

MACHINE LEARNING

On jupyter –ML_WCUBE

Date- 07/10/2k23

https://www.youtube.com/watch?v=O0Ka_nBRtN0

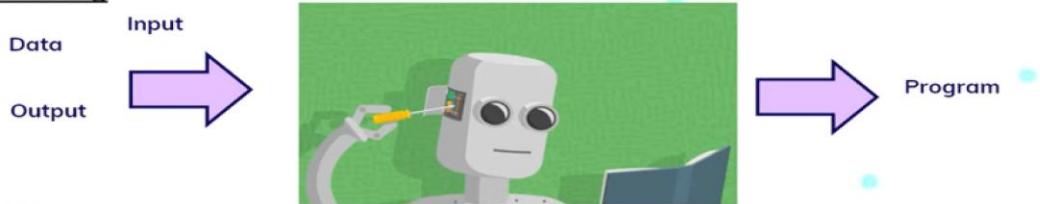
How ML is Different from Traditional Programming?



Traditional Programming

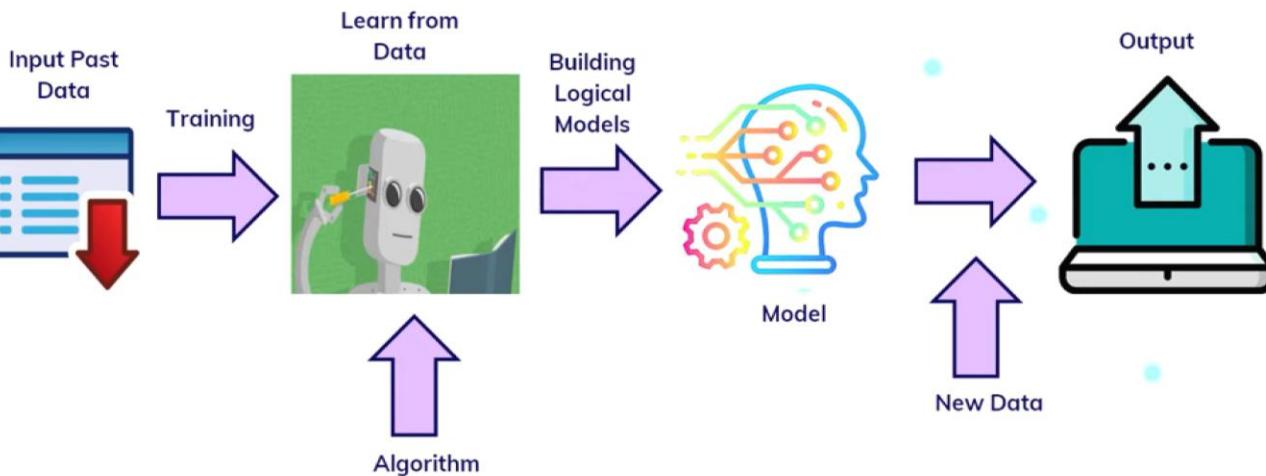


Machine Learning



www.wscubetech.com

How it Works?



www.wscubetech.com

Characteristics of Machine Learning

The Potential to Perform
Automated Data Visualization

Precise Data Analysis

Business Intelligence
at it's Finest

www.wscubetech.com



WSCUBE TECH™
System For Satisfaction

Advantages of Machine Learning

-  Resolving complex problems
-  Automation for everything
-  Trends and Pattern Identification
-  Wide Range of Applications

www.wscubetech.com

Disadvantages of Machine Learning

WSCUBE TECH™
System For Satisfaction

-  Data Acquisition
-  Highly Error-Prone
-  Algorithm Selection
-  Time-Consuming

www.wscubetech.com

Classification of Machine Learning

WSCUBE TECH™
System For Satisfaction



www.wscubetech.com

Timeline of Machine Learning

Alan Turning Proposed
The Turing Test.
Same Year Isaac Asimov Proposed
The Three Laws Of Robotics

The Term "Machine
Learning" Was First
Coined By Arthur Samuel.

The First Autonomous Vehicle
Was Created At Stanford AI Lab

1950

1959

1974

Arthur Samuel Who Was The Pioneer
Of Machine Learning, Created A
Program That Helped An IBM Computer
To Play A Checkers Game.

The First Robot
Was Introduced

IBM Deep Blue Intelligent
Computer Beats Expert
Garry Kasparov At Chess

www.wscubetech.com

Timeline of Machine Learning

Google Starts Building
Self-driving Car

DeepFace Was A Deep Neural Network
Created By Facebook, And They
Claimed That It Could Recognize A
Person With The Same Precision As A
Human Can Do

At Present
Machine Learning is everywhere
Around Us

2009

2014

At Present

2012

2017

Google Created A Deep Neural
Network Which Learned To
Recognise The Image Of Humans
And Cats

The Alphabet's Jigsaw Team Built An Intelligent System That Was
Able To Learn The Online Trolling. It Used To Read Millions Of
Comments Of Different Websites To Learn To Stop Online Trolling.

www.wscubetech.com

Career Opportunities



Machine
Learning
Engineer

Business
Intelligence
Developer

Data
Scientist

Human Centered
Machine Learning
Designer

NLP
Scientist

www.wscubetech.com

Prerequisites

If you are interested in learning more about machine learning, a few requirements should be met to be in order to excel in this field.

These requirements include:

- Basic knowledge of Python programming language
- Fundamental knowledge of statistics and probability
- Understanding of linear algebra and calculus

www.wscubetech.com

What is Machine Learning?

“Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.”

