Introduction to Machine Learning

Prof. Taher Hamza & Sara El-Metwally
Computer Science Department,
Faculty of Computers and Information,
Mansoura University, Egypt.

sarah_almetwally4@mans.edu.eg

Agenda

01 Needs for Machine Learning

02 What is Machine Learning

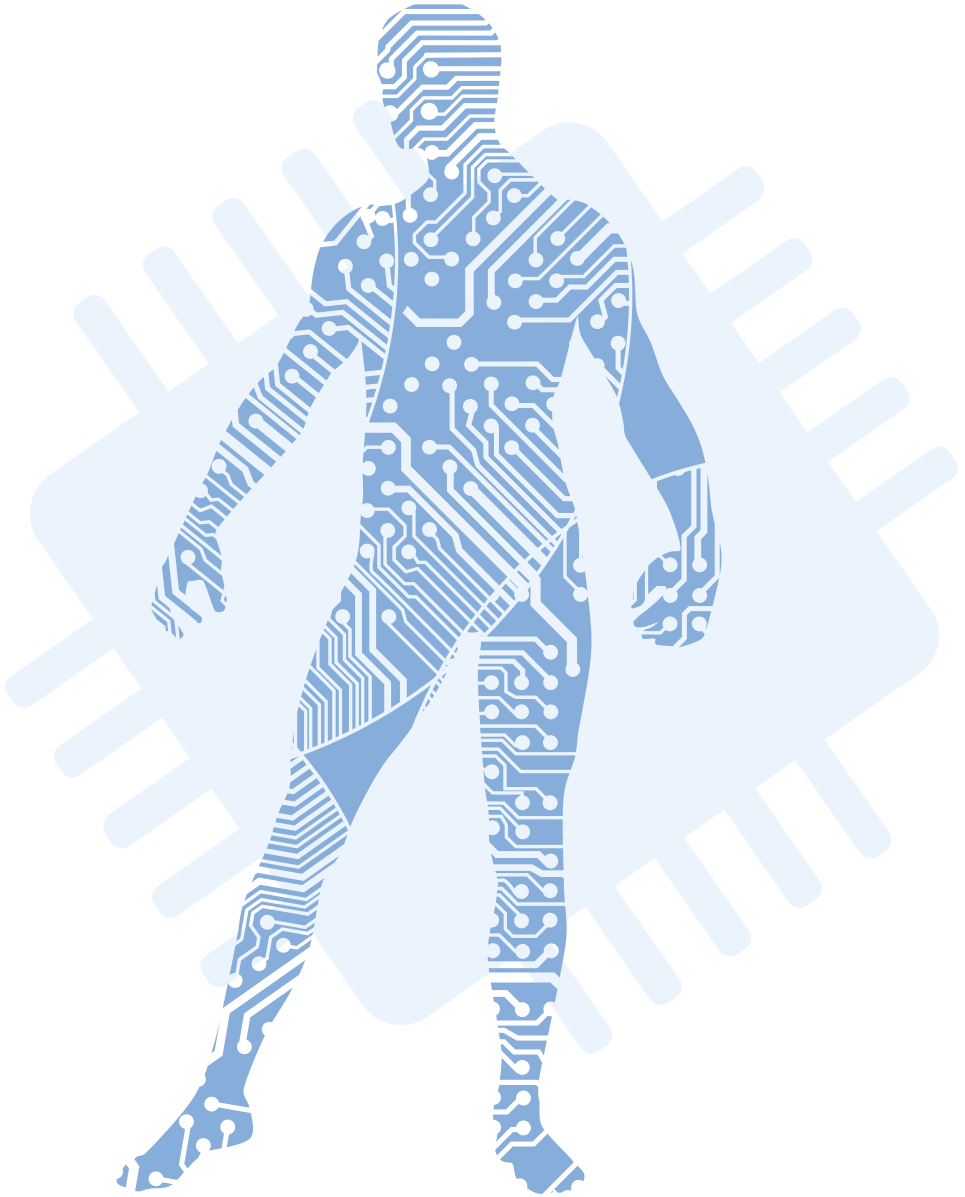
03 Machine Learning Definitions

04 Machine Learning Process

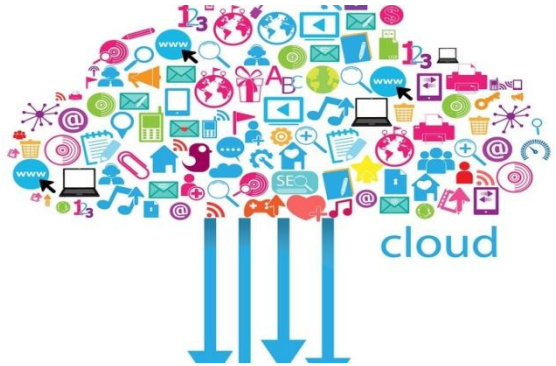
05 Types of Machine Learning

06 Linear Regression

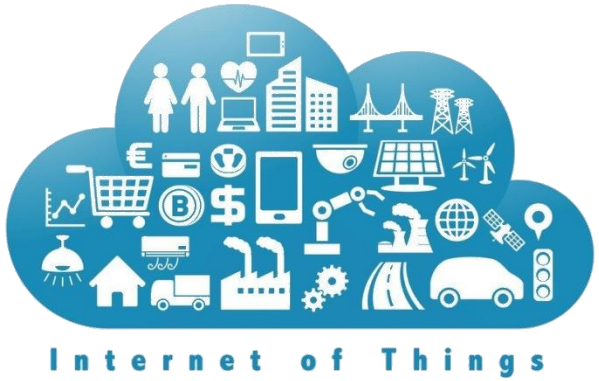
07 Simple Demo



Needs for Machine Learning



Cloud Data



IOT Data

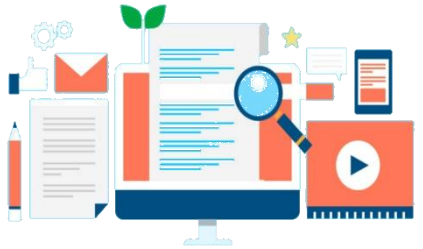
Social Media Data

- Over 2.5 Quintillion bytes of data is generated every single day.
- By 2020, it is estimated that every person on earth can create 1.7 MB on every second.

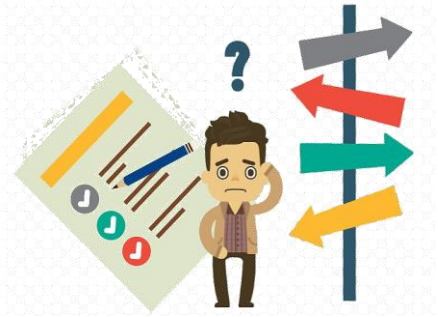
Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Needs for Machine Learning

