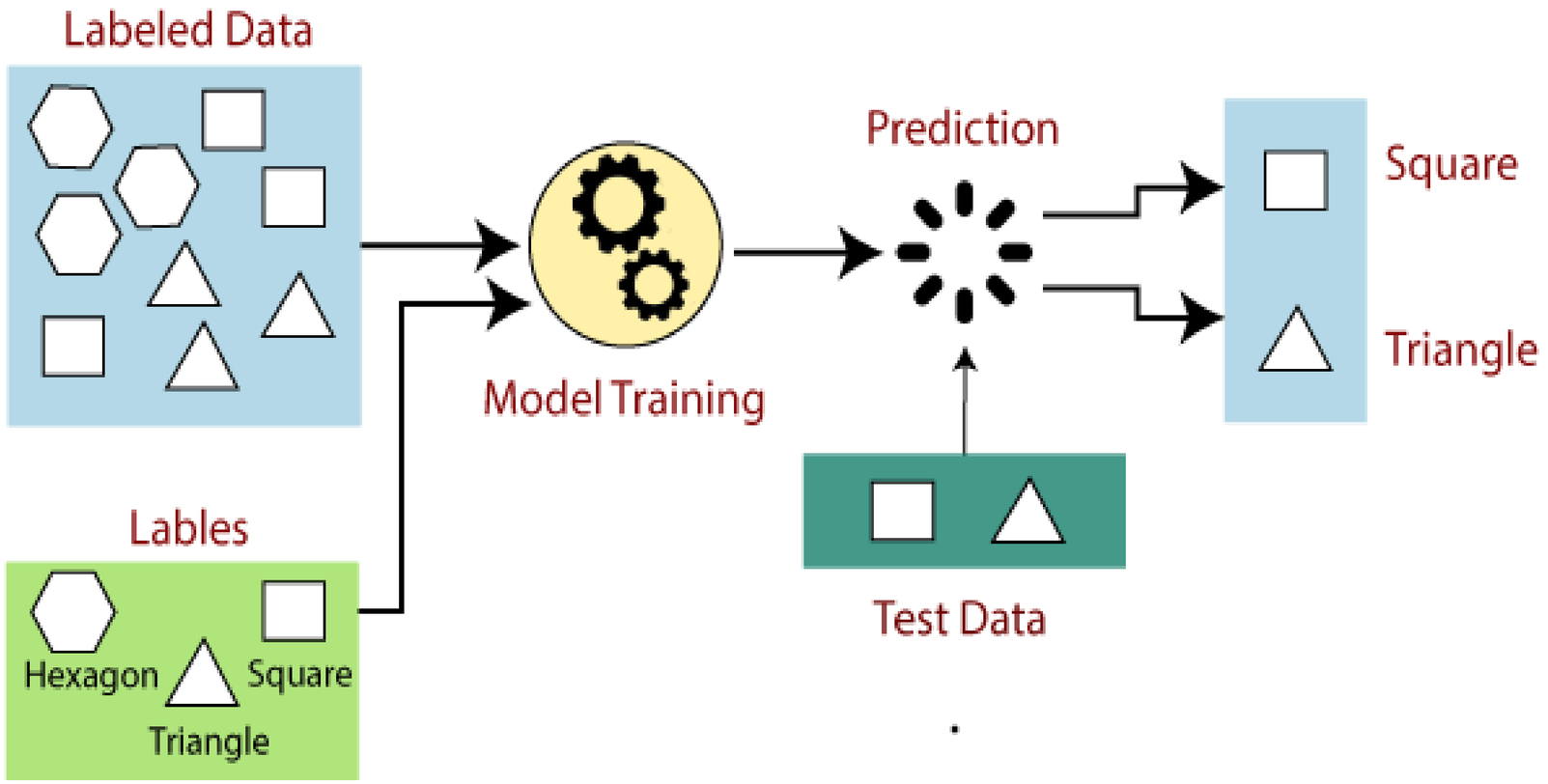
Supervised learning is the types of machine learning where a computer algorithm is trained on input data that has been labeled for a particular output.

The model is trained until it can detect the underlying patterns and relationships between the input data and the output labels, enabling it to yield accurate labeling results when presented with never-before-seen data.

By Shraddha Dalvi

Types of Machine Learning

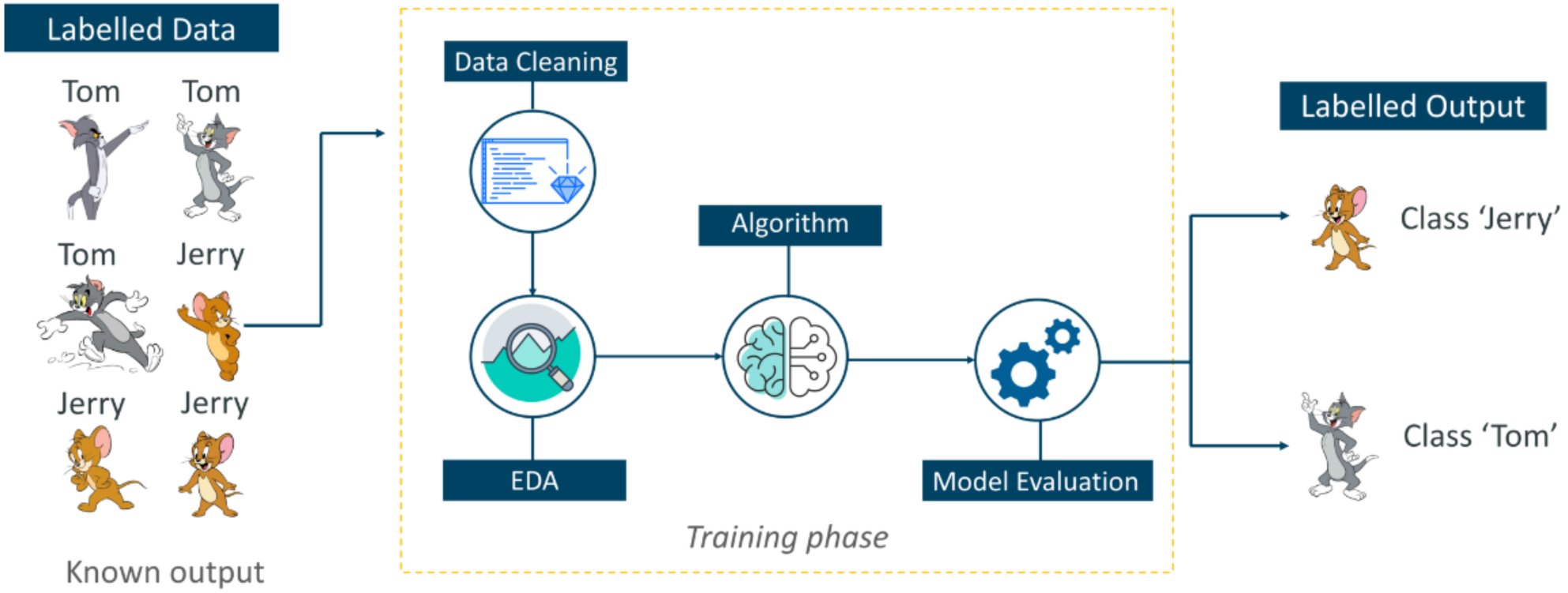
Supervised Learning



By Shraddha Dalvi

Types of Machine Learning

Supervised Learning



By Shraddha Dalvi

Types of Machine Learning

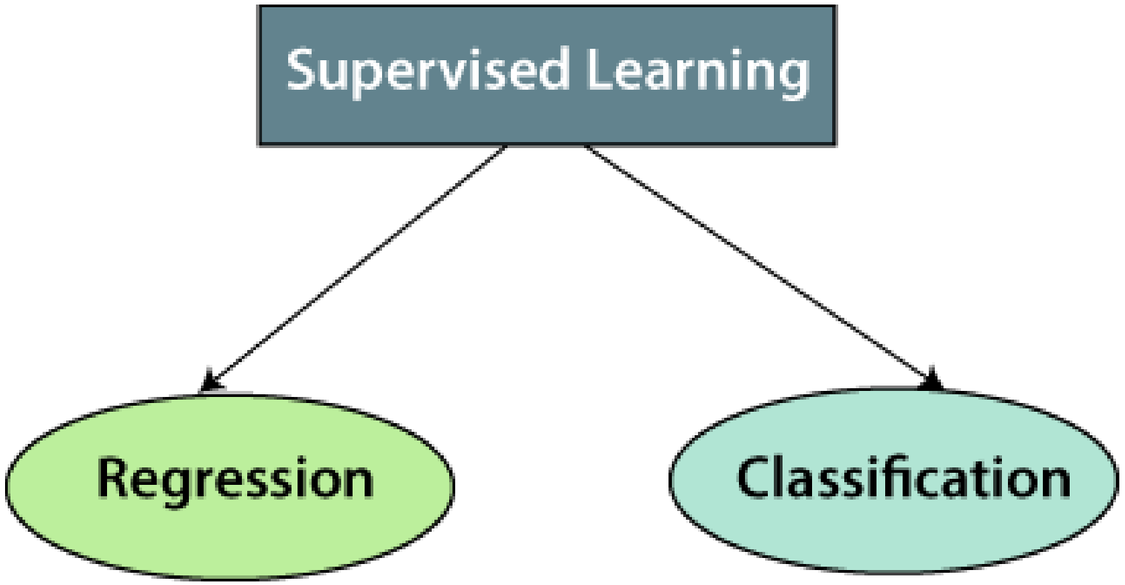
Supervised Learning

Mathematical definition of Supervised Learning

- Supervised learning is the one where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.
- $Y = f(X)$
- The goal is to approximate the mapping function so well that whenever you get some new input data (x), the machine can easily predict the output variables (Y) for that data.

Types of Machine Learning

Types of supervised Machine learning Algorithms:



By Shraddha Dalvi

Types of Machine Learning

Types of supervised Machine learning Algorithms:

Regression:

Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.