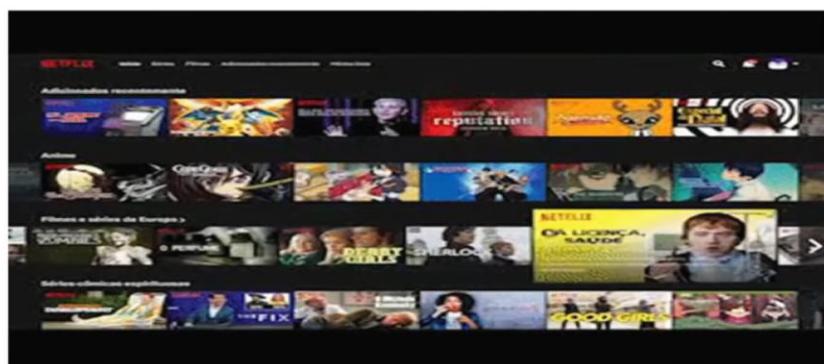


Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Recommendation Engines

Example: Netflix Viewing Suggestions

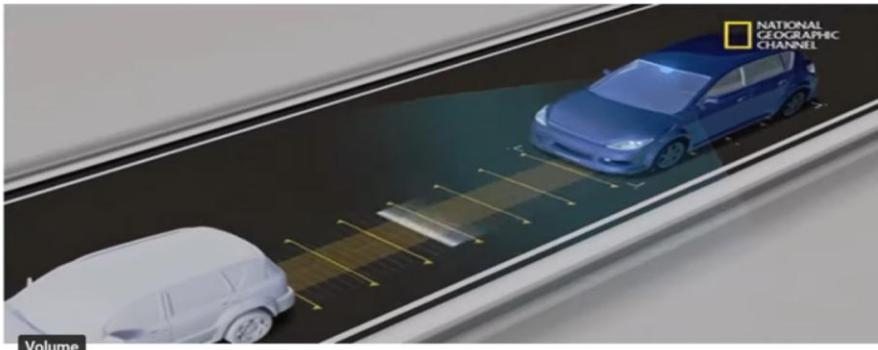
Application Area: Media + Entertainment + Shopping



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Self-Driving Cars

Example: Tesla Cars use ML to understand surrounding
Application Area: Automotive + Transportation



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Gamified Learning and Education

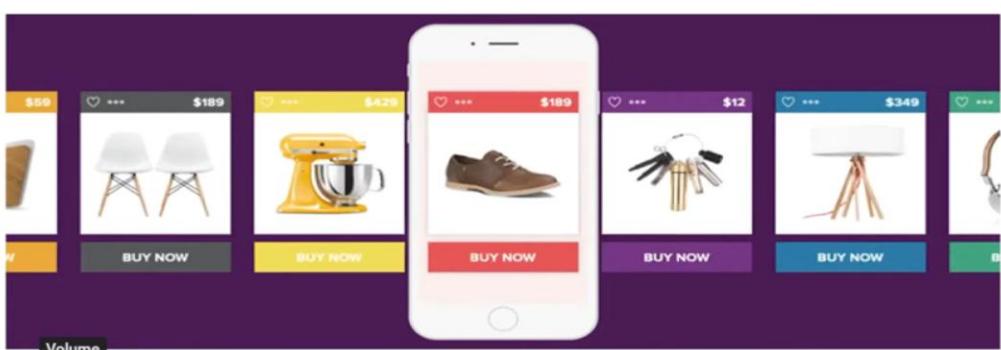
Example: Duolingo's Mobile Application
Application Area: Learning Language Application



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

E-Commerce Websites

Example: Myntra
Application Area: Fashion E-Commerce



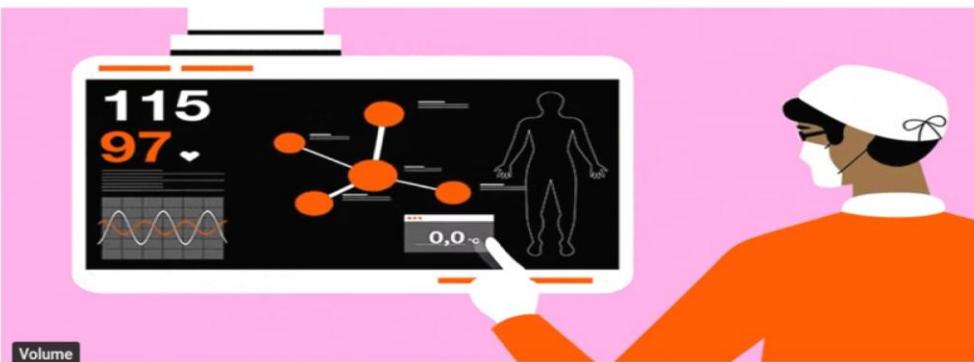
Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Medical Diagnosis



Example: Orderly Health

Application Area: HealthCare



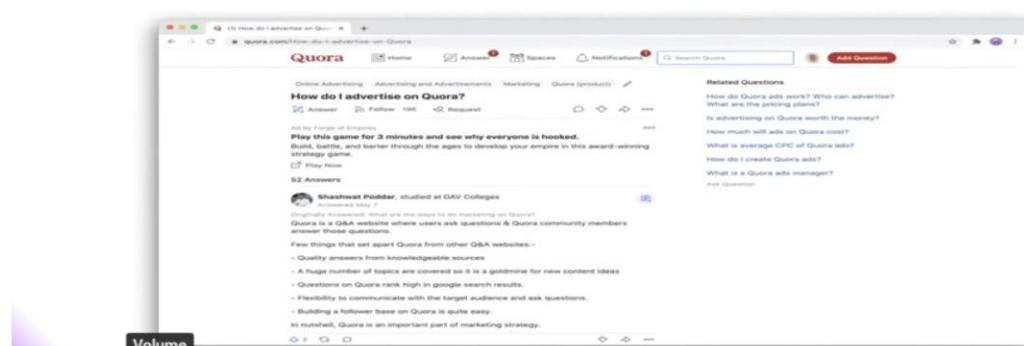
Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Getting Your Right Answers