Increase in Data Generation



Uncover Patterns and Trends in Data

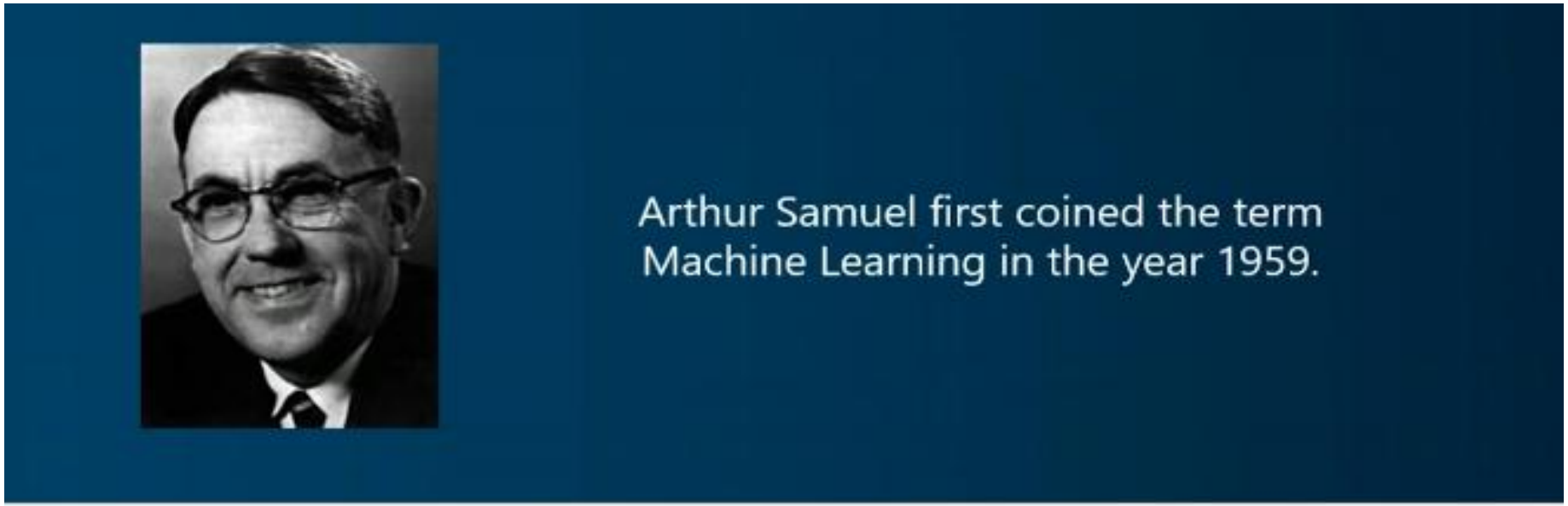


Improve Decision Making

Solve Complex Problems

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

What is Machine Learning

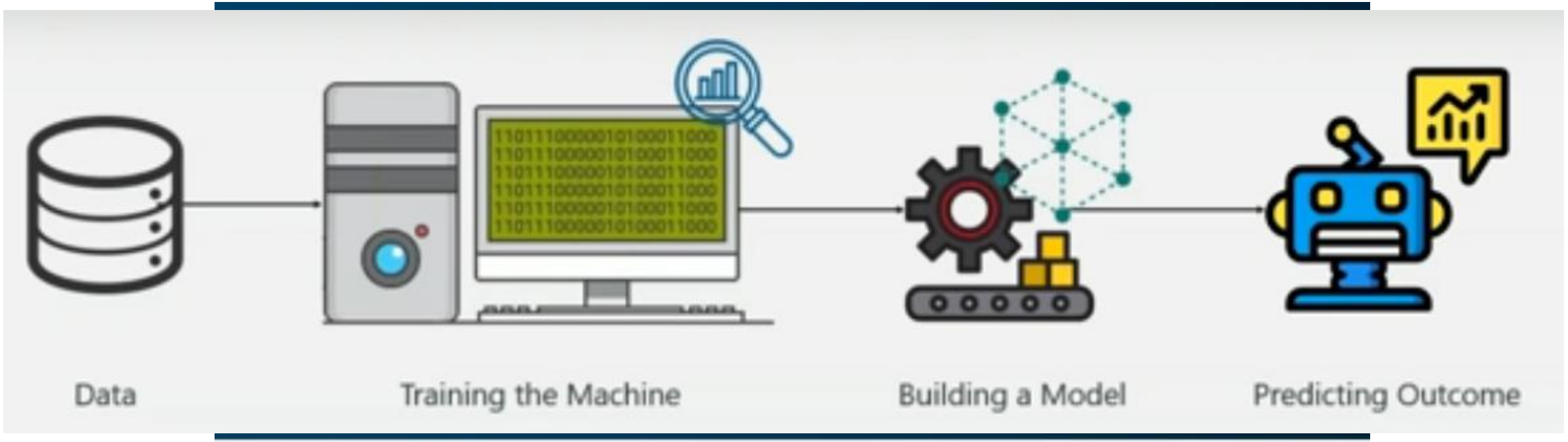


What Is Machine Learning?

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

What is Machine Learning



A simpler definition of Machine Learning

Machine learning is a subset of Artificial Intelligence (AI) which provides machines the ability to learn automatically & improve from experience without being explicitly programmed.

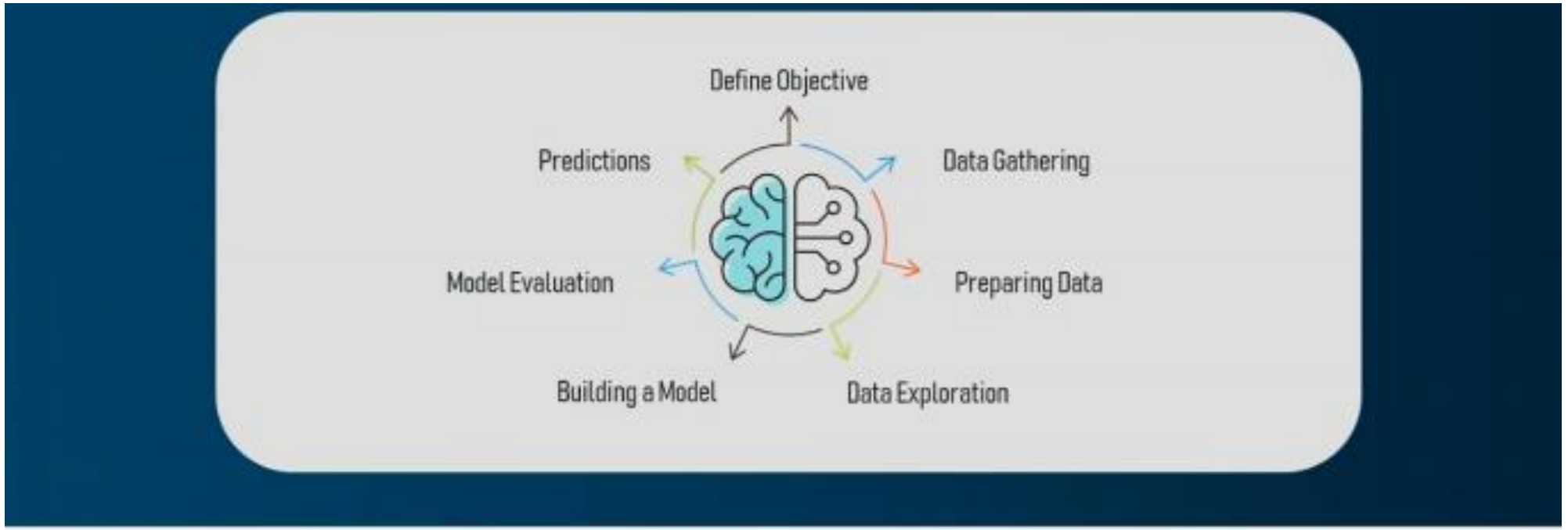
Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Definitions

Concept	Definition
Algorithm	Set of rules and statistical techniques used to learn patterns from the data (draw insights from data).
Model	A model in a computer is a mathematical function that represents a relationship or mapping between a set of inputs and a set of outputs.
Predictor Variable (Inputs)	It is features of the data that can be used to predict the output.
Response Variable (Outputs)	It is output variable that can predict by a machine learning model.
Training Data	A model is build using a set of training data.
Testing Data	A model is evaluated using a set of testing data.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process



MACHINE LEARNING PROCESS

*The Machine Learning process involves building a **Predictive model** that can be used to find a **solution** for a **Problem Statement.***

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process

Step 1: Define the objective of the Problem

To predict the possibility of rain by studying the weather conditions.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