Below are some popular Regression algorithms which come under supervised learning:

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression

By Shraddha Dalvi

Types of Machine Learning

Types of supervised Machine learning Algorithms:

Classification:

Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines

By Shraddha Dalvi

Types of Machine Learning

Unsupervised Learning

- As the name suggests, unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data. It can be compared to learning which takes place in the human brain while learning new things. It can be defined as:
- *"Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision."*
- Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

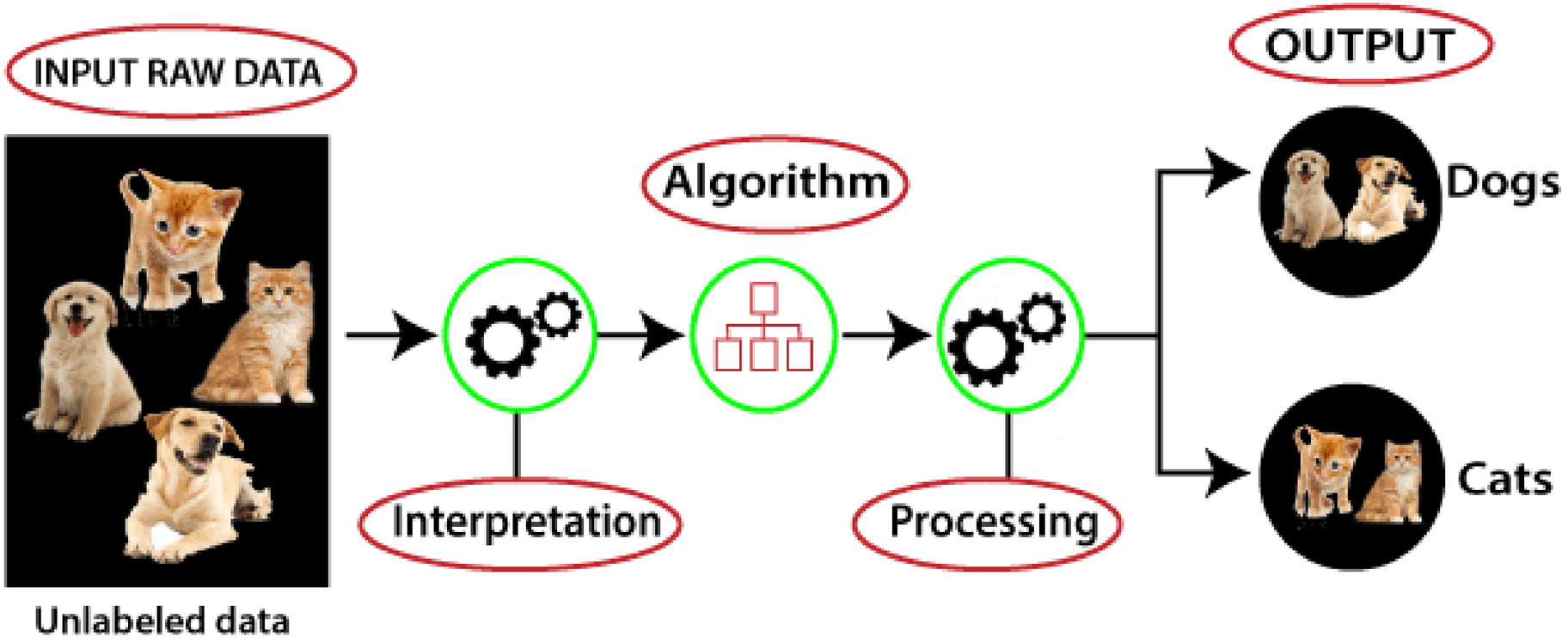
Types of Machine Learning

Example:

- Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs.
- The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset.
- The task of the unsupervised learning algorithm is to identify the image features on their own.
- Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.

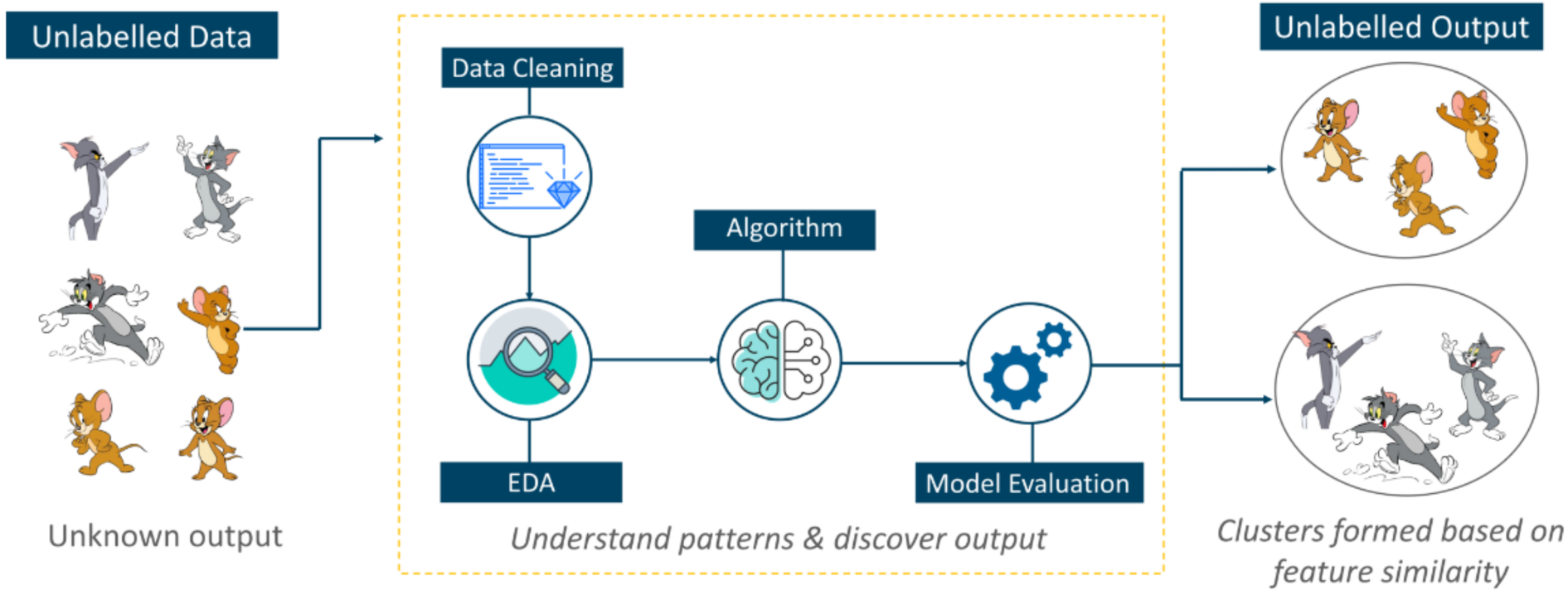
Types of Machine Learning

Example:



Types of Machine Learning

Example:



By Shraddha Dalvi

Types of Machine Learning

Why use Unsupervised Learning?

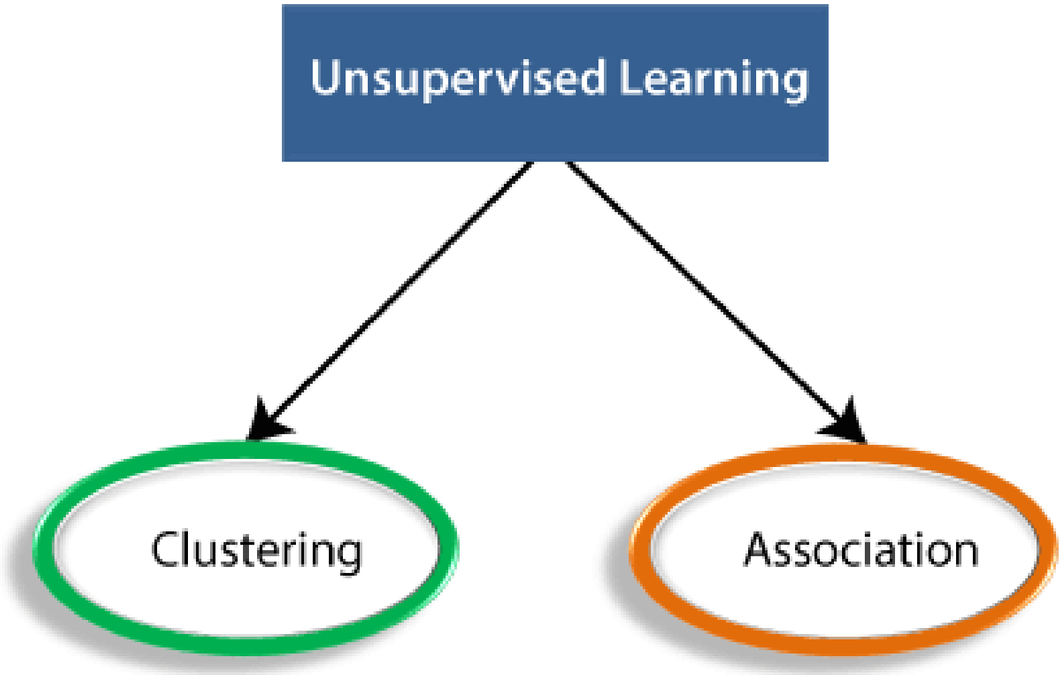
Below are some main reasons which describe the importance of Unsupervised Learning:

- Unsupervised learning is helpful for finding useful insights from the data.
- Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- Unsupervised learning works on unlabeled and uncategorized data which make unsupervised learning more important.
- In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

By Shraddha Dalvi

Types of Machine Learning

Types of Unsupervised Learning Algorithm:



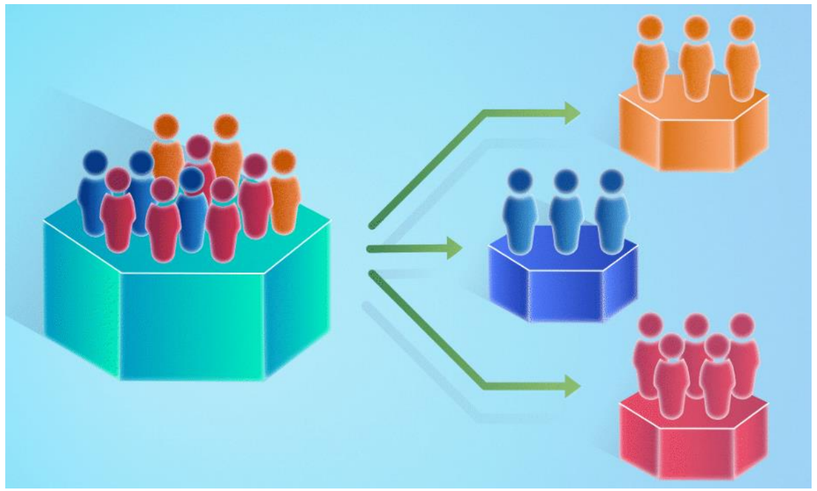
By Shraddha Dalvi

Types of Machine Learning

Types of Unsupervised Learning Algorithm:

Clustering:

- Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group.
- Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.