Example: Quora's Super-Specific Answer Ranking

Application Area: Search



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com



Machine Learning Life Cycle

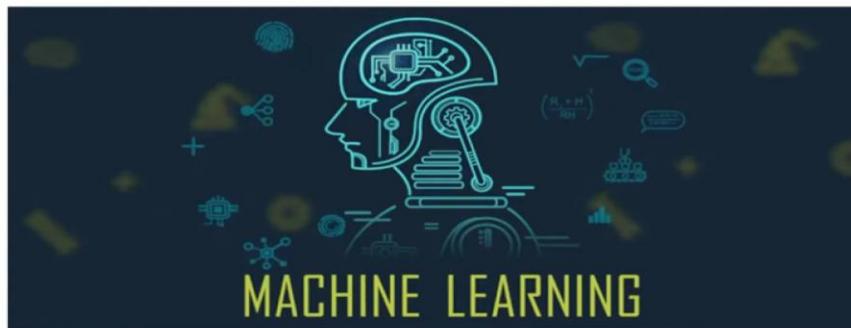


SUBSCRIBE

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

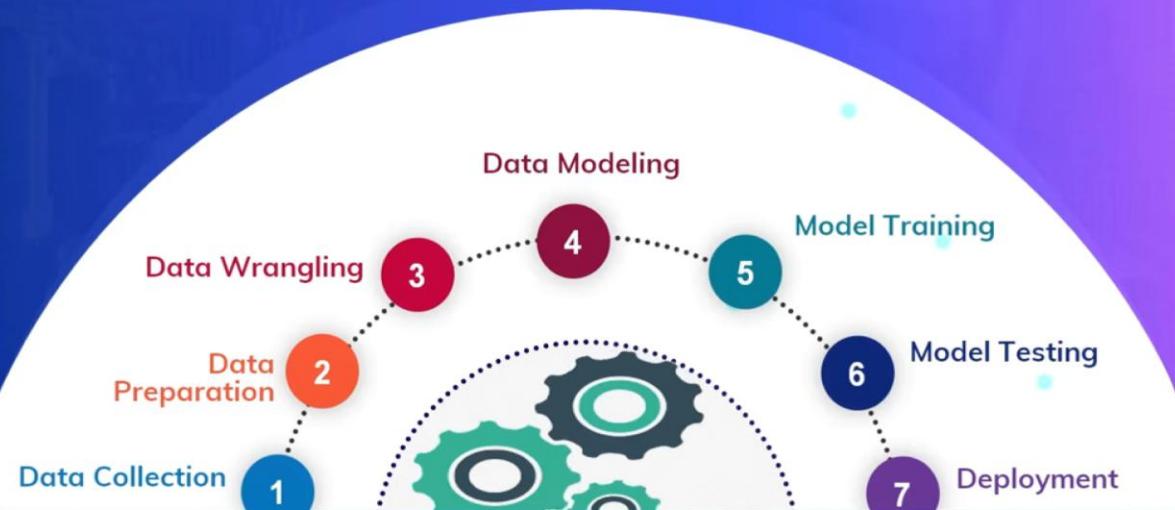
Introduction

“Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.”



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Machine Learning Life Cycle



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Data Collection

Goal over here is to gather as much as relevant data as possible



Identity various source of information



Gather data



Combine the data acquired from various data source

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Data Preparation

Data preparation deals with exploration of your data to generate far better results



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format.

This step involves:

-  Filtering/ cleaning up of raw data
-  Filtering Noise
-  Recognizing and removing outliers
-  Removing or filling the missing values

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Data Modeling

Data Modeling is the step in which we take the data and select a machine learning algorithm to built a model

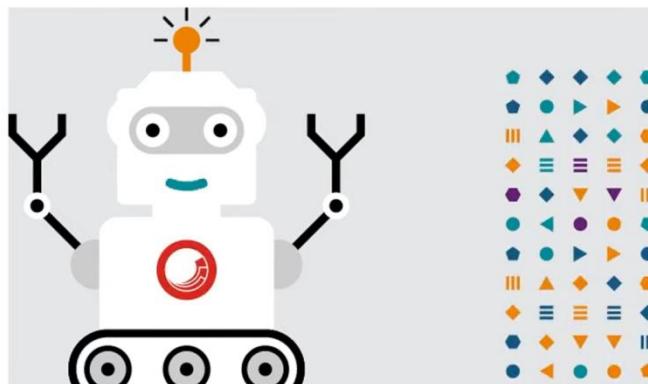
This step involves:

-  Selecting machine learning algorithm
-  Building the models
-  Validating the results

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Model Training

A machine learning training model is a process in which a machine learning (ML) algorithm is fed with sufficient training data to learn from.



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Model Verification

This stage of the machine learning lifecycle involves checking for the accuracy of the model by providing with the inputs that are unseen.



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

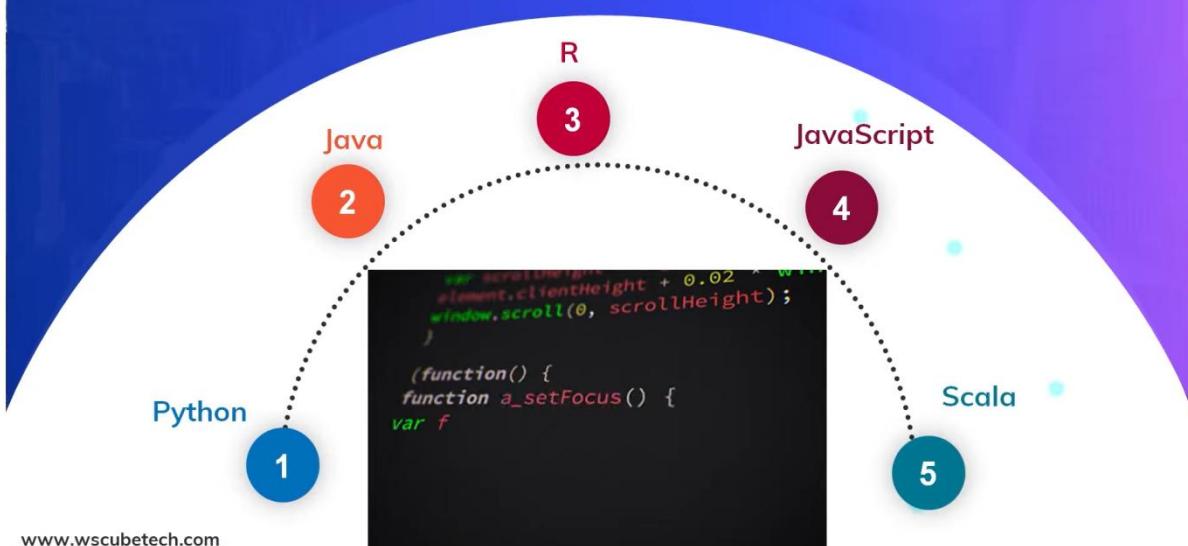
Model Deployment

This is the final step in the machine learning life cycle where we have a brilliant model ready to go to production