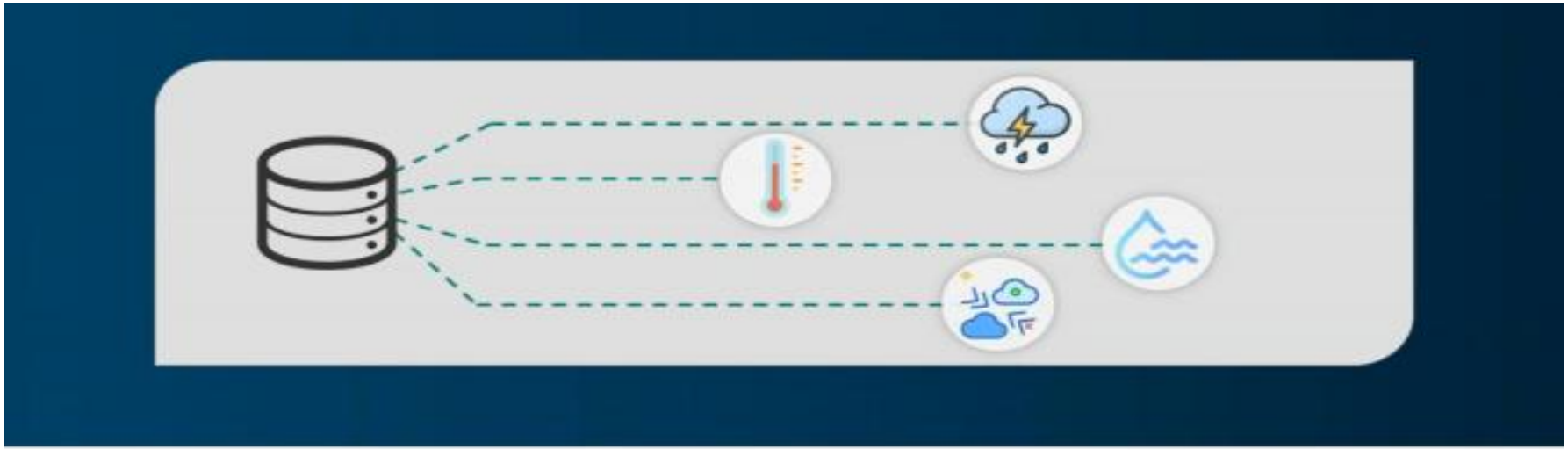
Machine Learning Process

Step 1: Define the objective of the Problem

To predict the possibility of rain by studying the weather conditions.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process

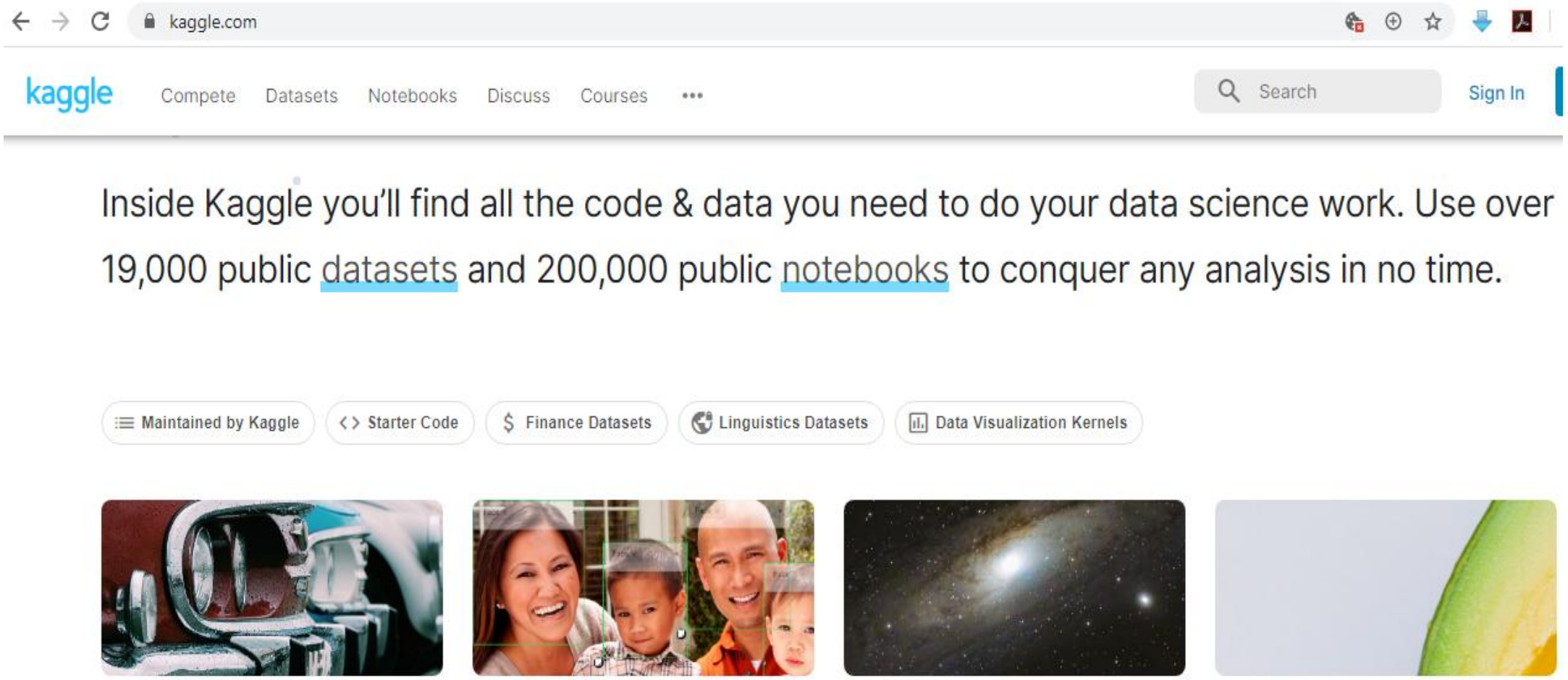


Step 2: Data Gathering

Data such as weather conditions, humidity level, temperature, pressure, etc are either collected manually or scarped from the web.

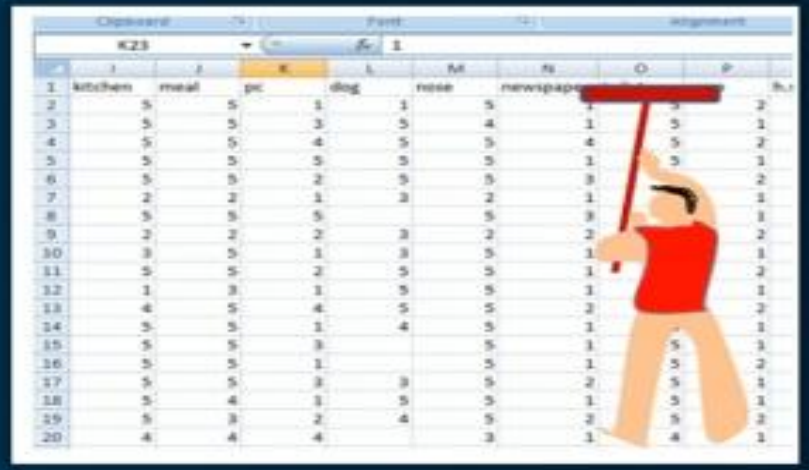
Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process



Machine Learning Process

- Transform data into desired format
- Data cleaning
  - Missing values
  - Corrupted data
  - Remove unnecessary data



Step 3: Preparing Data

Data Cleaning involves getting rid of inconsistencies in data such as missing values or redundant variables.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process

Step 4: Exploratory Data Analysis

Data Exploration involves understanding the patterns and trends in the data. At this stage all the useful insights are drawn and correlations between the variables are understood.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process

- Machine Learning model is built by using the training data set
- The model is the Machine Learning algorithm that predicts the output by using the data fed to it

Step 5: Building a Machine Learning Model

At this stage a Predictive Model is built by using Machine Learning Algorithms such as Linear Regression, Decision Trees, etc.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process