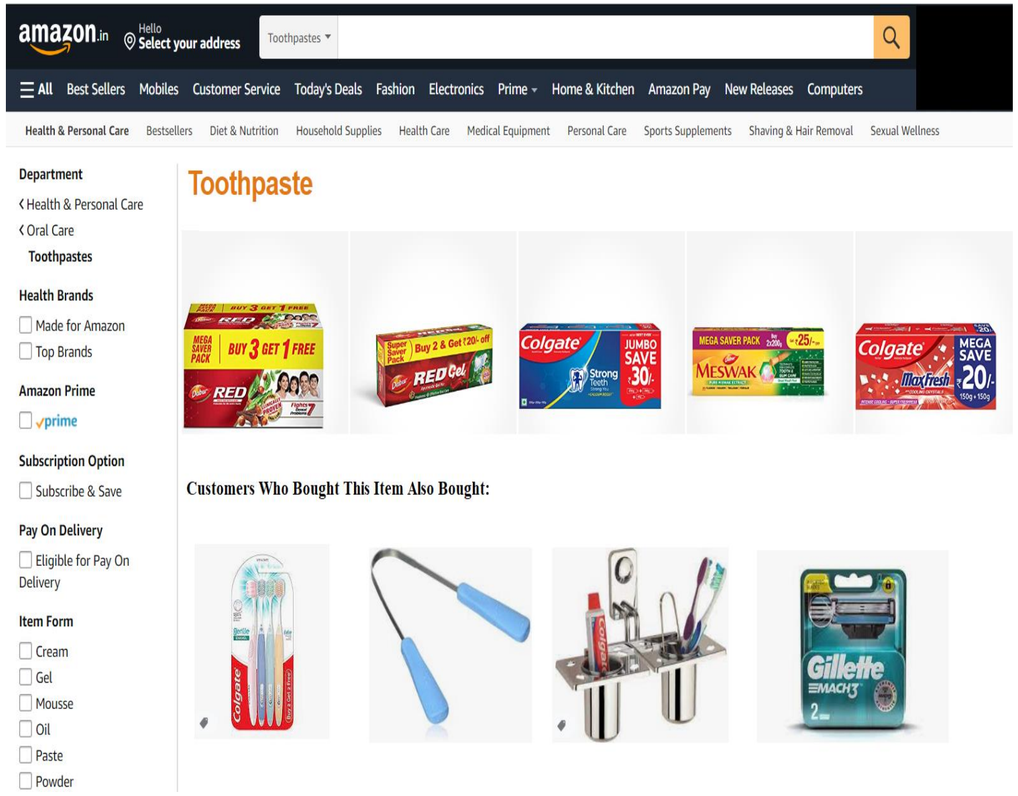


Types of Machine Learning

Types of Unsupervised Learning Algorithm:

Association:

- An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database.
- It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item.
- A typical example of Association rule is Market Basket Analysis.



By Shraddha Dalvi

Types of Machine Learning

Unsupervised Learning algorithms:

Below is the list of some popular unsupervised learning algorithms:

- K-means clustering
- KNN (k-nearest neighbors)
- Hierarchal clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

By Shraddha Dalvi

Types of Machine Learning

Reinforcement Learning

- Reinforcement Learning:

  Reinforcement Learning is a part of Machine learning where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from those actions.

Types of Machine Learning

What is Reinforcement Learning?

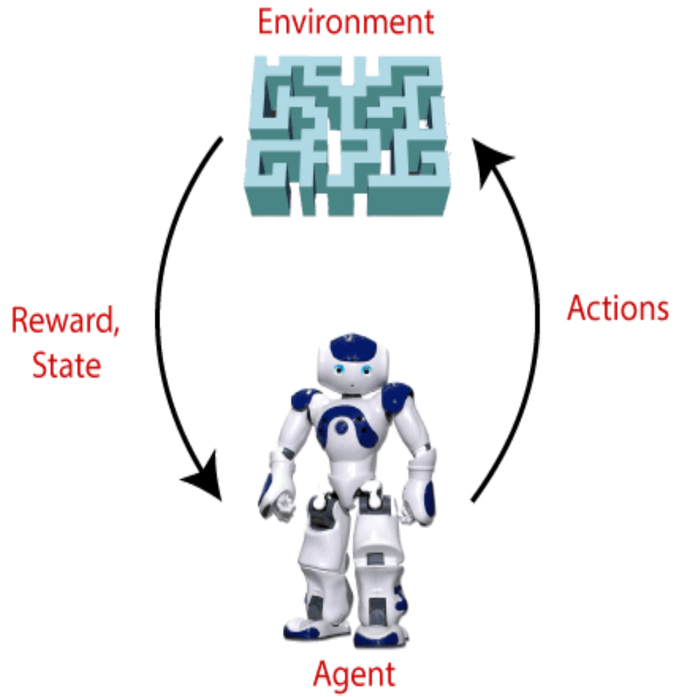
- Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions. For each good action, the agent gets positive feedback, and for each bad action, the agent gets negative feedback or penalty.
- In Reinforcement Learning, the agent learns automatically using feedbacks without any labeled data, unlike supervised learning.
- Since there is no labeled data, so the agent is bound to learn by its experience only.
- RL solves a specific type of problem where decision making is sequential, and the goal is long-term, such as game-playing, robotics, etc.

By Shraddha Dalvi

Types of Machine Learning

What is Reinforcement Learning?

- Example 1: Suppose there is an AI agent present within a maze environment, and his goal is to find the diamond. The agent interacts with the environment by performing some actions, and based on those actions, the state of the agent gets changed, and it also receives a reward or penalty as feedback.
- The agent continues doing these three things (take action, change state/remain in the same state, and get feedback), and by doing these actions, he learns and explores the environment.
- The agent learns that what actions lead to positive feedback or rewards and what actions lead to negative feedback or penalty.



Types of Machine Learning

What is Reinforcement Learning?

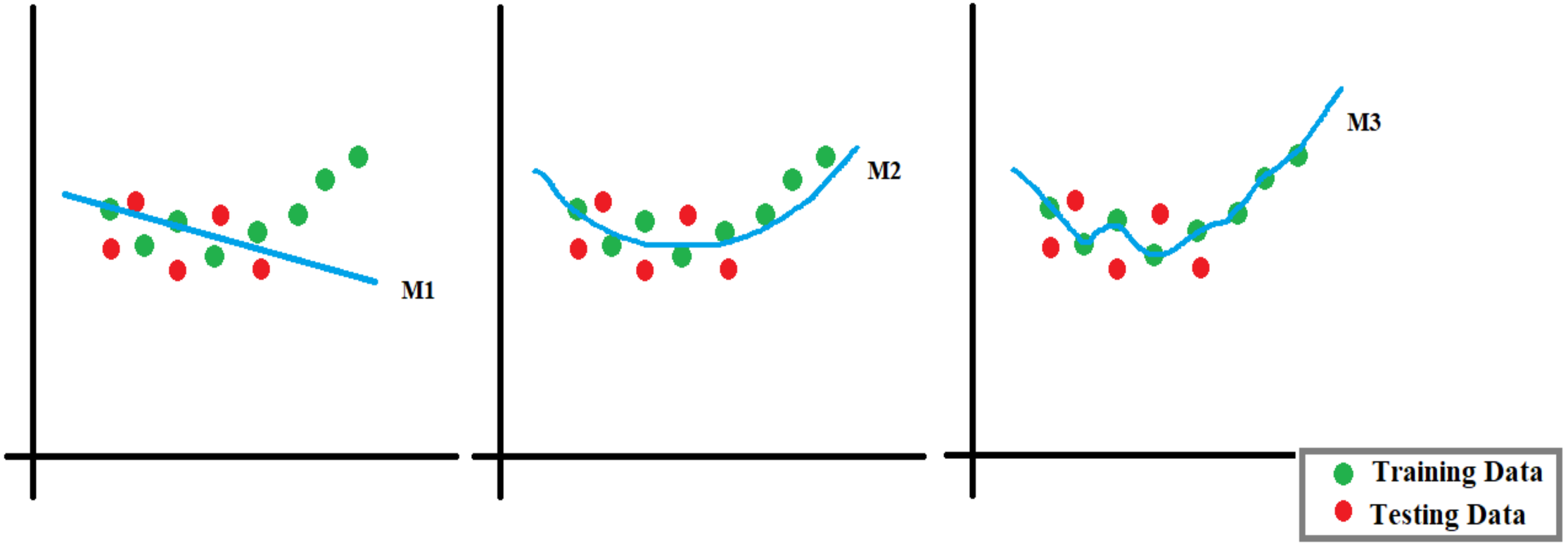
- Example 2: Imagine that you were dropped off at an isolated island! What would you do?
- Panic? Yes, initially we all would. But as time passes by, you will learn how to live on the island. You will explore the environment, understand the climate condition, the type of food that grows there, the dangers of the island, etc.
- This is exactly how Reinforcement Learning works, it involves an Agent that is put in an unknown environment where agent must learn by observing and performing actions that result in rewards.

Bias and Variance in ML

- Machine learning allows machines to perform data analysis and make predictions. However, if the machine learning model is not accurate, it can make predictions errors, and these prediction errors are usually known as Bias and Variance. In machine learning, these errors will always be present as there is always a slight difference between the model predictions and actual predictions.
- The main aim of ML/data science analysts is to reduce these errors in order to get more accurate results.

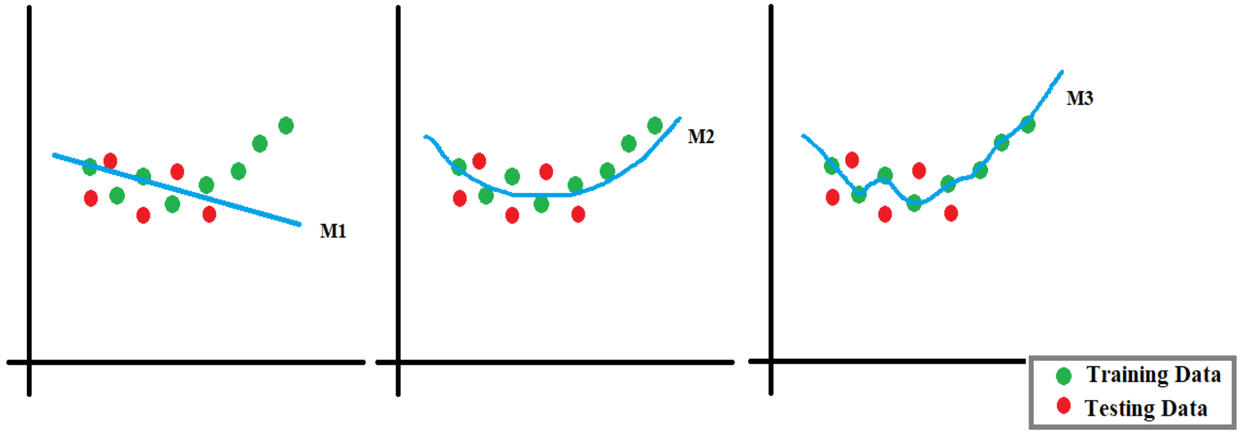
By Shraddha Dalvi

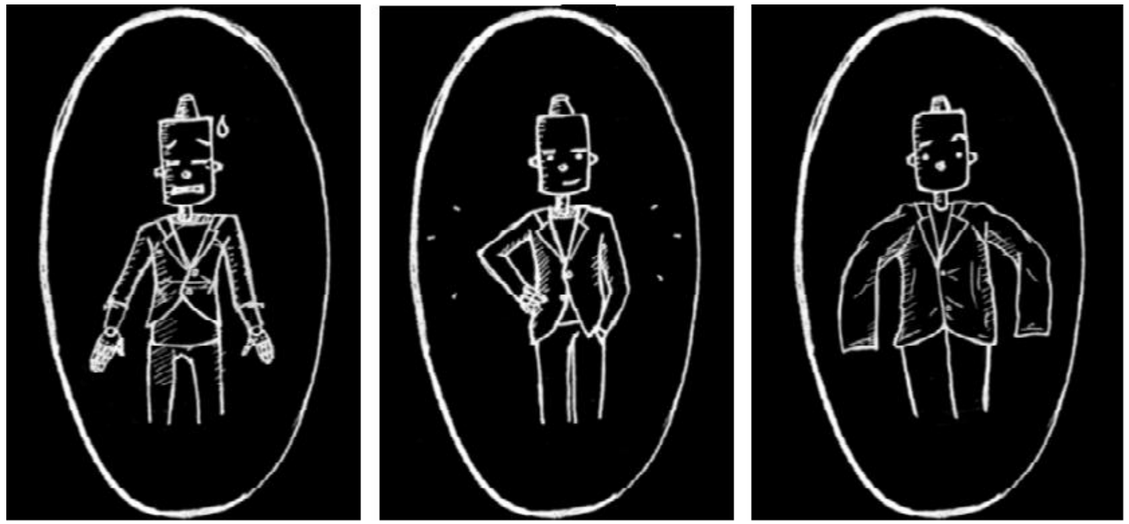
Bias and Variance in ML



By Shraddha Dalvi

Bias and Variance in ML





Underfitting | Best Fit | Overfitting

By Shraddha Dalvi

Bias and Variance in ML

What is Bias?