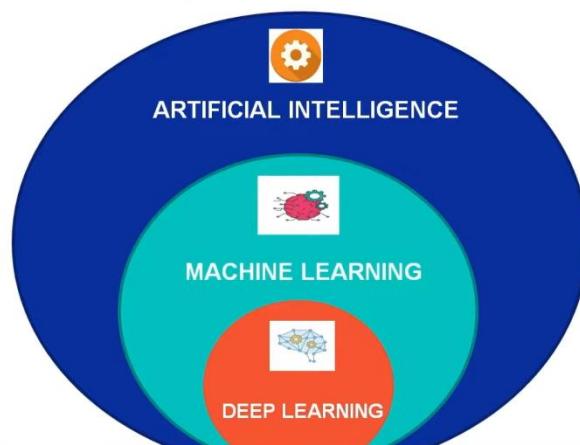
Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Best Languages for Machine Learning



Artificial Intelligence vs Machine Learning vs Deep Learning Machine Learning Tutorial

AI Vs ML Vs DL



Artificial Intelligence

“Artificial intelligence is the science and engineering of making computers capable of performing tasks that typically require human intelligence.”

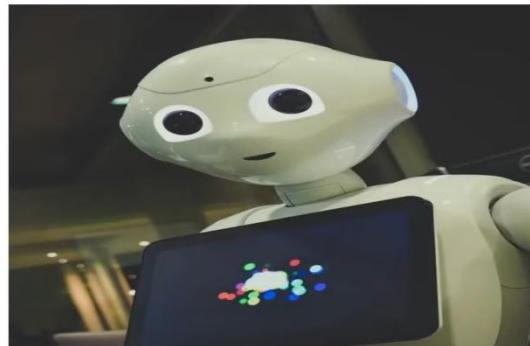
Based on the capabilities, AI classified as:



Applied AI (Weak AI)

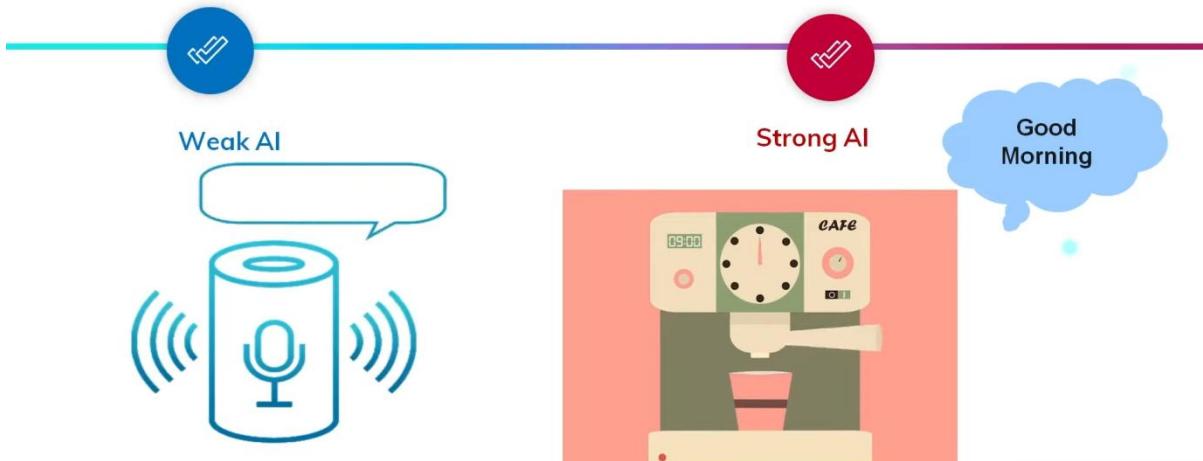


Generalized AI (Strong AI)



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Example of Weak AI & Strong AI



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Machine Learning

“Machine Learning is a subset of Artificial Intelligence which enables machines to learn from past data or experience without being explicitly programmed.”

Classification of Machine Learning:



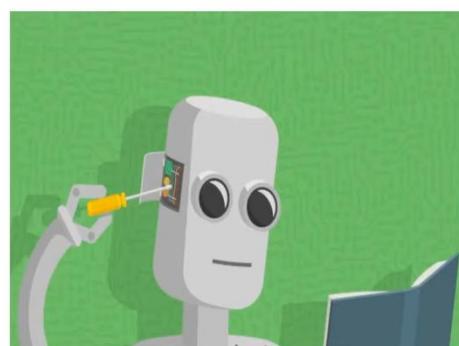
Supervised Learning



Unsupervised Learning

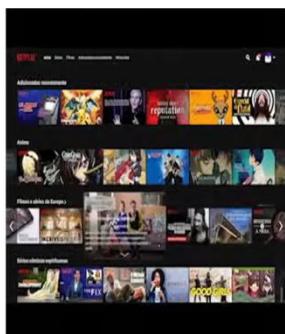


Reinforcement Learning

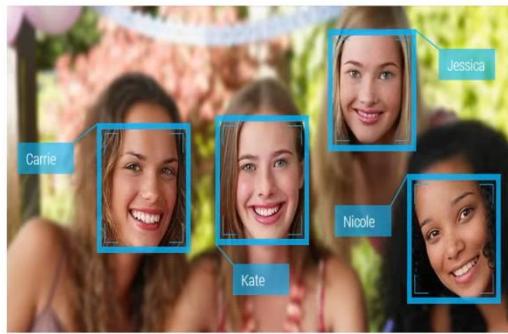


Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Applications of Machine Learning



Recommendation Engines



Auto Friend Tagging Suggestion



Spam Detection

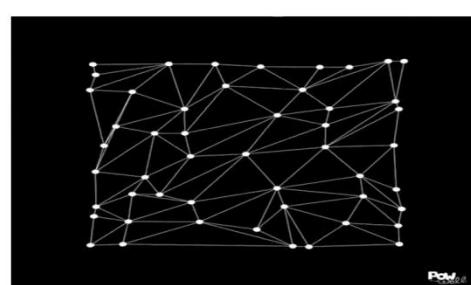
Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Deep Learning

“Deep Learning is a subset of Machine Learning concerned with the algorithms inspired by the structure and function of human brains.”



