

Machine Learning

Modified by: Yousef Elbaroudy

Instructor Information

- ❑ Yousef T. Elbaroudy
- ❑ Undergraduate Data Scientist student in AI Department at FCAI – Benha University
- ❑ Skilled at Data Analysis – Data Visualization – Machine/Deep Learning

[Linked in: Yousef Elbaroudy](#)

[Github: YousefTB](#)

[Kaggle: Yousef Elbaroudy](#)

[Data Science Playlist: Eng. Yousef Elbaroudy](#)



How to evaluate ?

- Must Exceed 65% of the evaluation to pass

Final Project	Tasks	Attendance
40%	40%	20%

- You must exceed 75% of attendance (attending 6 Sessions at least)
- The tasks will be for 2 approaches (in-session [20%] – assignment [20%])
- The final project will contain all topics of the training

Background

PREREQUISITES

Skill	Tool/Topics
Programming Skills	Python, R or MATLAB (Python is Preferred)
Mathematics	Differential Calculus, Linear Algebra
Probability & Statistics	Descriptive Measurements (Mean, Mode, Median), Distributions, Event probability, Conditional Probability, Joint Distributions, Bayes Theorem, A/B Testing ... etc
Database	SQL
Data Visualization	Python –Preferred- (Matplotlib, Seaborn), Tableau, PowerBI
Software Engineering	APIs, Data Structure, Algorithms, Web Development and Mobile Development

tools:

- statistical
- mathematical
- programming
- problem-solving
- data-management

OLD PLAN

Requirements	Topics	Duration [In Weeks]	Recommended Sources	Tasks
Python Programming Language (Jupyter Notebook)	Variables, I/O Commands Data Structures (Lists, Tuples, Dictionaries, Sets) Selective Structure (If - If else - If elif) Repetitive Structure (For & while loops) Functions Strings Standard Libraries (Math - csv - collection - json - random) Exception handling Debugging Files (JSON, CSV) Object-Oriented Programming Algorithms (Big-O) - Recursion - Searching - Sorting - DFS & BFS - Brute Force	3	Intro to Python for Computer Science & Data Science [Chapter 1 to 6 & Chapter 8 to 11] Book Python for Data Analysis [Chapter 2 & 3] Book Cisco CCNA Course [Basics & Advanced] Course Coursera Python Course [Basics & Advanced] Course Codecademy & AI-Dessay & AI-Zero playlists [Youtube]	Tic-Tac-Toe Game [Without Machine Learning] First Week Sheets Second Week Sheets Last Week Sheets Challenging Exercises [In last week] Discussion Each week
Basic Math (Pre-Calculus & Calculus)	Number Theory Functions Summations Exponents Logarithms Euler's Number & Natural Logarithms Limits Derivatives (Partial Derivative & Chain Rule) Integrals	1	Essential Math for Data Science [Chapter 1] Book Thomas Calculus [Chapter 1 to 5 & Chapter 7] Book	Book Exercises [at least one exercise for each topic] Discussion at the end of the week
Linear Algebra	NumPy Library (Using Python) * SciPy Library (Using Python) * Vectors basic & Vector Space Linear transformation Matrices Matrix multiplications Determinants Dimensions & Rank Special Matrices Elementary row operations & Row Reduction System of equations and inverse matrix Linear dependence Basis & Span Eigen Values & Eigen Vectors Orthogonality & Least-squares Gram-Schmidt process	3	Python for Data Science Handbook [Chapter 2] Book Python for Data Analysis [Chapter 4] Book Practical Linear Algebra for Data Science [Chapter 1 to 6 & Chapter 8 to 11] Book * Mathematics for Machine Learning [Chapter 1] Book Essential Math for Data Science [Chapter 4] Book Linear Algebra and Its Application [Chapter 1 to 6] Book * Intro to Python for Computer Science & Data Science [Chapter 7] Book	System of Linear equations Calculator (Using python) Multiple Sheets Discussion Each week

OLD PLAN

Probability & Statistics	Pandas Library (using Python) Statistical Library (using Python) * Statistical Models Library (using Python) * Populations, Samples and Bias Normal Distribution Joint Distribution Union Distribution Conditional Probability Bayes' Theorem Bernoulli Distribution Beta Distribution Poisson Distribution Descriptive Statistics (Mean - Median - Mode - Variance - Standard Deviation - Normal Distribution - Theoretical CDF - Z-Score - Empirical Rule - The Central Limit Theorem - Confidence Interval - Degrees of freedom - Hypothesis Testing - A/B Testing - T-Distribution - Chi-square Distribution - F-Distribution - Simple Linear Regression - Naive Bayes)	4	Dr. Ahmed Hegazy Playlist [Youtube] Python Data Science Handbook (Chapter 1 to 5) Book * Python for Data Analysis (Chapter 5) Book Essential Math for Data Science (Chapter 2 & 3) Book * Mathematics for Machine Learning (Chapter 7) Book	Multiple Sheets Discussion Each Week
SQL (Database) -Optional-	SQL Datatypes Tables creation Populating and Modifying Data Query Mechanics The (Select - where - from) clauses Condition Evaluation & Types Set operators & Rules Data generation, selection and generation Grouping & Aggregation Subqueries Querying multiple Tables	2	Learning SQL (Chapter 1 to 8) Book	
Data Visualization using Python	Matplotlib (using python) * Seaborn Library (using Python) * Simple Line Plots Simple Scatter Plots Visualizing Errors Density Estimation Plots Histograms and Density Customizing Plot legends Multiple subplots Text & Annotation Customizing Ticks Three-Dimensional Plotting Visualization with Seaborn	2	Python Data Science Handbook (Chapter 5) Book * Python for Data Analysis (Chapter 5) Book *	
Total Weeks			15	