- While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors or Errors due to bias.

Low Bias:

- A low bias model will make fewer assumptions about the form of the target function.

High Bias:

- A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. A high bias model also cannot perform well on new data.

By Shraddha Dalvi

Bias and Variance in ML

What is a Variance?

- The variance would specify the amount of variation in the prediction if the different training data was used. In simple words, .

- Low variance :
  - Low variance means there is a small variation in the prediction of the target function with changes in the training data set.
- High variance :
  - High variance shows a large variation in the prediction of the target function with changes in the training dataset.

variance quantifies the dispersion or spread of a set of data points. It measures how much individual data points deviate from the mean (expected value). Here's a brief explanation:

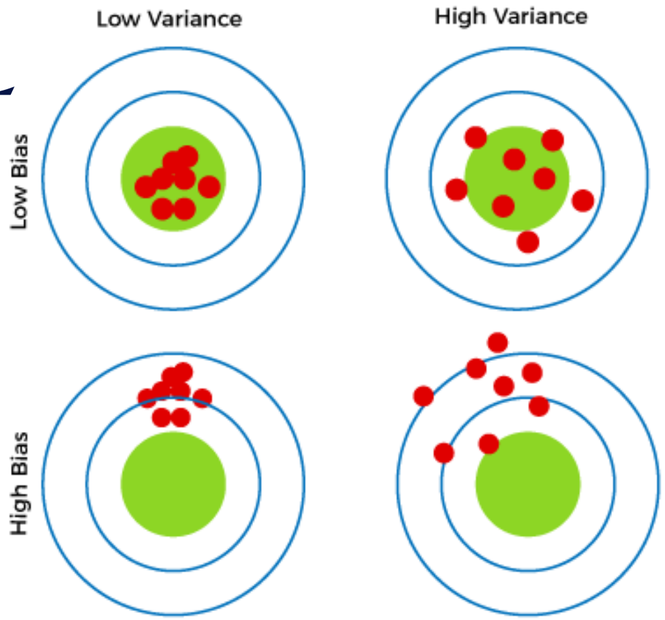
Bias and Variance in ML

Different Combinations of Bias-Variance

- There are four possible combinations of bias and variances, which are represented by the below diagram:

- Low-Bias, Low-Variance
- Low-Bias, High-Variance
- High-Bias, Low-Variance
- High-Bias, High-Variance

Bias and Variance in ML



Different Combinations of Bias-Variance

1. Low-Bias, Low-Variance

- The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.

2. Low-Bias, High-Variance

- With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an overfitting

A high variance indicates that the data points are widely spread out from the mean.
A low variance suggests that the data points are closely clustered around the mean.

3. High-Bias, Low-Variance

- With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to underfitting problems in the model.

4. High-Bias, High-Variance

- With high bias and high variance, predictions are inconsistent and also inaccurate on average.

By Shraddha Dalvi

Overfitting and Underfitting in ML

- Overfitting and Underfitting are the two main problems that occur in machine learning and degrade the performance of the machine learning models.
- Before understanding the overfitting and underfitting, let's understand some basic term that will help to understand this topic well:

Signal: It refers to the true underlying pattern of the data that helps the machine learning model to learn from the data.

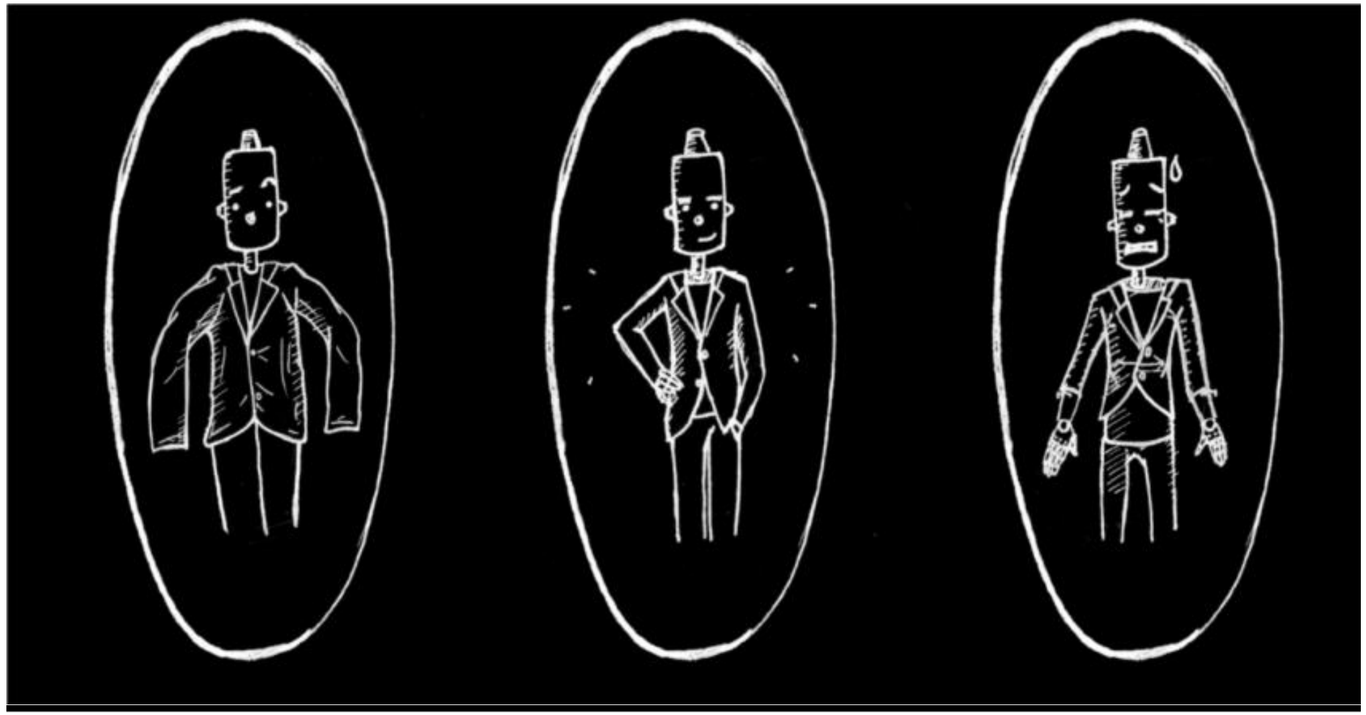
Noise: Noise is unnecessary and irrelevant data that reduces the performance of the model.

Bias: Bias is a prediction error that is introduced in the model due to oversimplifying the machine learning algorithms. Or it is the difference between the predicted values and the actual values.

Variance: If the machine learning model performs well with the training dataset, but does not perform well with the test dataset, then variance occurs.

By Shraddha Dalvi

Overfitting and Underfitting in ML



Overfitting **Best-Fit** **Underfitting**

- https://www.analyticsvidhya.com/blog/2020/02/underfitting-overfitting-best-fitting-machine-learning/

By Shraddha Dalvi

Overfitting and Underfitting in ML

Overfitting:

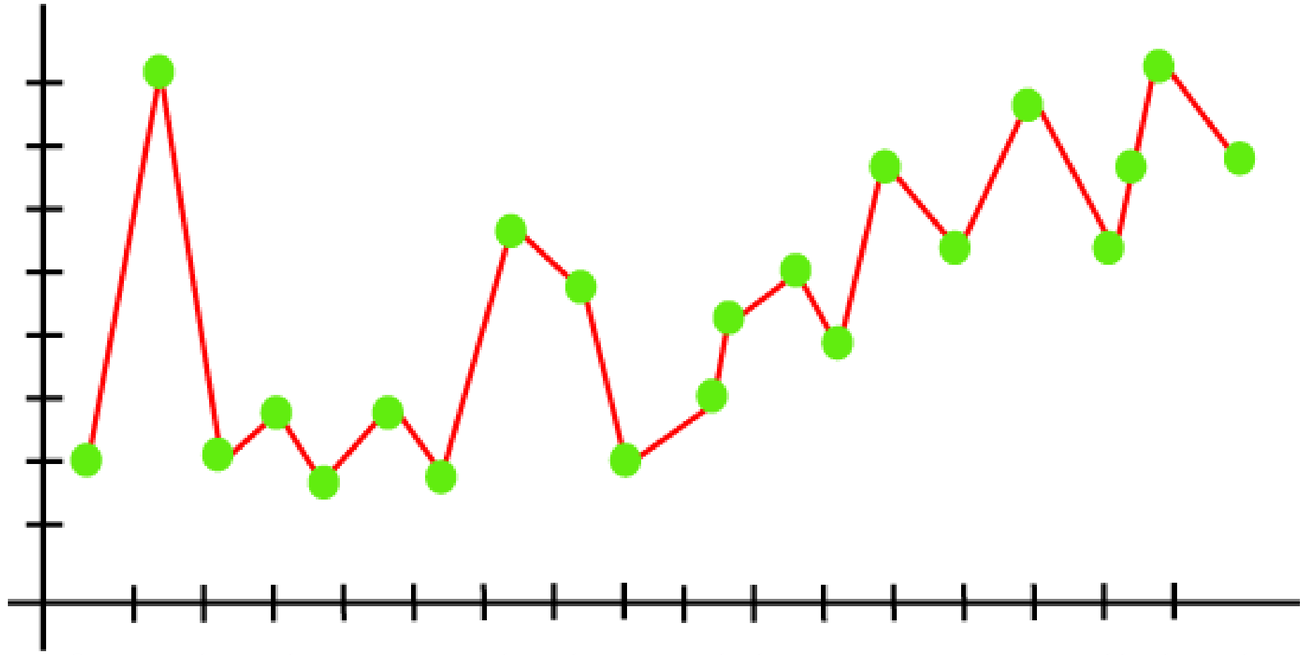
- Overfitting occurs when our machine learning model tries to cover all the data points or more than the required data points present in the given dataset. Because of this, the model starts caching noise and inaccurate values present in the dataset, and all these factors reduce the efficiency and accuracy of the model. The overfitted model has low bias and high variance.
- The chances of occurrence of overfitting increase as much we provide training to our model. It means the more we train our model, the more chances of occurring the overfitted model.