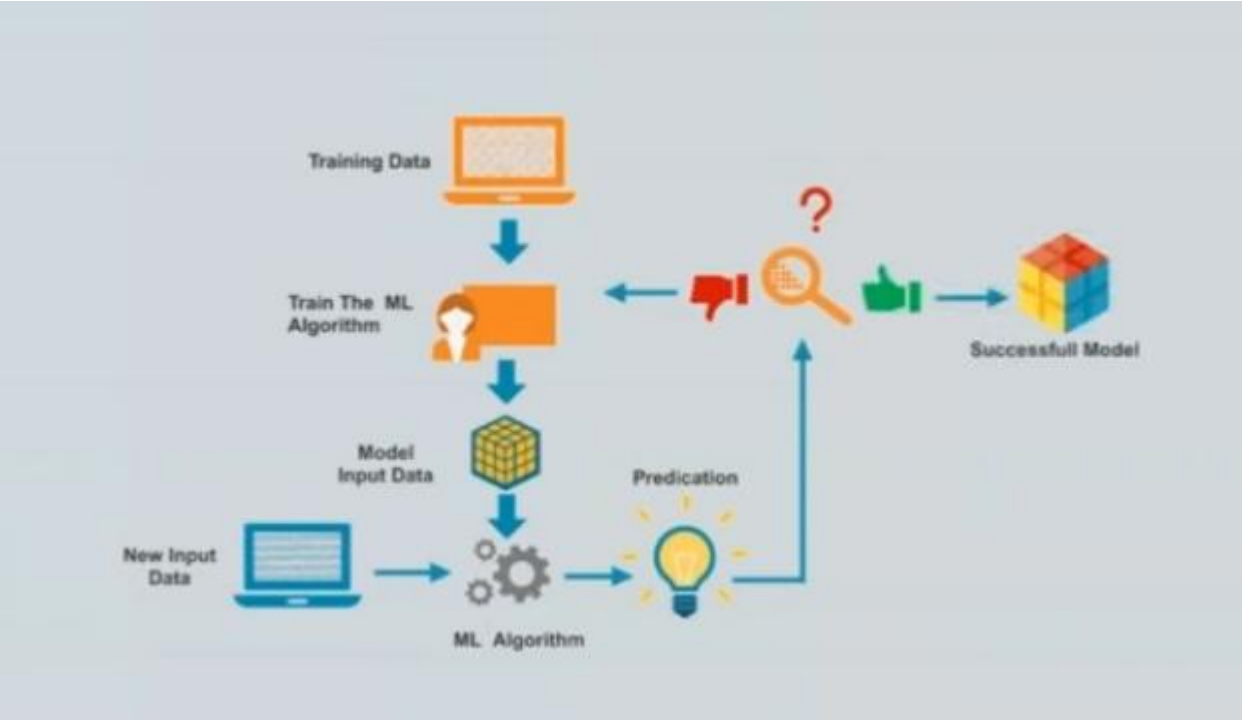
- Machine Learning model is evaluated by using the testing data set
- The accuracy of the model is calculated
- Further improvement in the model are done by using techniques like Parameter tuning

Step 6: Model Evaluation & Optimization

The efficiency of the model is evaluated and any further improvement in the model are implemented.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Process

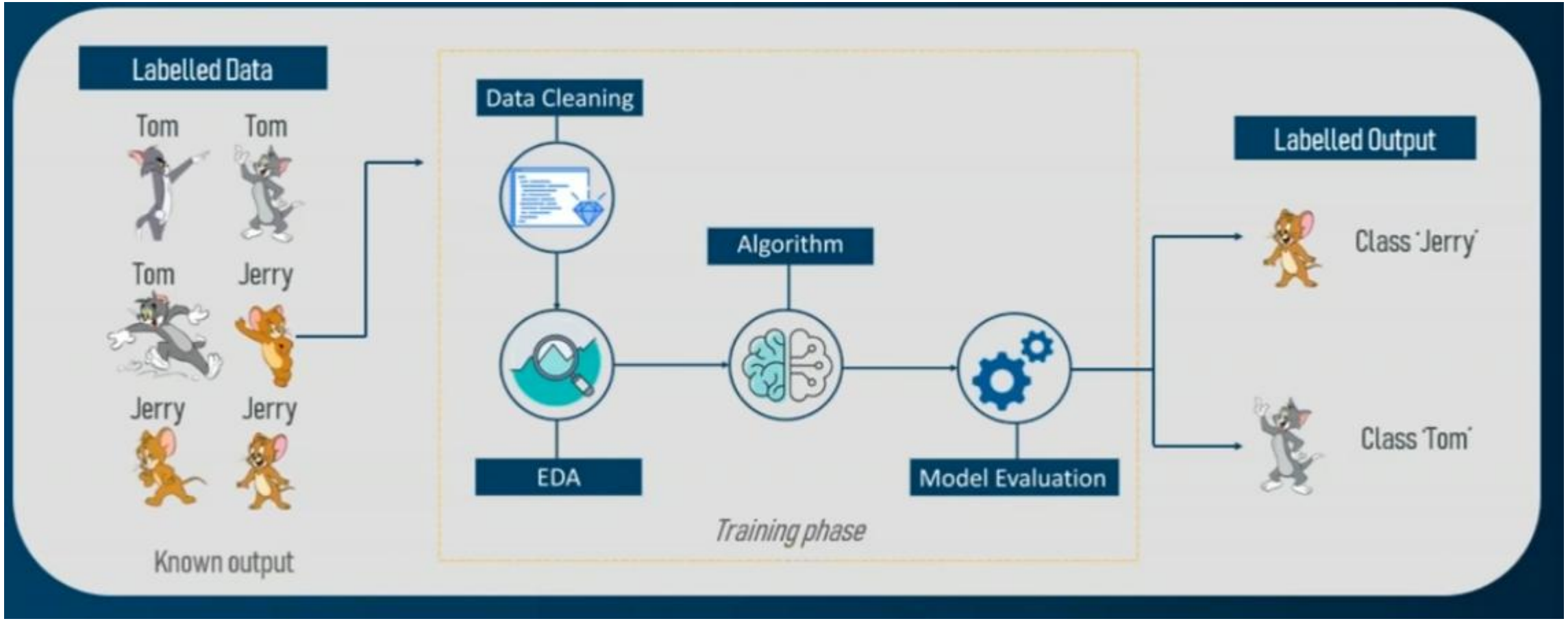


Step 7: Predictions

The final outcome is predicted after performing parameter tuning and improving the accuracy of the model.

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

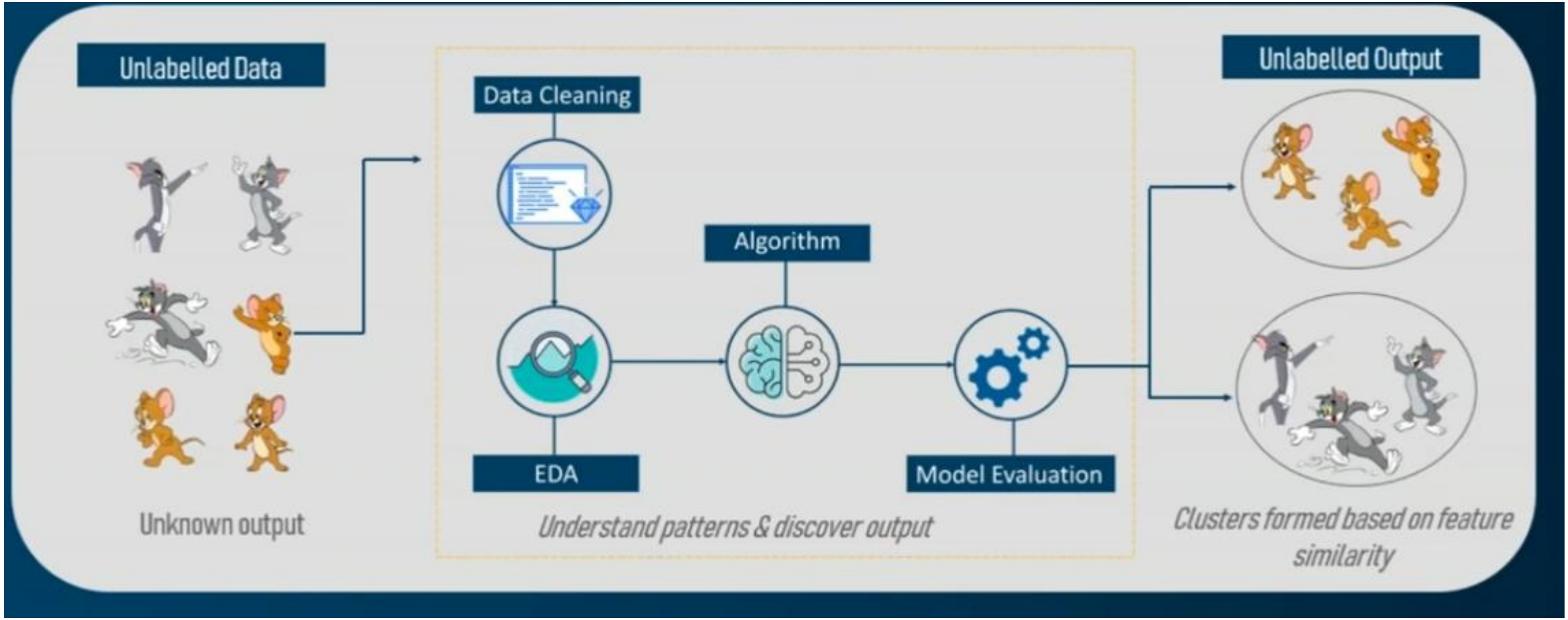
Machine Learning Types



Supervised Learning

Supervised learning is a technique in which we teach or train the machine using data which is well labelled.