Human Brain with Billions of Tiny Neurons



Artificial Neural Network

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Applications of Deep Learning



Automatic Machine Translation



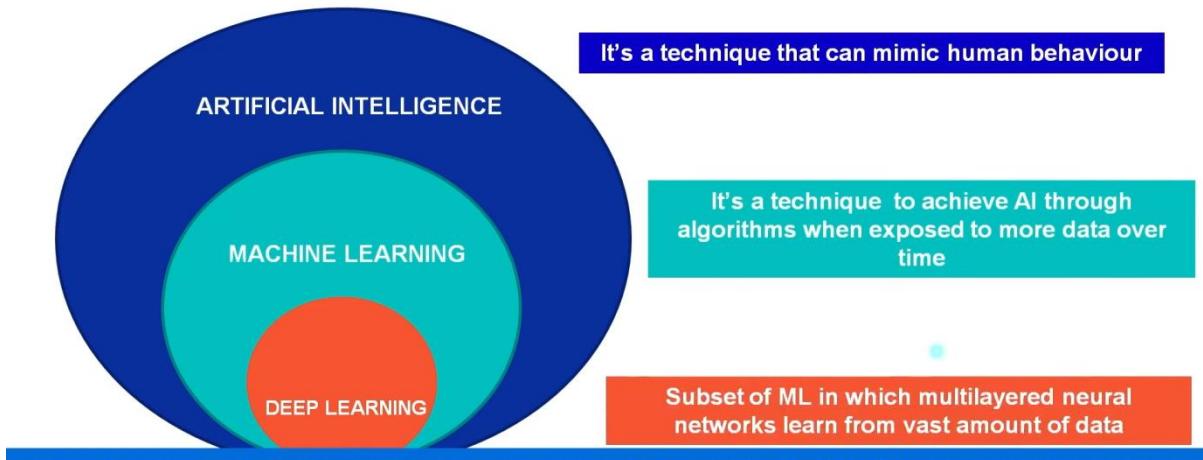
Colorization of Black & White



Fraud News Detection

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Recap



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com



How to get Datasets for Machine Learning?



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

What is Dataset?

“A dataset is a collection of records usually presented in tabular form.”

Name	Age	City	Department
Sam	25	Jodhpur	Digital Marketing
Sharon	27	Jaipur	Developer
Jack	32	Pune	SEO
Mark	42	Mumbai	Content Writer
Diana	50	Hyderabad	Trainer

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

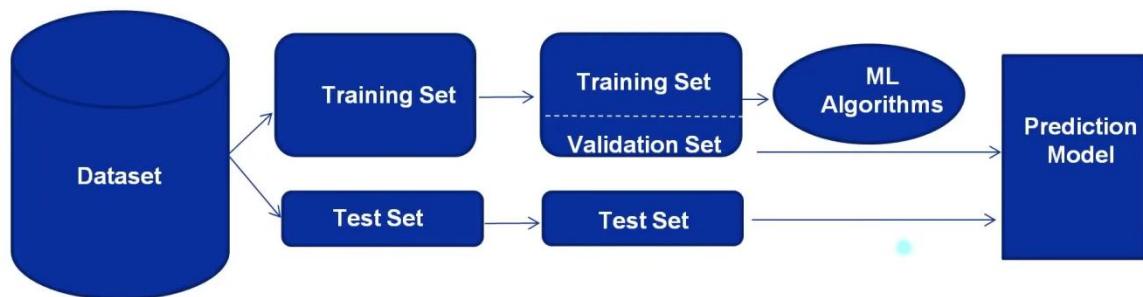
Types of Datasets



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Need of Dataset

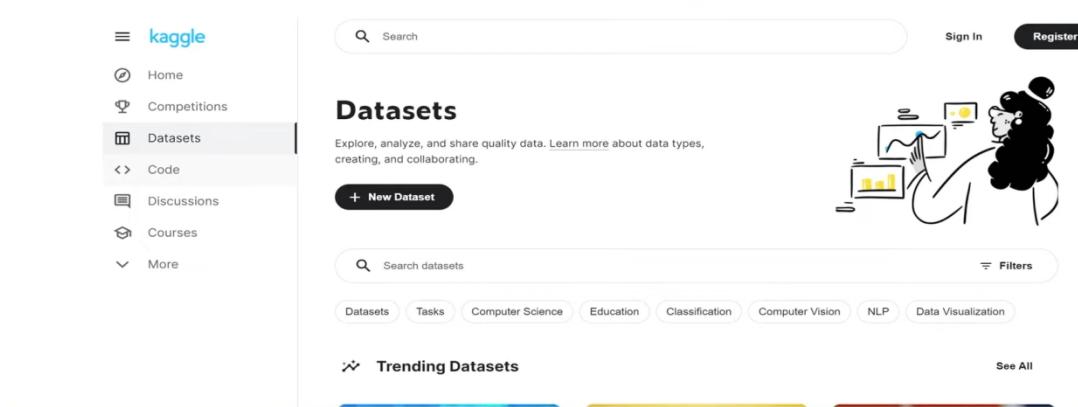
In the development phase of ML projects, datasets are classified as
Training dataset and Test Dataset



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Kaggle Datasets

Link <https://www.kaggle.com/datasets>.



The screenshot shows the Kaggle Datasets homepage. The left sidebar includes links for Home, Competitions, Datasets (which is selected), Code, Discussions, Courses, and More. The main area features a search bar, a 'Datasets' section with a brief description, a 'New Dataset' button, and a 'Trending Datasets' section. Below these are filters for Datasets, Tasks, Computer Science, Education, Classification, Computer Vision, NLP, and Data Visualization. A cartoon character is shown working with a computer screen displaying graphs.

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Amzon datasets



Link <https://registry.opendata.aws/>

Registry of Open Data on AWS

About
This registry exists to help people discover and share datasets that are available via AWS resources. Learn more about sharing data on AWS. See all usage examples for datasets listed in this registry.
See datasets from Digital Earth Africa, Facebook Data for Good, NASA Space Act Agreement, NIH STRIDES, NOAA Big Data Program, Space Telescope Science Institute, and Amazon Sustainability Data Initiative.