By Shraddha Dalvi

Overfitting and Underfitting in ML

Overfitting:

- Example: The concept of the overfitting can be understood by the below graph of the linear regression output:



- As we can see from the above graph, the model tries to cover all the data points present in the scatter plot. It may look efficient, but in reality, it is not so. Because the goal of the regression model to find the best fit line, but here we have not got any best fit, so, it will generate the prediction errors.

Overfitting and Underfitting in ML

Underfitting

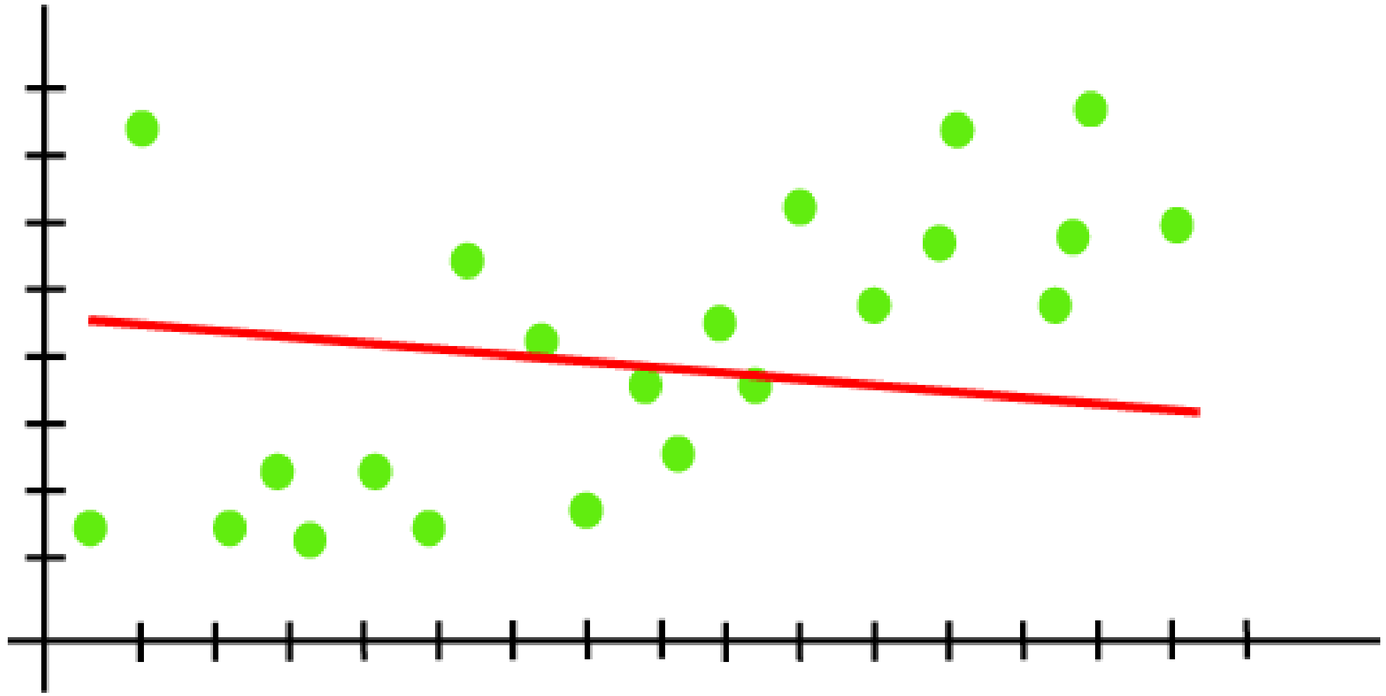
- Underfitting occurs when our machine learning model is not able to capture the underlying trend of the data. To avoid the overfitting in the model, the fed of training data can be stopped at an early stage, due to which the model may not learn enough from the training data. As a result, it may fail to find the best fit of the dominant trend in the data.
- In the case of underfitting, the model is not able to learn enough from the training data, and hence it reduces the accuracy and produces unreliable predictions.

By Shraddha Dalvi

Overfitting and Underfitting in ML

Underfitting

- Example: We can understand the underfitting using below output of the linear regression model:



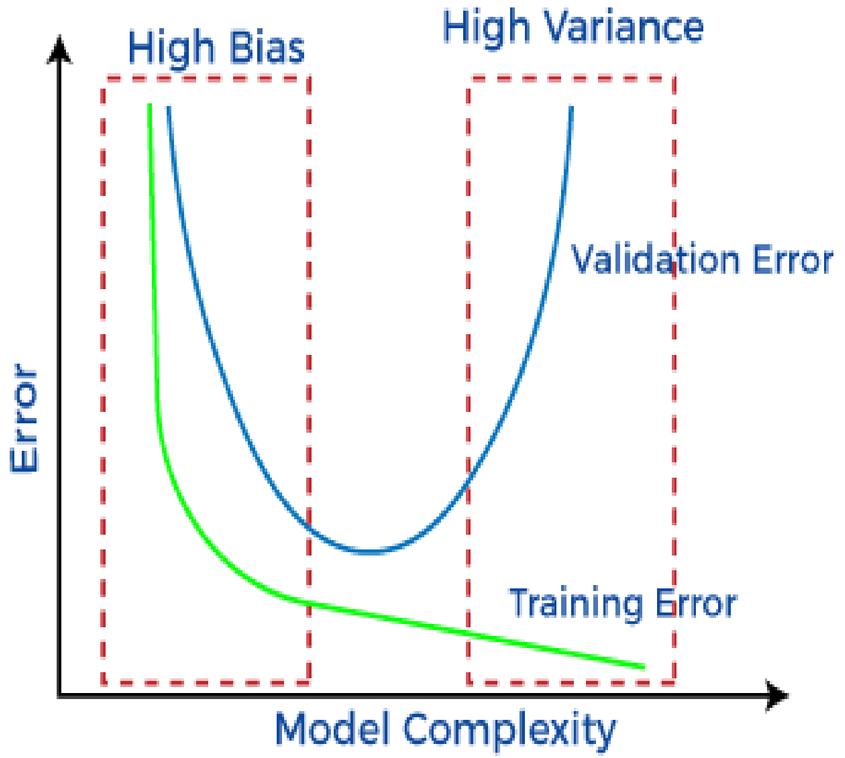
- As we can see from the above diagram, the model is unable to capture the data points present in the plot.

By Shraddha Dalvi

Bias and Variance in ML

Bias-Variance Trade-Off

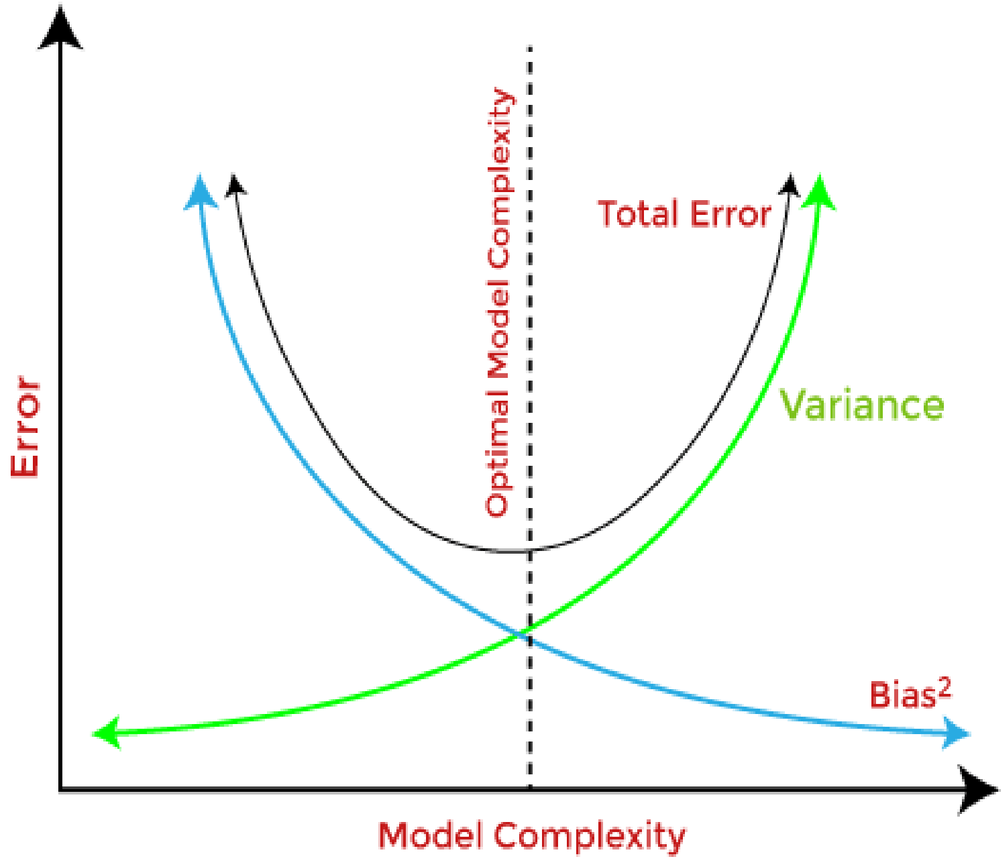
- While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model.
- If the model is very simple with fewer parameters, it may have low variance and high bias. Whereas, if the model has a large number of parameters, it will have high variance and low bias.
- So, it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as the Bias-Variance trade-off.



Bias and Variance in ML

Bias-Variance Trade-Off

- For an accurate prediction of the model, algorithms need a low variance and low bias. But this is not possible because bias and variance are related to each other:
  - If we decrease the variance, it will increase the bias.
  - If we decrease the bias, it will increase the variance.