Machine Learning Types



Unsupervised Learning

Unsupervised learning is the training of machine using information that is unlabeled and allowing the algorithm to act on that information without guidance.

Machine Learning Types

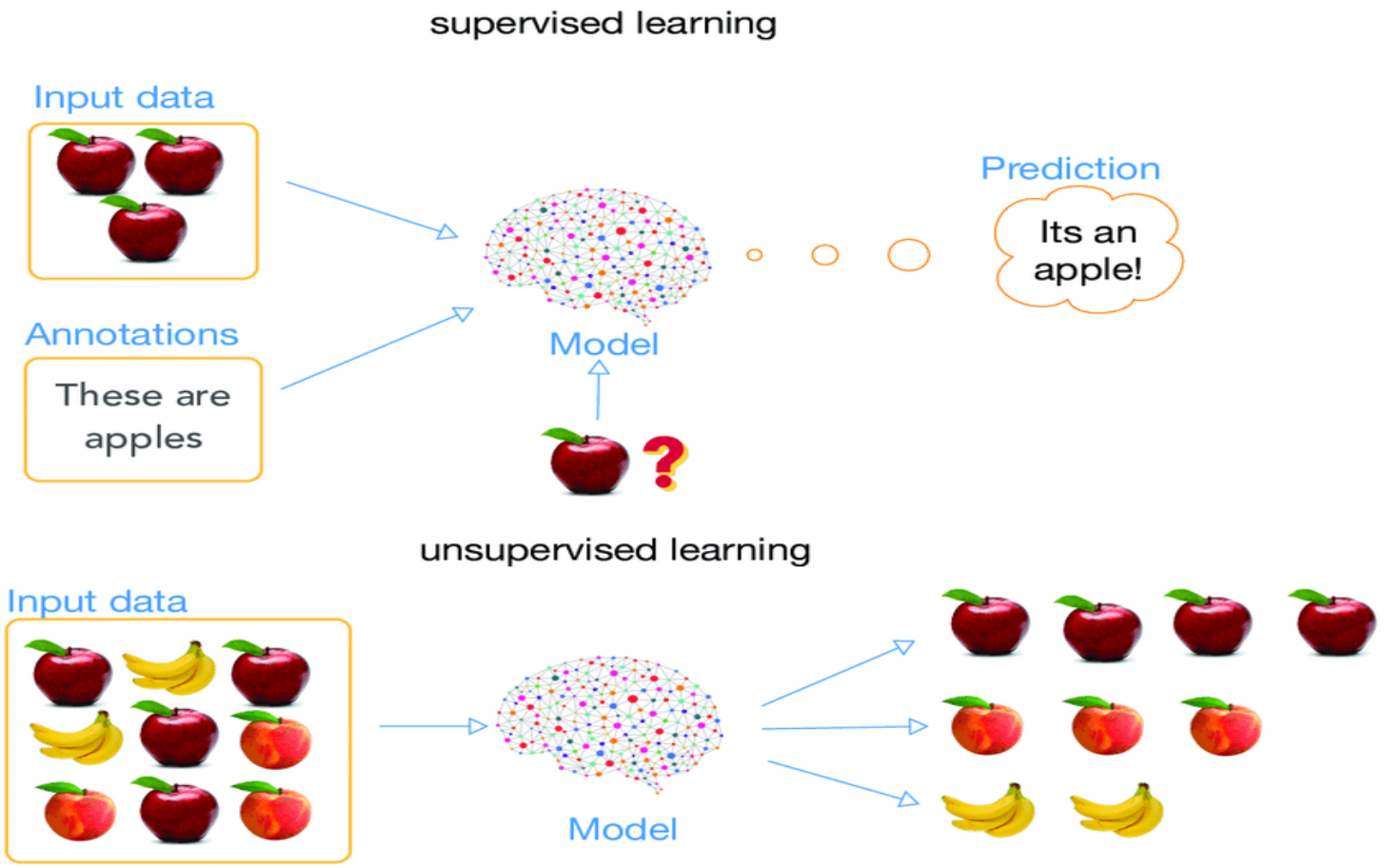
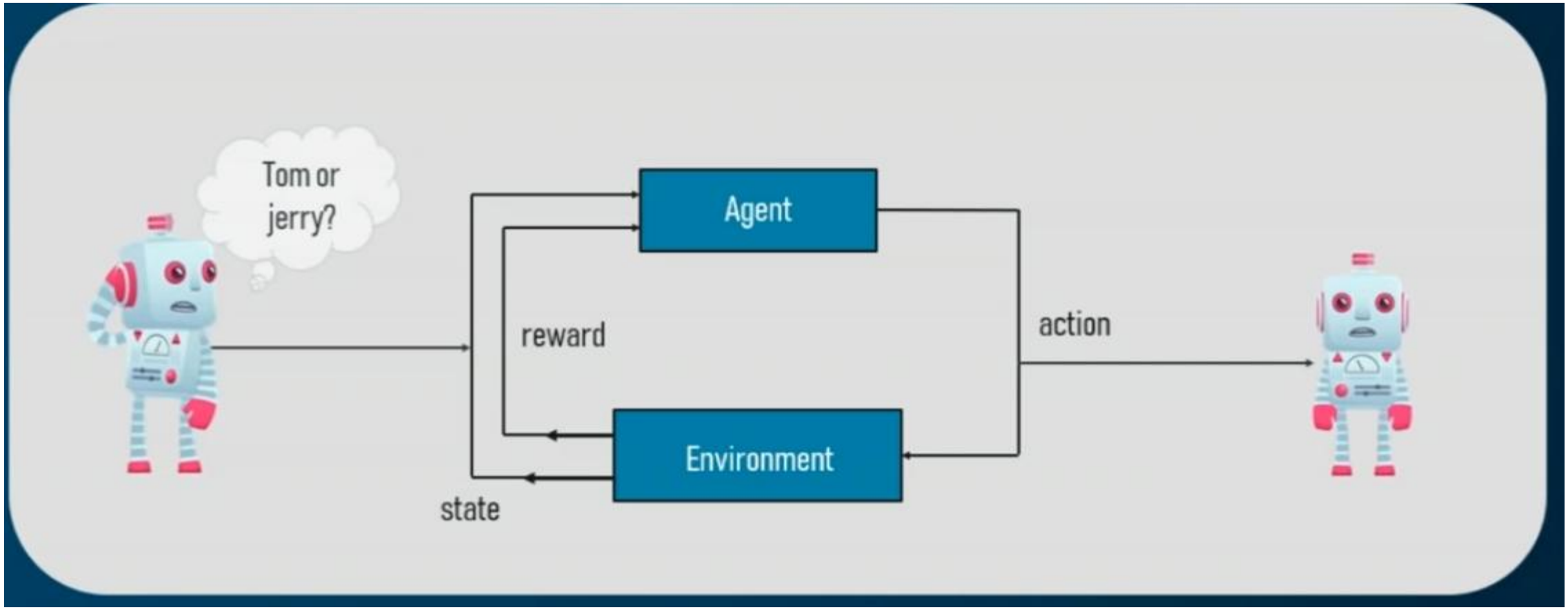


Photo credits: Yan et al., Background Augmentation Generative Adversarial Networks (BAGANs): Effective Data Generation Based on GAN-Augmented 3D Synthesizing, Symmetry, 2018.

Machine Learning Types



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.