Search datasets (currently 259 matching datasets)
Search datasets

Add to this registry
If you want to add a dataset or example of how to use a dataset to this registry, please follow the instructions on the Registry of Open Data on AWS GitHub repository.
Unless specifically stated in the applicable dataset documentation, datasets available through the Registry of Open Data on AWS are not provided and maintained by AWS. Datasets are provided and maintained by a variety of third parties under a variety of licenses. Please check dataset licenses and related documentation to determine if a dataset may

The Cancer Genome Atlas
The Cancer Genome Atlas (TCGA), a collaboration between the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI), aims to generate comprehensive, multi-dimensional maps of the key genomic changes in major types and subtypes of cancer. TCGA has analyzed matched tumor and normal tissues from 11,000 patients, allowing for the comprehensive characterization of 33 cancer types and subtypes, including 10 rare cancers. The dataset contains open Clinical Supplement, Biospecimen Supplement, RNA-Seq Gene Expression Quantification, miRNA-Seq Isoform Expression Quantificati...

Details →
Usage examples

- Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines by Kyle Elliott, Matthew H. Bailey, et al.
- The Immune Landscape of Cancer by Vésteinn Thorsson, David L. Gibbs, et al.
- Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas by Theo A. Knijnenburg, Linghua Wang, et al.
- Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas by Joshua D. Campbell, Christina Yau, et al.
- Cancer Genomics Cloud by Seven Bridges

See 29 usage examples →

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

UCI Machine Learning Repository



Link <http://archive.ics.uci.edu/ml/index.php>

UCI Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 588 data sets as a service to the machine learning community. You may [view all data sets](#) through our searchable interface. For a general overview of the Repository, please visit our [About](#) page. For information about citing data sets in publications, please read our [citation policy](#). If you wish to donate a data set, please consult our [donation policy](#). For any other questions, feel free to contact the Repository librarians.

Supported By: In Collaboration With:

Latest News:	Newest Data Sets:	Most Popular Data Sets (hits since 2007):
<p>09-24-2018: Welcome to the new Repository admins Dheeru Dua and Elia Karra Taniskidou!</p> <p>04-04-2013: Welcome to the new Repository admins Kevin Bache and Moshe Lichman!</p> <p>03-01-2010: Note from donor regarding Netflix data</p> <p>10-16-2009: Two new data sets have been added.</p> <p>09-14-2009: Several data sets have been added.</p> <p>03-24-2008: New data sets have been added!</p> <p>06-25-2007: Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope</p>	<p>04-21-2021: Synchronous Machine Data Set</p> <p>04-20-2021: Wikipedia Math Essentials</p> <p>04-20-2021: Wikipedia Math Essentials</p> <p>02-17-2021: Hungarian Chickenpox Cases</p>	<p>4013853: Iris</p> <p>2164221: Adult</p> <p>1674476: Wine</p> <p>1539946: Heart Disease</p>

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Google's Data Search Engine



Link <https://datasetsearch.research.google.com/>

Google

Flower

Last updated Download format Usage rights Topic Free Saved datasets

100+ datasets found

kaggle Iris Flower Dataset www.kaggle.com zip Updated Mar 22, 2018

kaggle Flowers Recognition www.kaggle.com wb.n3ncloud.co.kr zip Updated Jun 28, 2018

Flowers Classification Dataset

kaggle Iris Flower Dataset
Iris flower data set used for multi-class classification.
Explore at Kaggle
zip(1010 bytes)
Dataset updated Mar 22, 2018
Authors MathNerd
License CC0 1.0 Universal Public Domain Dedication

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Microsoft datasets

Link <https://msropendata.com/>

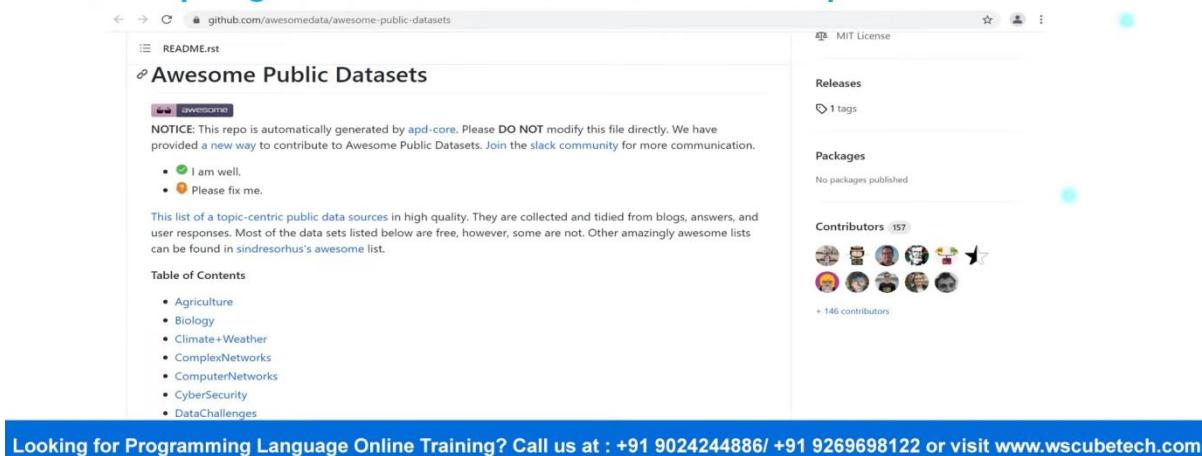


Dataset Categories

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Awesome Public Datasets Collections

Link <https://github.com/awesomedata/awesome-public-datasets>



NOTICE: This repo is automatically generated by apd-core. Please DO NOT modify this file directly. We have provided a new way to contribute to Awesome Public Datasets. Join the slack community for more communication.

- I am well.
- Please fix me.

This list of a topic-centric public data sources in high quality. They are collected and tidied from blogs, answers, and user responses. Most of the data sets listed below are free, however, some are not. Other amazingly awesome lists can be found in sindresrhus's awesome list.