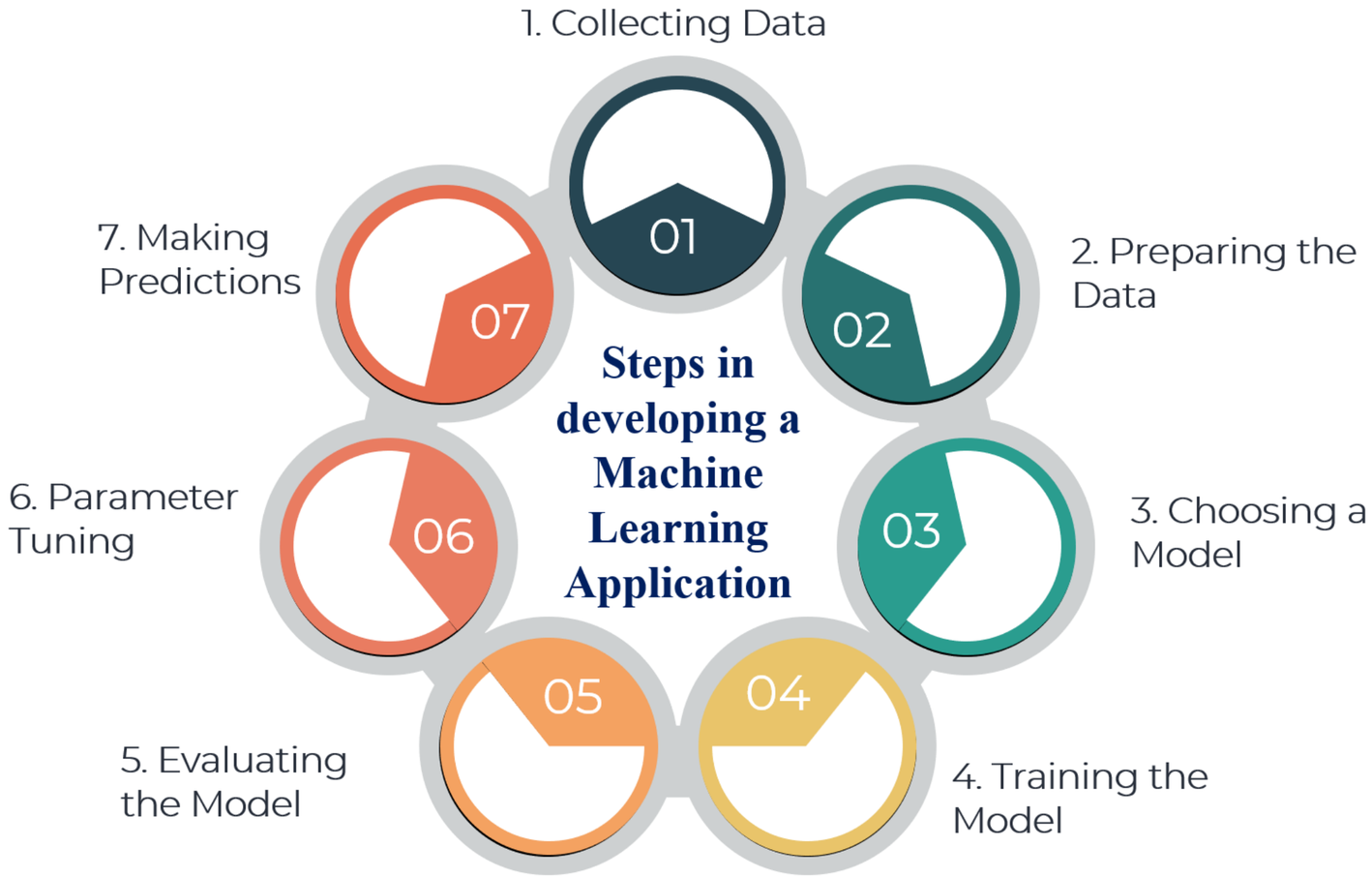
Hence, the Bias-Variance trade-off is about finding the sweet spot to make a balance between bias and variance errors. https://youtu.be/EuBBz3bI-aA?feature=shared

By Shraddha Dalvi

Steps in developing a Machine Learning Application

There are 7 major steps to develop a Machine Learning Application

1. Collecting Data:
2. Preparing the Data
3. Choosing a Model
4. Training the Model
5. Evaluating the Model
6. Parameter Tuning
7. Making Predictions



By Shraddha Dalvi

Steps in developing a Machine Learning Application

There are 7 major steps to develop a Machine Learning Application

1. Collecting Data:
2. Preparing the Data
3. Choosing a Model
4. Training the Model
5. Evaluating the Model
6. Parameter Tuning
7. Making Predictions

By Shraddha Dalvi

Steps in developing a Machine Learning Application

1. Collecting Data

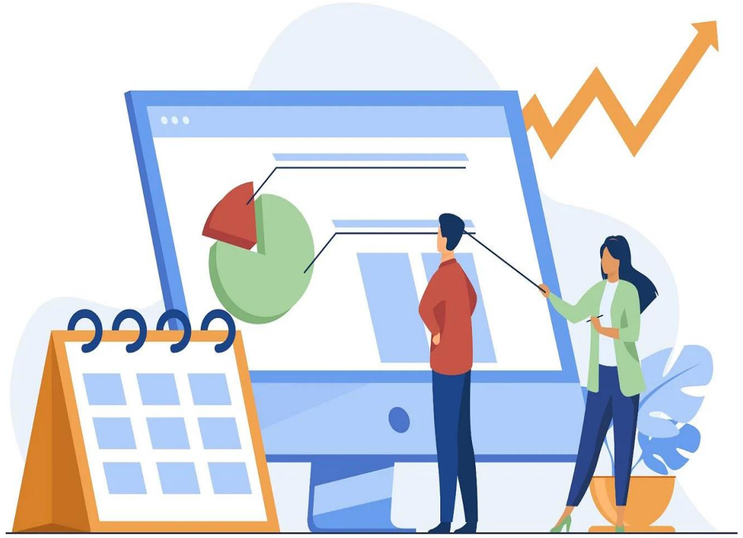
- As you know, machines initially learn from the data that you give them.
- It is of the utmost importance to collect reliable data so that your machine learning model can find the correct patterns.
- The quality of the data that you feed to the machine will determine how accurate your model is.
- If you have incorrect or outdated data, you will have wrong outcomes or predictions which are not relevant.

By Shraddha Dalvi

Steps in developing a Machine Learning Application

2. Preparing the Data

- Putting together all the data you have and randomizing it. This helps make sure that data is evenly distributed, and the ordering does not affect the learning process.
- Cleaning the data to remove unwanted data, missing values, rows, and columns, duplicate values, data type conversion, etc.
- Visualize the data to understand how it is structured and understand the relationship between various variables and classes present.
- Splitting the cleaned data into two sets - a training set and a testing set. The training set is the set your model learns from. A testing set is used to check the accuracy of your model after training.



By Shraddha Dalvi

Steps in developing a Machine Learning Application

3. Choosing a Model

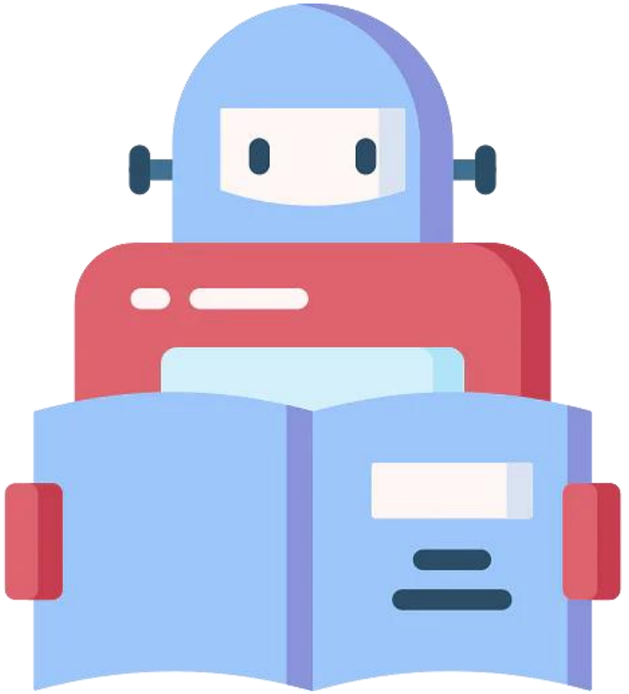
- A machine learning model determines the output you get after running a machine learning algorithm on the collected data.
- It is important to choose a model which is relevant to the task at hand. Over the years, scientists and engineers developed various models suited for different tasks like speech recognition, image recognition, prediction, etc.
- Apart from this, you also have to see if your model is suited for numerical or categorical data and choose accordingly.

By Shraddha Dalvi

Steps in developing a Machine Learning Application

4. Training the Model

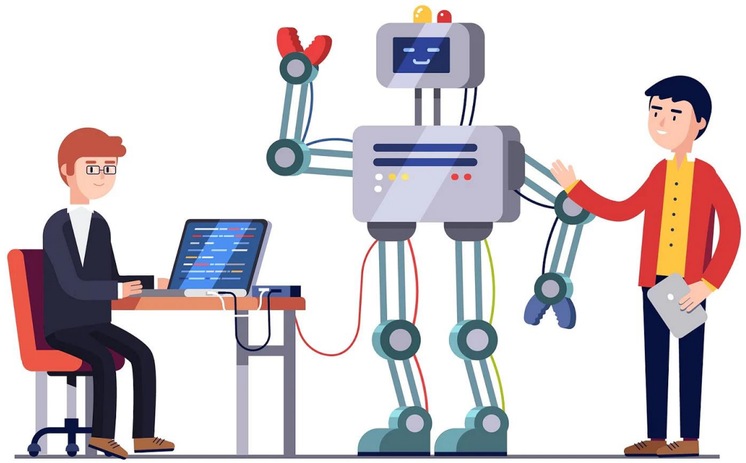
- Training is the most important step in machine learning. In training, you pass the prepared data to your machine learning model to find patterns and make predictions.
- It results in the model learning from the data so that it can accomplish the task set. Over time, with training, the model gets better at predicting.



Steps in developing a Machine Learning Application

5. Evaluating the Model

- After training your model, you have to check to see how it's performing. This is done by testing the performance of the model on previously unseen data.
- The unseen data used is the testing set that you split our data into earlier.
- If testing was done on the same data which is used for training, you will not get an accurate measure, as the model is already used to the data, and finds the same patterns in it, as it previously did.
- This will give you disproportionately high accuracy.
- When used on testing data, you get an accurate measure of how your model will perform and its speed.