Machine Learning Types

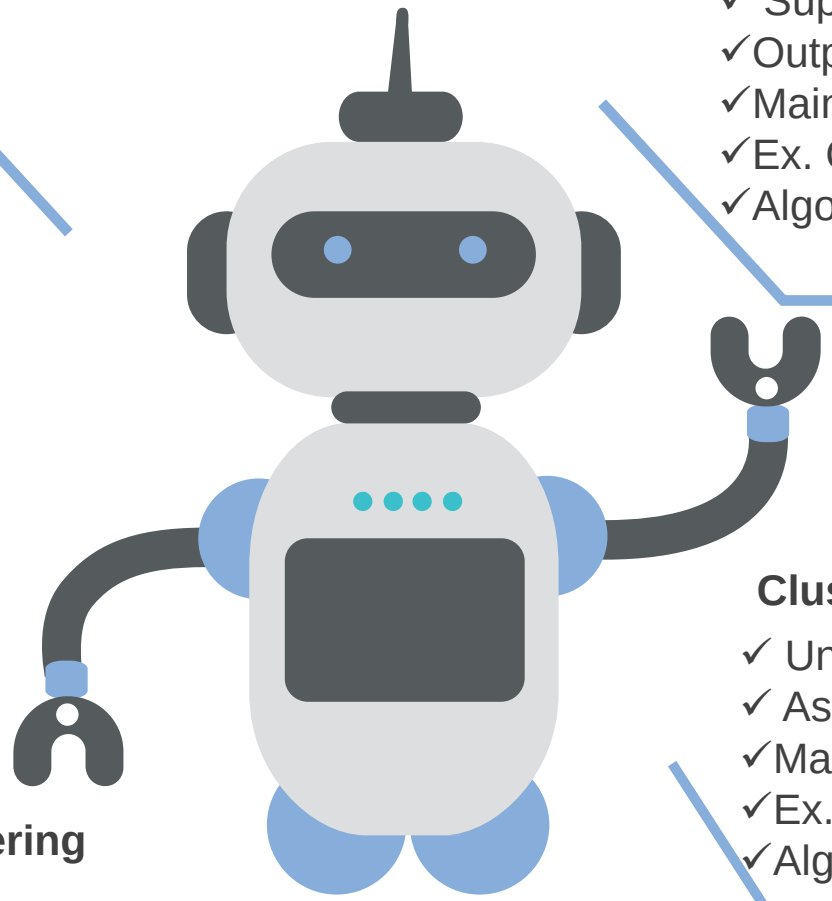
	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	Machine learns by using labeled data	Machine is trained using unlabeled data without any guidance.	Agent interacts with its environments by producing actions and discovers errors or rewards.
Type of Problems	Regressions and Classifications	Associations and Clustering	Reward based
Type of Data	Labeled data	Unlabeled data	No pre-defined data
Training	External Supervision	No supervision	No supervision
Approach	Map labeled input to know output	Understand patterns and discover output	Follow trial and error method.
Popular Algorithms	Linear regression Logistic regression SVM, KNN	K-means C-means	Q-learning

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Types

Regression

- ✓ Supervised Machine Learning.
- ✓Output is a continuous quantity.
- ✓Main goal is forecast /predict.
- ✓Ex. Predict Stock Market Price.
- ✓Algorithms like Linear Regression



Classification

- ✓ Supervised Machine Learning.
- ✓Output is a categorical quantity.
- ✓Main goal is compute category (class) of data.
- ✓Ex. Classify Emails as spam or not spam.
- ✓Algorithms like Logistic Regression

Clustering

- ✓ Unsupervised Machine Learning.
- ✓ Assigns data points into clusters.
- ✓Main goal is group similar items together
- ✓Ex. Clusters Movies, Users, etc.
- ✓Algorithms like K-means

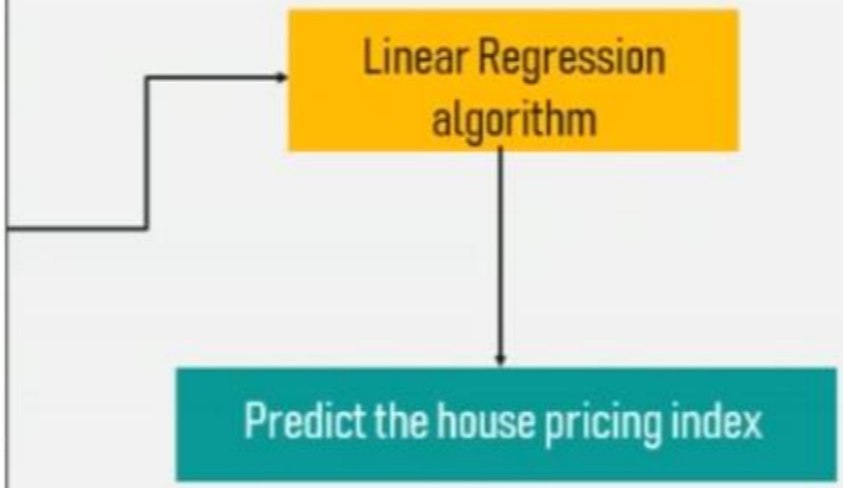
Regression Vs. Classification Vs. Clustering

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Types

Problem Statement: To study the House Sales dataset and build a Machine Learning model that predicts the house pricing index.

```
> str(data)
'data.frame':   21613 obs. of  21 variables:
 $ id            : num  7.13e+09 6.41e+09 5.63e+09 2.49e+09 1.95e+09 ...
 $ date          : Factor w/ 372 levels "20140502T000000",..: 165 221 291 221 284 11 57
 $ price         : num  221900 538000 180000 604000 510000 ...
 $ bedrooms      : int  3 3 2 4 3 4 3 3 3 3 ...
 $ bathrooms     : num  1 2.25 1 3 2 4.5 2.25 1.5 1 2.5 ...
 $ sqft_living   : int  1180 2570 770 1960 1680 5420 1715 1060 1780 1890 ...
 $ sqft_lot      : int  5650 7242 10000 5000 8080 101930 6819 9711 7470 6560 ...
 $ floors        : num  1 2 1 1 1 1 2 1 1 2 ...
 $ waterfront    : int  0 0 0 0 0 0 0 0 0 0 ...
 $ view          : int  0 0 0 0 0 0 0 0 0 0 ...
 $ condition     : int  3 3 3 5 3 3 3 3 3 3 ...
 $ grade         : int  7 7 6 7 8 11 7 7 7 7 ...
 $ sqft_above    : int  1180 2170 770 1050 1680 3890 1715 1060 1050 1890 ...
 $ sqft_basement: int  0 400 0 910 0 1530 0 0 730 0 ...
 $ yr_built      : int  1955 1951 1933 1965 1987 2001 1995 1963 1960 2003 ...
 $ yr_renovated : int  0 1991 0 0 0 0 0 0 0 0 ...
 $ zipcode       : int  98178 98125 98028 98136 98074 98053 98003 98198 98146 98038 ...
 $ lat           : num  47.5 47.7 47.7 47.5 47.6 ...
 $ long          : num  -122 -122 -122 -122 -122 ...
 $ sqft_living15: int  1340 1690 2720 1360 1800 4760 2238 1650 1780 2390 ...
 $ sqft_lot15    : int  5650 7639 8062 5000 7503 101930 6819 9711 8113 7570 ...
```