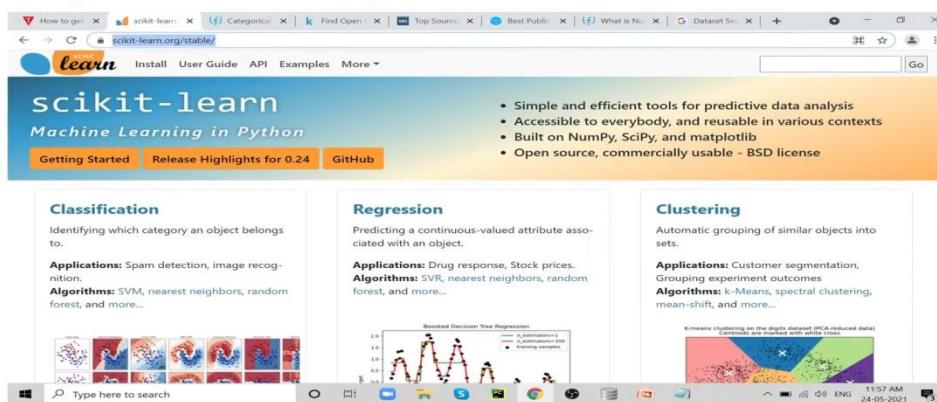
Table of Contents

- Agriculture
- Biology
- Climate+Weather
- ComplexNetworks
- ComputerNetworks
- CyberSecurity
- DataChallenges

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Scikit-learn Datasets

Link <https://scikit-learn.org/stable/>



Classification
Identifying which category an object belongs to.
Applications: Spam detection, image recognition.
Algorithms: SVM, nearest neighbors, random forest, and more...

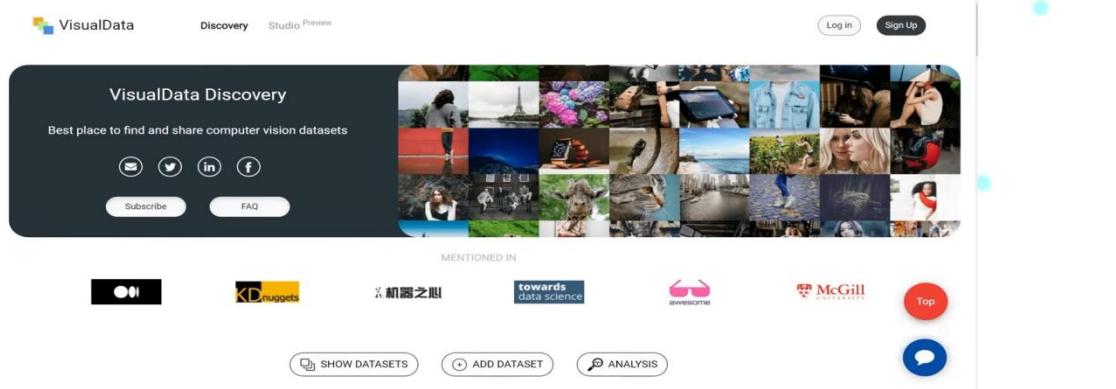
Regression
Predicting a continuous-valued attribute associated with an object.
Applications: Drug response, Stock prices.
Algorithms: SVR, nearest neighbors, random forest, and more...

Clustering
Automatic grouping of similar objects into sets.
Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, and more...

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Computer Vision Datasets

Link <https://www.visualdata.io/discovery>



Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Government Datasets

Indian Government Datasets <https://data.gov.in/>

US Government Datasets <https://www.data.gov/>

Northern Ireland Public Sector Datasets <https://www.opendatani.gov.uk/>

European Union Open Data Portal <https://data.europa.eu/euodp/data/dataset>

Looking for Programming Language Online Training? Call us at : +91 9024244886/ +91 9269698122 or visit www.wscubetech.com

Techniques & Steps of Data Preprocessing

Data Preprocessing in Machine Learning

Module 2 | Techniques & Steps of Data Preprocessing

What is Data Pre-processing?



“Data Preprocessing is a process of converting your raw data into suitable form.”



Data Preprocessing Steps Involves



Time -2 .31

Features & Labels in Machine Learning



Feature & Label in Machine Learning?

FEATURE: In Machine Learning feature means property of your training data.

LABEL: In Machine Learning label means the output you get from your model after training.

	A	B	C	D
1	Person	Height (in feet)	Weight (in kg)	Foot Size (in inches)
2	male	6	81	12
3	male	5.92	86	11
4	male	5.58	90	12
5	female	5.92	77	10
6	female	5	45	6
7	female	5.5	68	8
8	female	5.52	58	7
9	female	5.75	68	9

www.wscubetech.com

Example: Feature & Label

Case 1: The prediction(Y) Calories here is a Label.
 Calories is the column that you want to predict using various features like
 x1: Gender, x2: Height and x3: Weight .

	A	B	C	D	E	F	G	H
1	Person	Age	Height (in feet)	Weight (in kg)	Duration	Heart_Rate	Body_Temp	Calories
2	male	68	6	94	29	105	40.8	231
3	male	70	5.92	79	5	88	38.7	26
4	female	20	5.44	60	14	94	40.3	66

Case 2: The prediction(Y) HeartRate here is a Label.
 HeartRate is the column that you want to predict using various features like
 x1: Gender and x2: Weight .

Example: Feature & Label

The prediction(Y) PetName here is a Label.

PetName is the column that you want to predict using various features like
x1: Age, x2: HomeRegion and x3: FamilyIncome .

	A	B	C	D
1	Age	Home_Region	Family_Income	Pet_Name
2		12 City	1 Lakh	Aquarium
3		45 Village	10 Lakh	Horse
4		28 City	15 Lakh	Dog

Question: Feature & Label?

WScube Tech



Question – which is label you have to take for prediction ?

Supervised Machine Learning



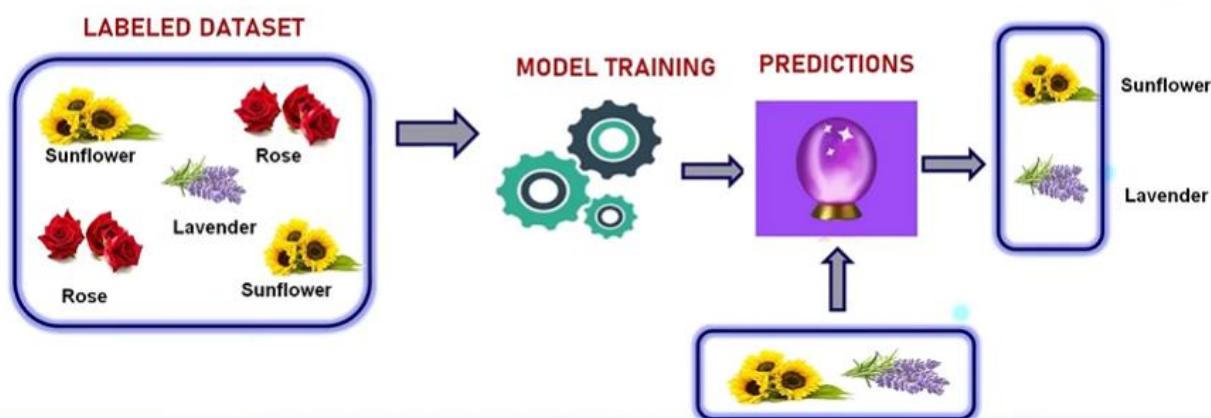
What is Supervised Learning?