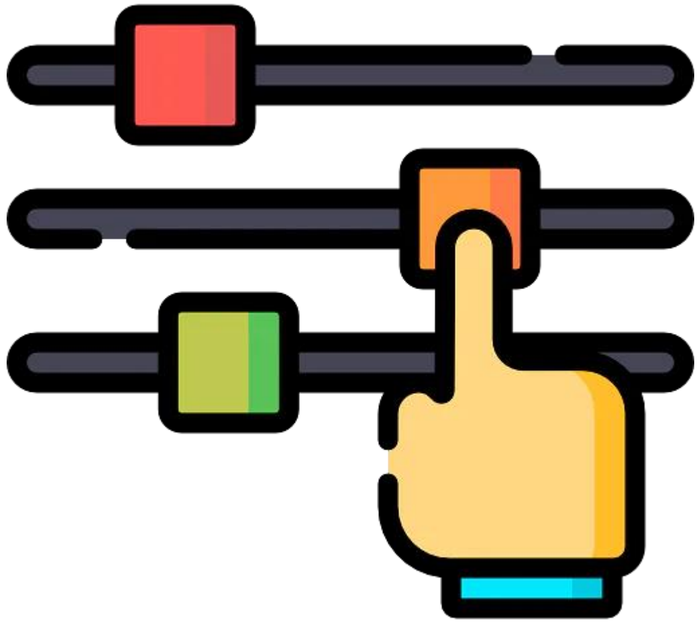


By Shraddha Dalvi

Steps in developing a Machine Learning Application

6. Parameter Tuning

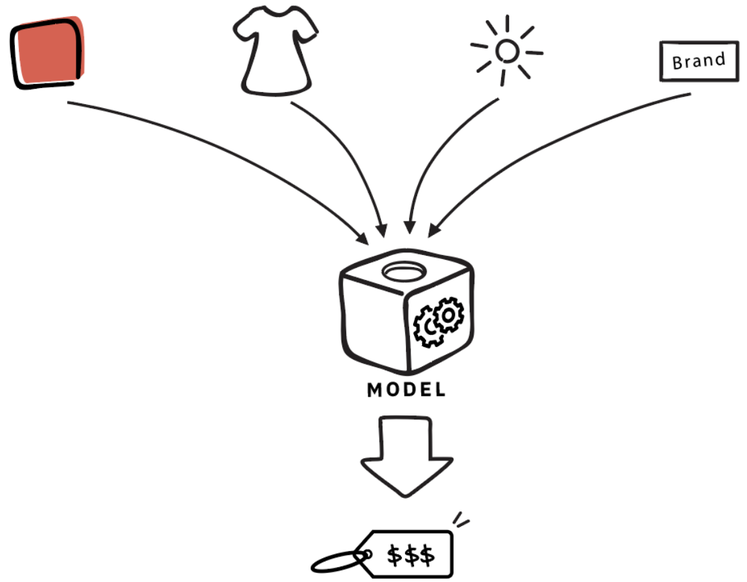
- Once you have created and evaluated your model, see if its accuracy can be improved in any way.
- This is done by tuning the parameters present in your model. Parameters are the variables in the model that the programmer generally decides.
- At a particular value of your parameter, the accuracy will be the maximum. Parameter tuning refers to finding these values.



Steps in developing a Machine Learning Application

7. Making Predictions

- In the end, you can use your model on unseen data to make predictions accurately.



Issues in Machine Learning

Although machine learning is being used in every industry and it helps organizations to make more informed and data-driven choices, it still has so many problems that cannot be ignored.

Here are some common issues in Machine Learning

- 1. Inadequate Training Data
- 2. Poor quality of data
- 3. Non-representative training data
- 4. Overfitting and Underfitting
- 5. Monitoring and maintenance
- 6. Getting bad recommendations
- 7. Lack of skilled resources
- 8. Customer Segmentation
- 9. Process Complexity of Machine Learning
- 10. Data Bias
- 11. Lack of Explainability
- 12. Slow implementations and results
- 13. Irrelevant features

By Shraddha Dalvi

Issues in Machine Learning

1. Inadequate Training Data

- The major issue that comes while using machine learning algorithms is the lack of quality as well as quantity of data. Although data plays a vital role in the processing of machine learning algorithms, many data scientists claim that inadequate data, noisy data, and unclean data are extremely exhausting the machine learning algorithms.
- Data quality can be affected by some factors as follows:
  - ❑ Noisy Data
  - ❑ Incorrect data
  - ❑ Generalizing of output data

By Shraddha Dalvi

Issues in Machine Learning

2. Poor quality of data

- As we have discussed above, data plays a significant role in machine learning, and it must be of good quality as well. Noisy data, incomplete data, inaccurate data, and unclean data lead to less accuracy in classification and low-quality results.
- Hence, data quality can also be considered as a major common problem while processing machine learning algorithms.

Issues in Machine Learning

3. Non-representative training data

- To make sure our training model is generalized well or not, we have to ensure that sample training data must be representative of new cases that we need to generalize. The training data must cover all cases that are already occurred as well as occurring.
- Further, if we are using non-representative training data in the model, it results in less accurate predictions.

Issues in Machine Learning

4. Overfitting and Underfitting

The major issue that comes while using machine learning algorithms is the lack of quality as well as quantity of data.

Issues in Machine Learning

5. Monitoring and maintenance

- As we know that generalized output data is mandatory for any machine learning model; hence, regular monitoring and maintenance become compulsory for the same.
- Different results for different actions require data change; hence editing of codes as well as resources for monitoring them also become necessary.

Issues in Machine Learning

6. Getting bad recommendations

- A machine learning model operates under a specific context which results in bad recommendations and concept drift in the model.
- Let's understand with an example where at a specific time customer is looking for some gadgets, but now customer requirement changed over time but still machine learning model showing same recommendations to the customer while customer expectation has been changed.
- This incident is called a Data Drift. It generally occurs when new data is introduced or interpretation of data changes. However, we can overcome this by regularly updating and monitoring data according to the expectations.

By Shraddha Dalvi

Issues in Machine Learning

7. Lack of skilled resources

- Although Machine Learning and Artificial Intelligence are continuously growing in the market, still these industries are fresher in comparison to others.
- The absence of skilled resources in the form of manpower is also an issue.
- Hence, we need manpower having in-depth knowledge of mathematics, science, and technologies for developing and managing scientific substances for machine learning.

Issues in Machine Learning

8. Customer Segmentation

- Customer segmentation is also an important issue while developing a machine learning algorithm.
- To identify the customers who paid for the recommendations shown by the model and who don't even check them.
- Hence, an algorithm is necessary to recognize the customer behavior and trigger a relevant recommendation for the user based on past experience.

Issues in Machine Learning

9. Process Complexity of Machine Learning

- The machine learning process is very complex, which is also another major issue faced by machine learning engineers and data scientists.
- However, Machine Learning and Artificial Intelligence are very new technologies but are still in an experimental phase and continuously being changing over time.
- There is the majority of hits and trial experiments; hence the probability of error is higher than expected.
- Further, it also includes analyzing the data, removing data bias, training data, applying complex mathematical calculations, etc., making the procedure more complicated and quite tedious.

Issues in Machine Learning

10. Data Bias

Data Biasing is also found a big challenge in Machine Learning. These errors exist when certain elements of the dataset are heavily weighted or need more importance than others. Biased data leads to inaccurate results, skewed outcomes, and other analytical errors.

By Shraddha Dalvi

Issues in Machine Learning

11. Lack of Explainability

This basically means the outputs cannot be easily comprehended as it is programmed in specific ways to deliver for certain conditions. Hence, a lack of explainability is also found in machine learning algorithms which reduce the credibility of the algorithms.

By Shraddha Dalvi

Issues in Machine Learning

12. Slow implementations and results

- This issue is also very commonly seen in machine learning models. However, machine learning models are highly efficient in producing accurate results but are time-consuming.
- Slow programming, excessive requirements' and overloaded data take more time to provide accurate results than expected.
- This needs continuous maintenance and monitoring of the model for delivering accurate results.

Issues in Machine Learning

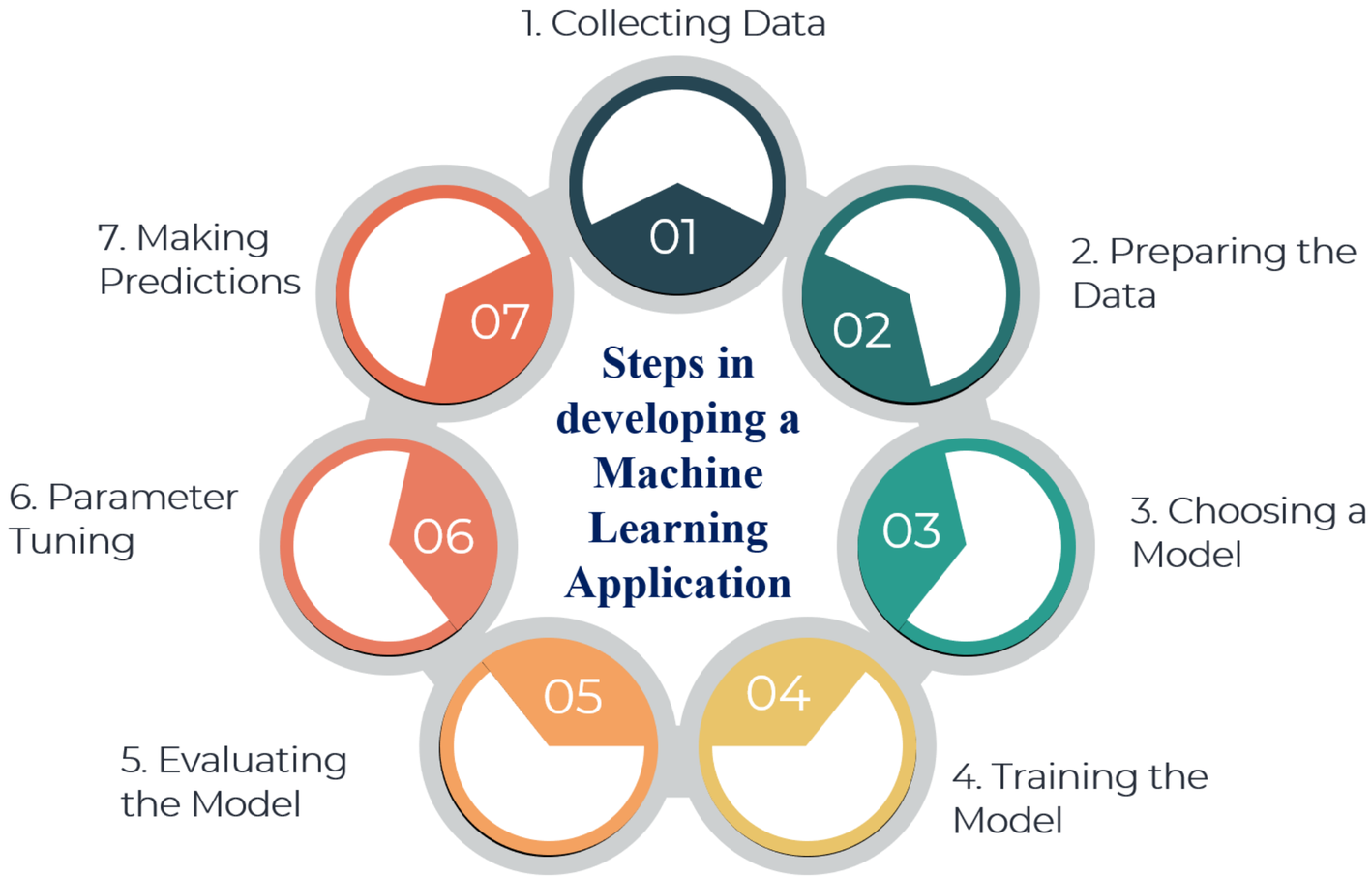
13. Irrelevant features

- Although machine learning models are intended to give the best possible outcome, if we feed garbage data as input, then the result will also be garbage.
- Hence, we should use relevant features in our training sample. A machine learning model is said to be good if training data has a good set of features or less to no irrelevant features.