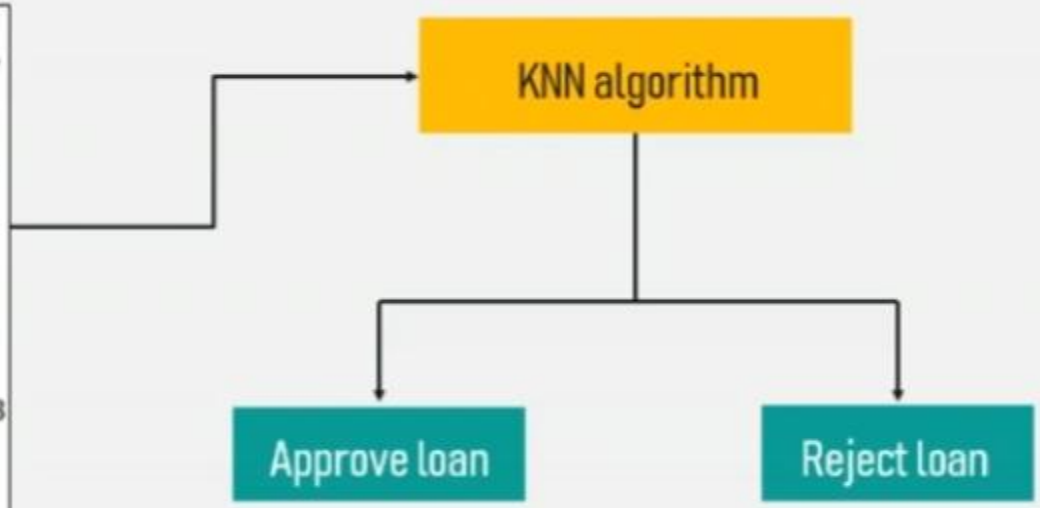


Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Types

Problem Statement: Study a bank credit dataset and make a decision about whether to approve the loan of an applicant based on his profile

```
$ Account.Balance                  : int  1 1 2 1 1 1 1 1 4 2 ...
$ Duration.of.Credit..month.       : int  18 9 12 12 12 10 8 6 18 24 ..
$ Payment.Status.of.Previous.Credit: int  4 4 2 4 4 4 4 4 4 2 ...
$ Purpose                          : int  2 0 9 0 0 0 0 0 3 3 ...
$ Credit.Amount                    : int  1049 2799 841 2122 2171 2241
$ Value.Savings.Stocks             : int  1 1 2 1 1 1 1 1 1 3 ...
$ Length.of.current.employment     : int  2 3 4 3 3 2 4 2 1 1 ...
$ Instalment.per.cent              : int  4 2 2 3 4 1 1 2 4 1 ...
$ Sex...Marital.Status             : int  2 3 2 3 3 3 3 3 2 2 ...
$ Guarantors                       : int  1 1 1 1 1 1 1 1 1 1 ...
$ Duration.in.Current.address      : int  4 2 4 2 4 3 4 4 4 4 ...
$ Most.valuable.available.asset    : int  2 1 1 1 2 1 1 1 3 4 ...
$ Age..years.                      : int  21 36 23 39 38 48 39 40 65 23
$ Concurrent.Credits               : int  3 3 3 3 1 3 3 3 3 3 ...
$ Type.of.apartment                : int  1 1 1 1 2 1 2 2 2 1 ...
$ No.of.Credits.at.this.Bank       : int  1 2 1 2 2 2 2 1 2 1 ...
$ Occupation                       : int  3 3 2 2 2 2 2 2 1 1 ...
$ No.of.dependents                 : int  1 2 1 2 1 2 1 2 1 1 ...
$ Telephone                        : int  1 1 1 1 1 1 1 1 1 1 ...
$ Foreign.Worker                   : int  1 1 1 2 2 2 2 2 1 1 ...
```

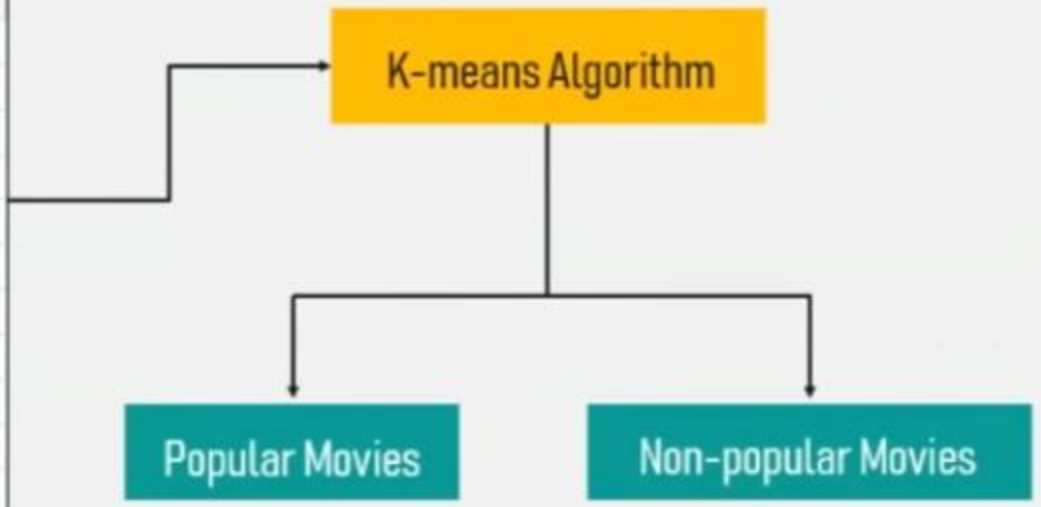


Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Machine Learning Types

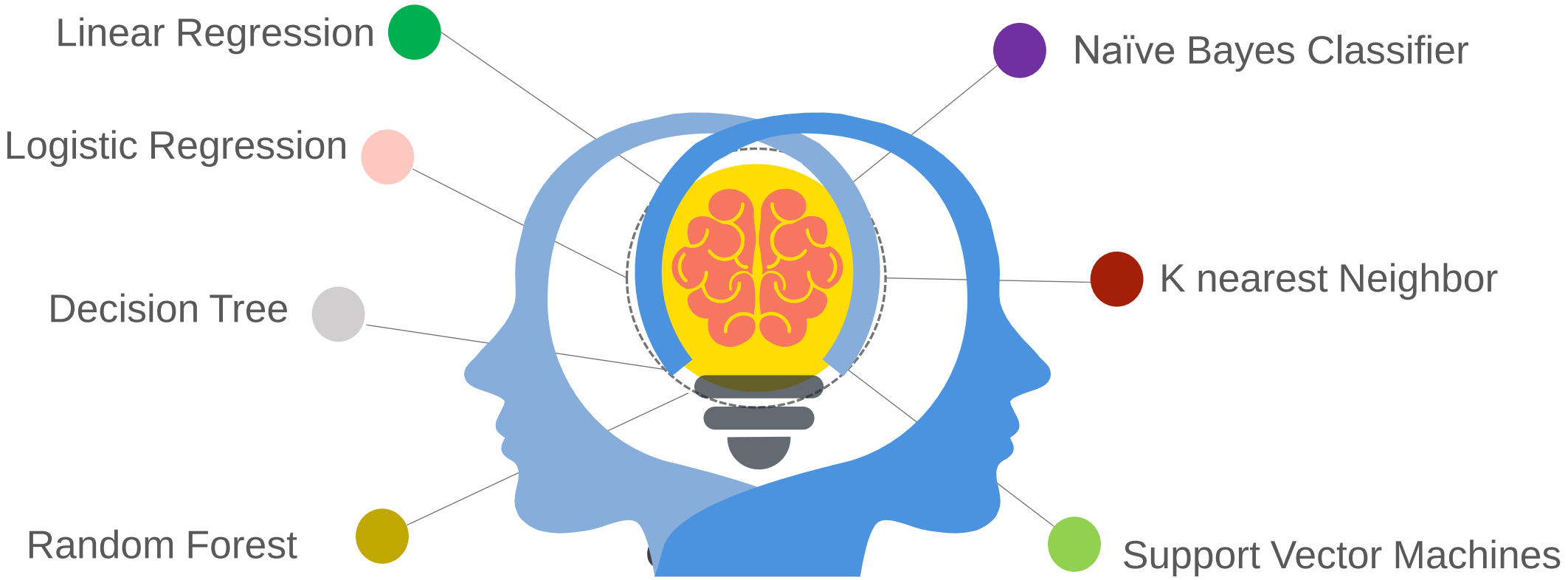
Problem Statement: To cluster a set of movies as either good or average based on their social media out reach

	director_facebook_likes	actor_3_facebook_likes	actor_1_facebook_likes	cast_total_facebook_likes
AvatarÂ	0	855	1000	4834
Pirates of the C...	563	1000	40000	48350
SpectreÂ	0	161	11000	11700
The Dark Knigh...	22000	23000	27000	106759
John CarterÂ	475	530	640	1873
Spider-Man 3Â	0	4000	24000	46055
TangledÂ	15	284	799	2036
Avengers: Age ...	0	19000	26000	92000
Harry Potter an...	282	10000	25000	58753
Batman v Super...	0	2000	15000	24450
Superman Retur...	0	903	18000	29991
Quantum of Sol...	395	393	451	2023
Pirates of the C...	563	1000	40000	48486