"In Supervised learning, you train the machine using data which is well "labeled", i.e. some input data is already tagged with correct answer and this algorithm learns from labeled training data that helps you to predict the further outcomes."



How Does Supervised Learning Works?

The goal of Supervised Learning is to map the input variable (x) with the output variable (y)



Types of Supervised ML Algorithms



Linear Regression
Polynomial Regression
Regression Trees



Random Forest
Decision Trees
Logistic Regression
Support Vector machines

Advantages of Supervised Learning



You have full control over what the machine is learning



You can easily test and debug your model



You can determine the number of classes

Disadvantages of Supervised Learning



Have limited scope



Collecting labelled dataset is expensive and time-consuming

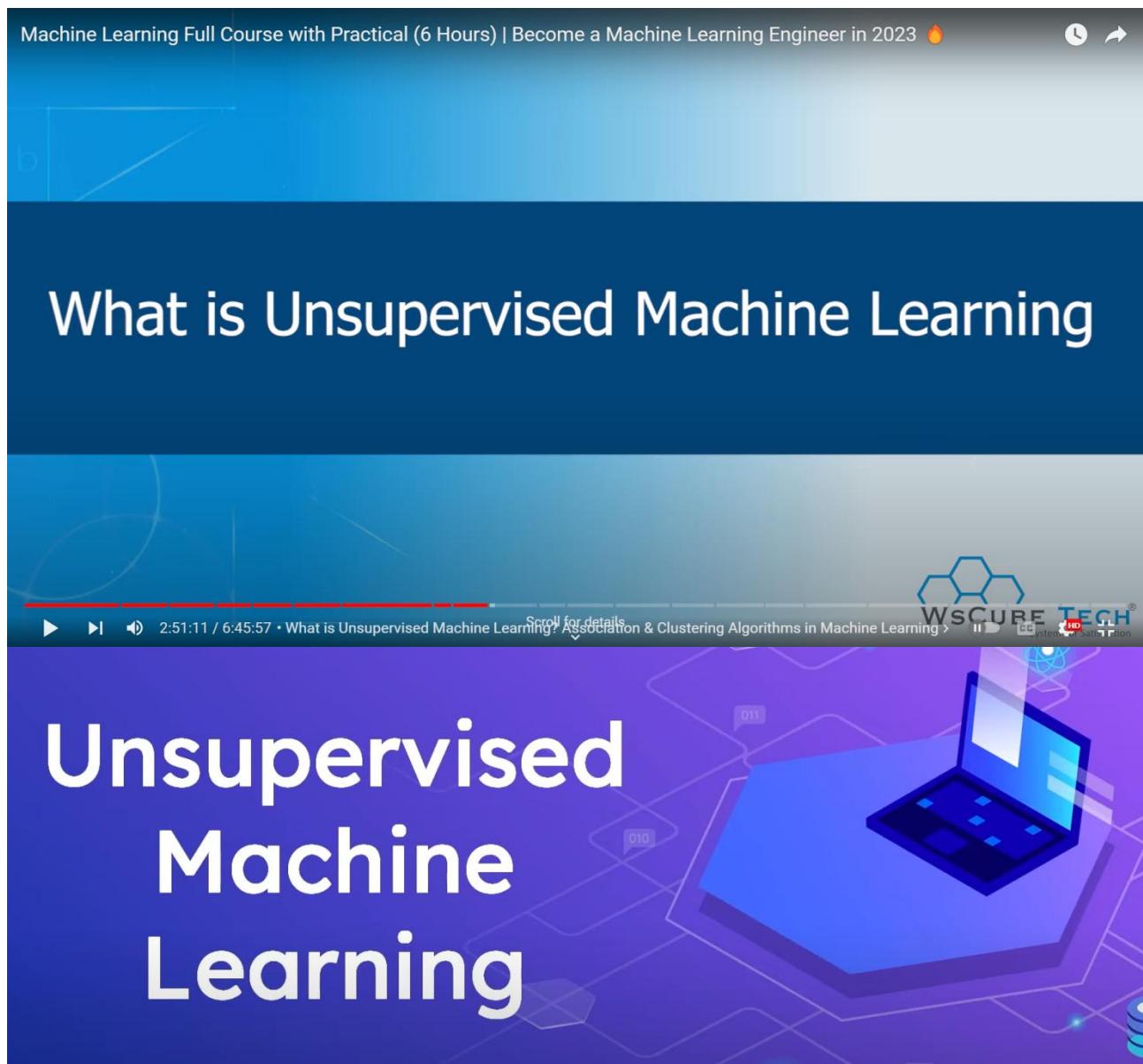


Wrong Prediction

Time – 2.47

Conclusion

- ❑ In Supervised learning, you train the machine using data which is well "labeled."
- ❑ Regression and Classification are two types of supervised machine learning techniques.
- ❑ The biggest challenge in supervised learning is that if your training set doesn't have examples that you want to have in a class, could give inaccurate results.



The screenshot shows a video player interface. At the top, it displays the title "Machine Learning Full Course with Practical (6 Hours) | Become a Machine Learning Engineer in 2023" with a fire emoji. On the right side of the title bar are icons for a clock and a share button. Below the title, the main content area has a dark blue header with the text "What is Unsupervised Machine Learning". The main body of the video features a purple background with a 3D rendering of a laptop displaying binary code (010, 011) and a network of lines and dots, representing data and connections. The WsCUBE TECH logo is visible in the bottom right corner of the video frame. At the very bottom of the screen, there is a dark footer bar with video control icons (play, pause, volume) and the text "2:51:11 / 6:45:57 • What is Unsupervised Machine Learning? Association & Clustering Algorithms in Machine Learning >".