Steps in developing a Machine Learning Application

There are 7 major steps to develop a Machine Learning Application

1. Collecting Data:
2. Preparing the Data
3. Choosing a Model
4. Training the Model
5. Evaluating the Model
6. Parameter Tuning
7. Making Predictions



By Shraddha Dalvi

Steps in developing a Machine Learning Application

There are 7 major steps to develop a Machine Learning Application

1. Collecting Data:
2. Preparing the Data
3. Choosing a Model
4. Training the Model
5. Evaluating the Model
6. Parameter Tuning
7. Making Predictions

By Shraddha Dalvi

Issues in Machine Learning

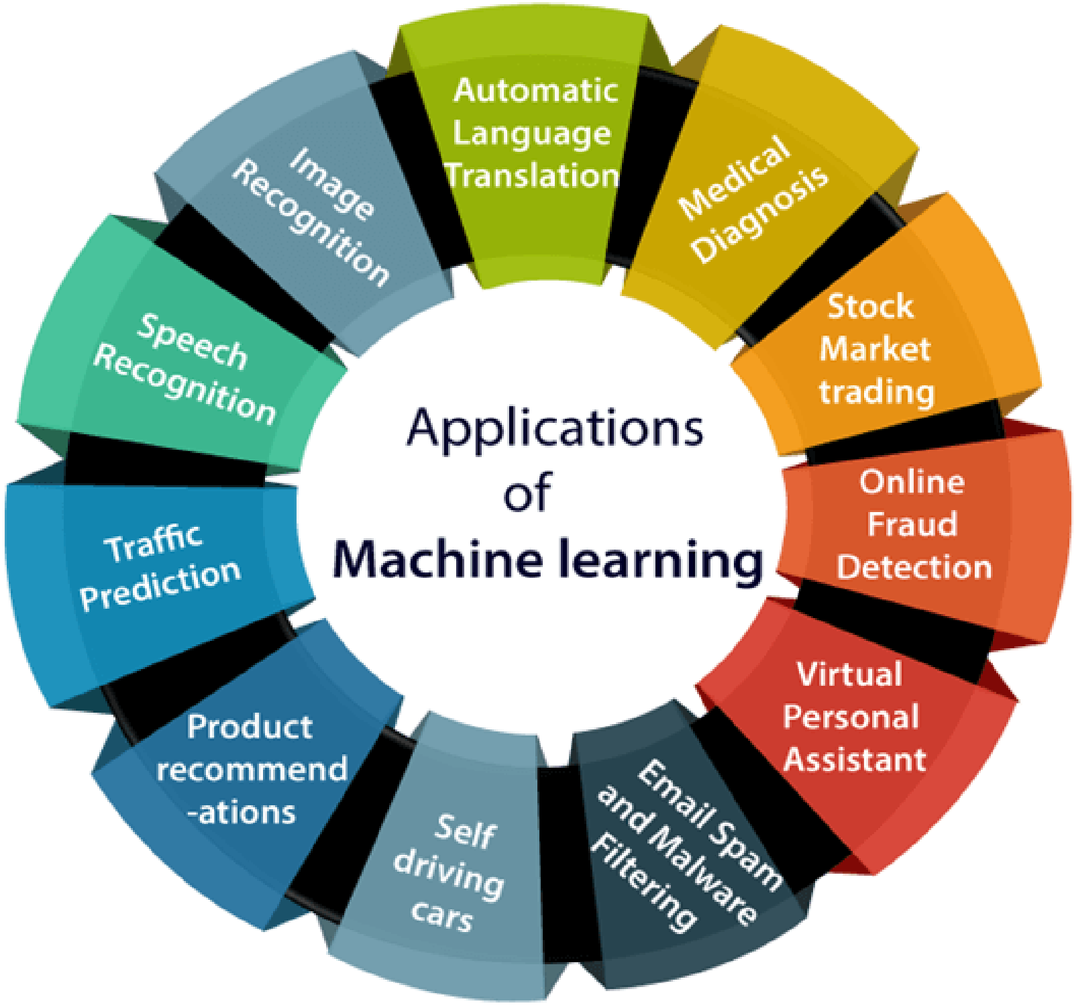
Although machine learning is being used in every industry and it helps organizations to make more informed and data-driven choices, it still has so many problems that cannot be ignored.

Here are some common issues in Machine Learning

- 1. Inadequate Training Data
- 2. Poor quality of data
- 3. Non-representative training data
- 4. Overfitting and Underfitting
- 5. Monitoring and maintenance
- 6. Getting bad recommendations
- 7. Lack of skilled resources
- 8. Customer Segmentation
- 9. Process Complexity of Machine Learning
- 10. Data Bias
- 11. Lack of Explainability
- 12. Slow implementations and results
- 13. Irrelevant features

By Shraddha Dalvi

Applications of Machine learning

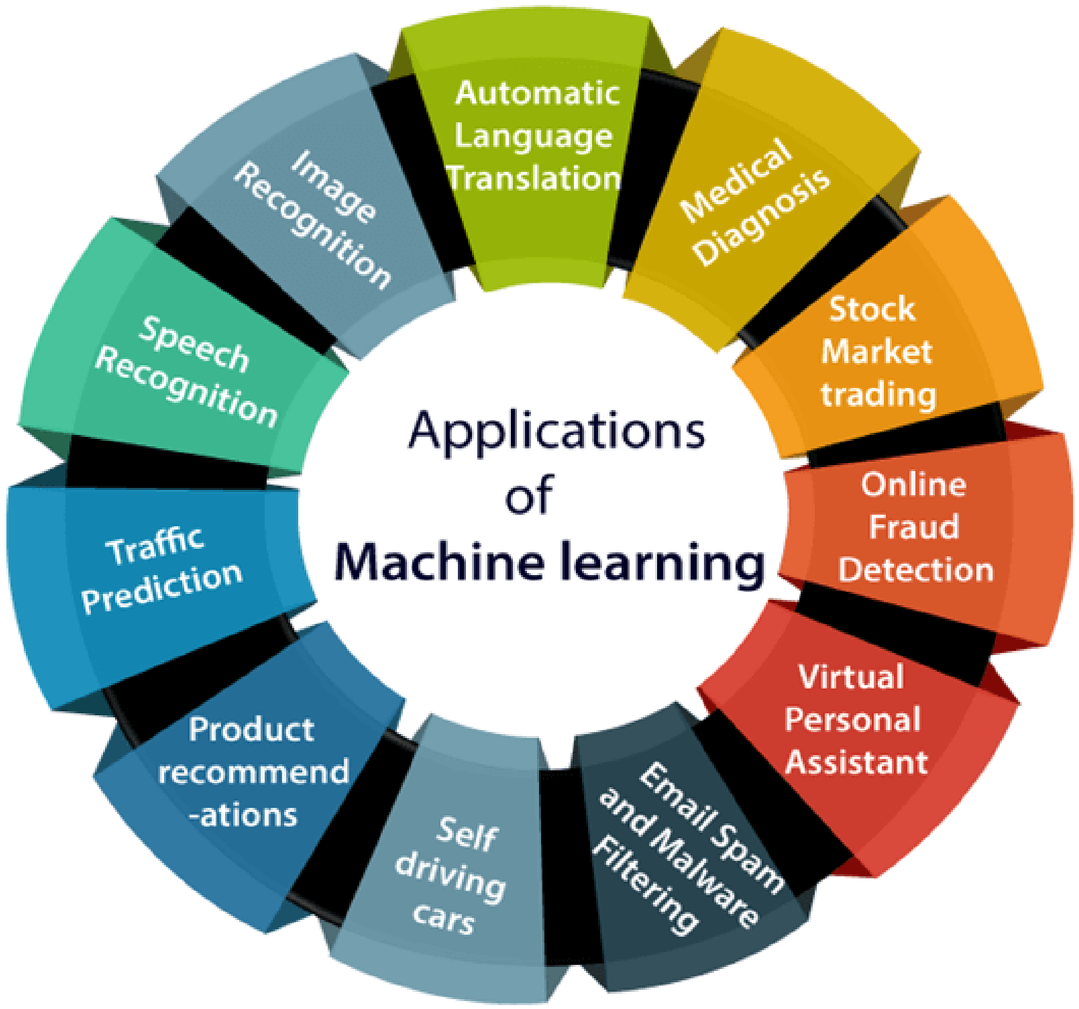


- 1. Image Recognition
- 2. Speech Recognition
- 3. Traffic prediction
- 4. Product recommendations
- 5. Self-driving cars
- 6. Email Spam and Malware Filtering
- 7. Virtual Personal Assistant
- 8. Online Fraud Detection
- 9. Stock Market trading
- 10. Medical Diagnosis
- 11. Automatic Language Translation

By Shraddha Dalvi

Applications of Machine learning

- Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc.
- Below are some most trending real-world applications of Machine Learning:



By Shraddha Dalvi

Applications of Machine learning

1. Image Recognition

- Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, Automatic friend tagging suggestion:
- Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's face detection and recognition algorithm.
- It is based on the Facebook project named "Deep Face," which is responsible for face recognition and person identification in the picture.

By Shraddha Dalvi

Applications of Machine learning

2. Speech Recognition

- While using Google, we get an option of "Search by voice," it comes under speech recognition, and it's a popular application of machine learning.
- Speech recognition is a process of converting voice instructions into text, and it is also known as "Speech to text", or "Computer speech recognition." At present, machine learning algorithms are widely used by various applications of speech recognition. Google assistant, Siri, Cortana, and Alexa are using speech recognition technology to follow the voice instructions.

Applications of Machine learning

3. Traffic prediction

- If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.
- It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:
  - ❑ Real Time location of the vehicle form Google Map app and sensors
  - ❑ Average time has taken on past days at the same time.
- Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

Applications of Machine learning

4. Product recommendations

- Machine learning is widely used by various e-commerce and entertainment companies such as Amazon, Netflix, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.
- Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.
- As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

Applications of Machine learning

5. Self-driving cars

- One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car.
- It is using unsupervised learning method to train the car models to detect people and objects while driving.

Applications of Machine learning

6. Email Spam and Malware Filtering

- Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:
    - Content Filter
    - Header filter
    - General blacklists filter
    - Rules-based filters
    - Permission filters
- Some machine learning algorithms such as Multi-Layer Perceptron, Decision tree, and Naïve Bayes classifier are used for email spam filtering and malware detection.

Applications of Machine learning

7. Virtual Personal Assistant

- We have various virtual personal assistants such as Google assistant, Alexa, Cortana, Siri. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.
- These virtual assistants use machine learning algorithms as an important part.
- These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

Applications of Machine learning

8. Online Fraud Detection

- Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as fake accounts, fake ids, and steal money in the middle of a transaction. So to detect this, Feed Forward Neural network helps us by checking whether it is a genuine transaction or a fraud transaction.
- For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

Applications of Machine learning

9. Stock Market trading:

- Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's long short term memory neural network is used for the prediction of stock market trends.

Applications of Machine learning

10. Medical Diagnosis:

- In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.
- It helps in finding brain tumors and other brain-related diseases easily.

Applications of Machine learning

11. Automatic Language Translation:

- Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages. Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.
- The technology behind the automatic translation is a sequence to sequence learning algorithm, which is used with image recognition and translates the text from one language to another language.