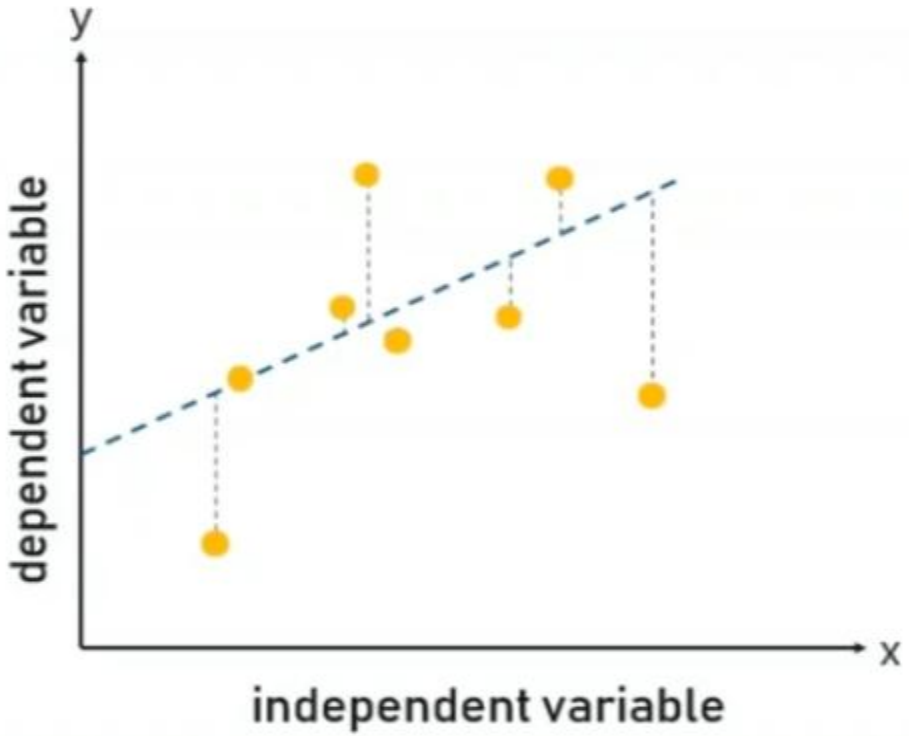
Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Supervised Machine Learning Algorithms



Linear Regression

Linear Regression is a method to predict dependent variable (Y) based on values of independent variables (X). It can be used for the cases where we want to predict some continuous quantity.



- *Dependent variable (Y):*

The response variable who's value needs to be predicted.

- *Independent variable (X):*

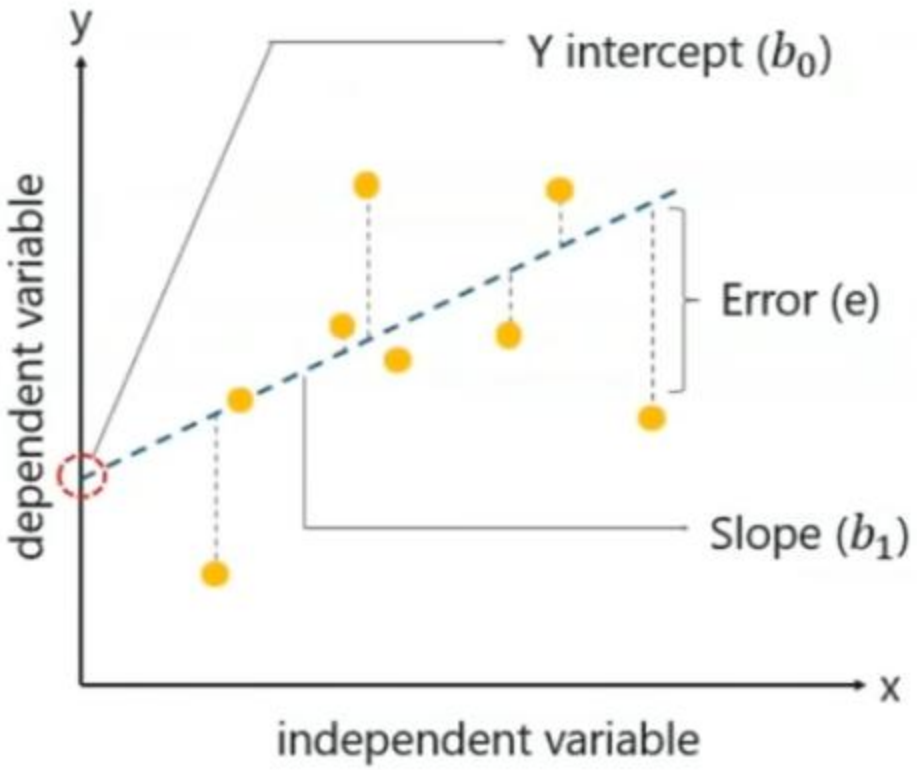
The predictor variable used to predict the response variable.

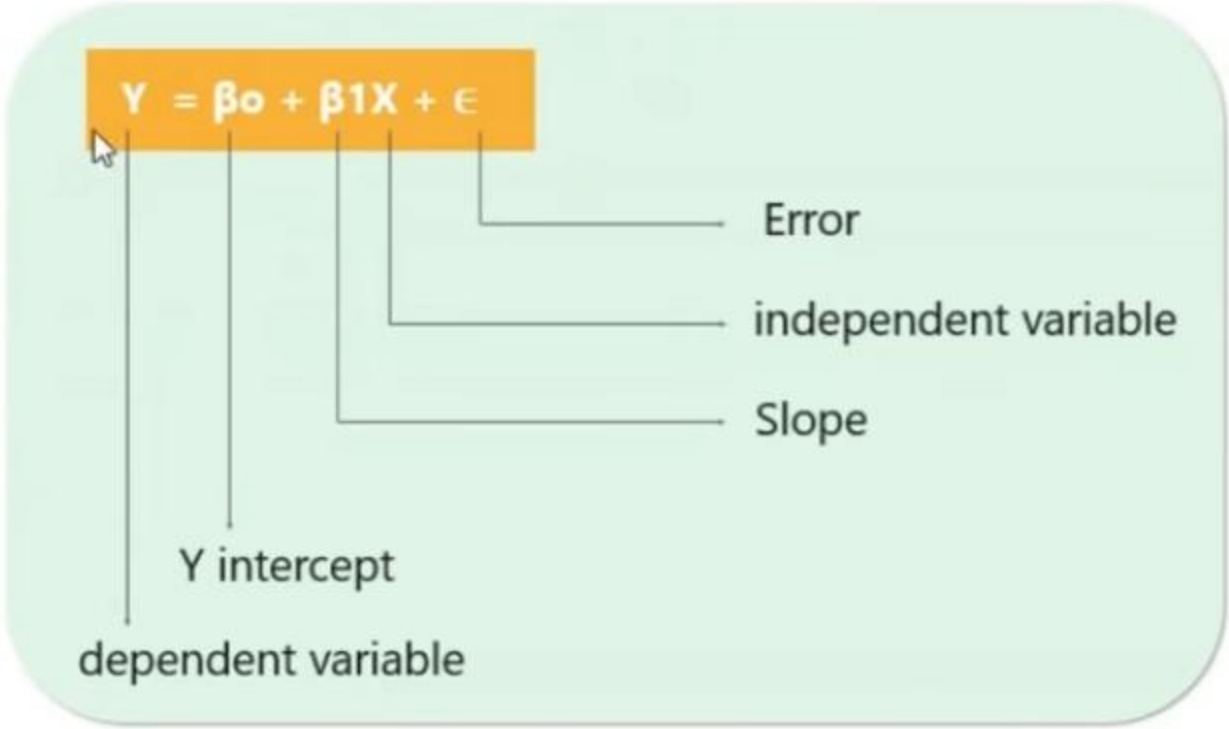
The following equation is used to represent a linear regression model:

$$Y = \beta o + \beta 1X + \epsilon$$

Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Linear Regression





Slides are adapted from Artificial Intelligence Full Course | Artificial Intelligence Tutorial for Beginners | Edureka

Thank You