WORLD OF LEARNING

What is Machine Learning

- Machine learning is a field of artificial intelligence (AI) that involves developing algorithms and statistical models that enable computers to learn and improve from experience, without being explicitly programmed.

Traditional Programming Vs Machine Learning

■ Traditional Programming



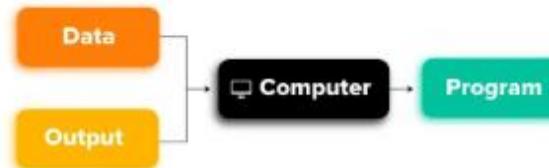
■ Machine Learning



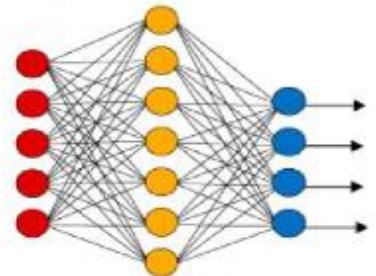
TRADITIONAL PROGRAMMING



MACHINE LEARNING



Simple Neural Network

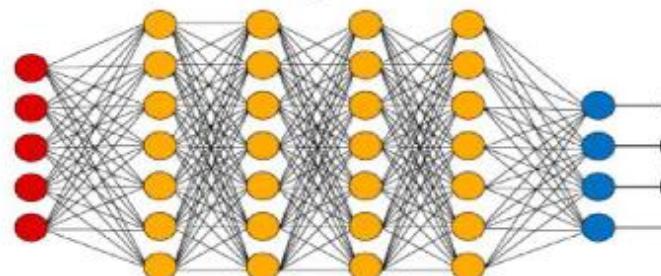


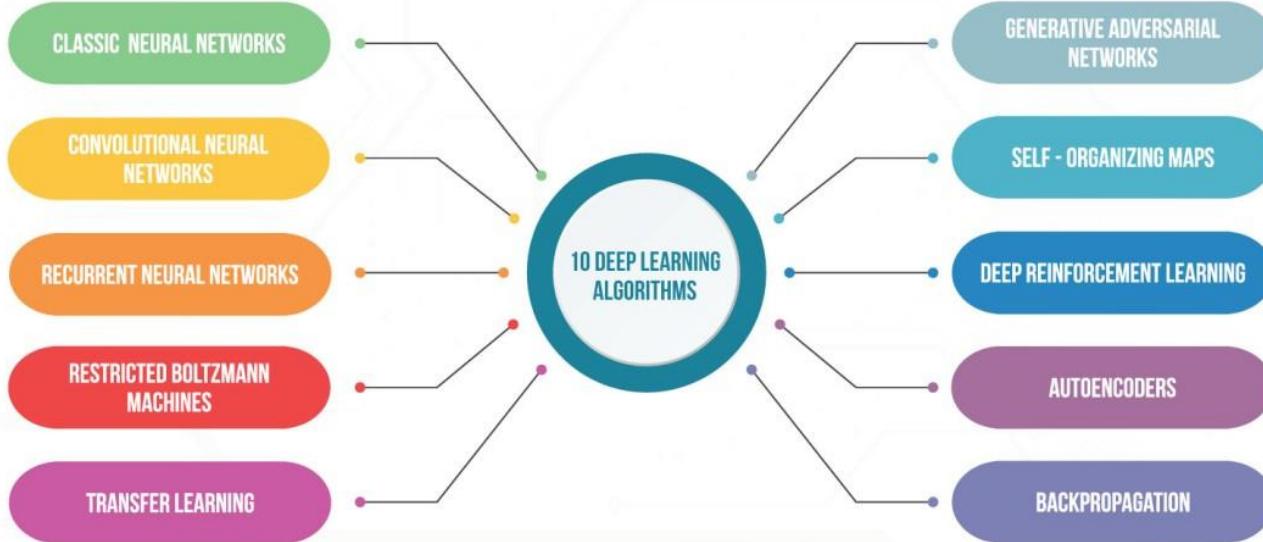
● Input Layer

● Hidden Layer

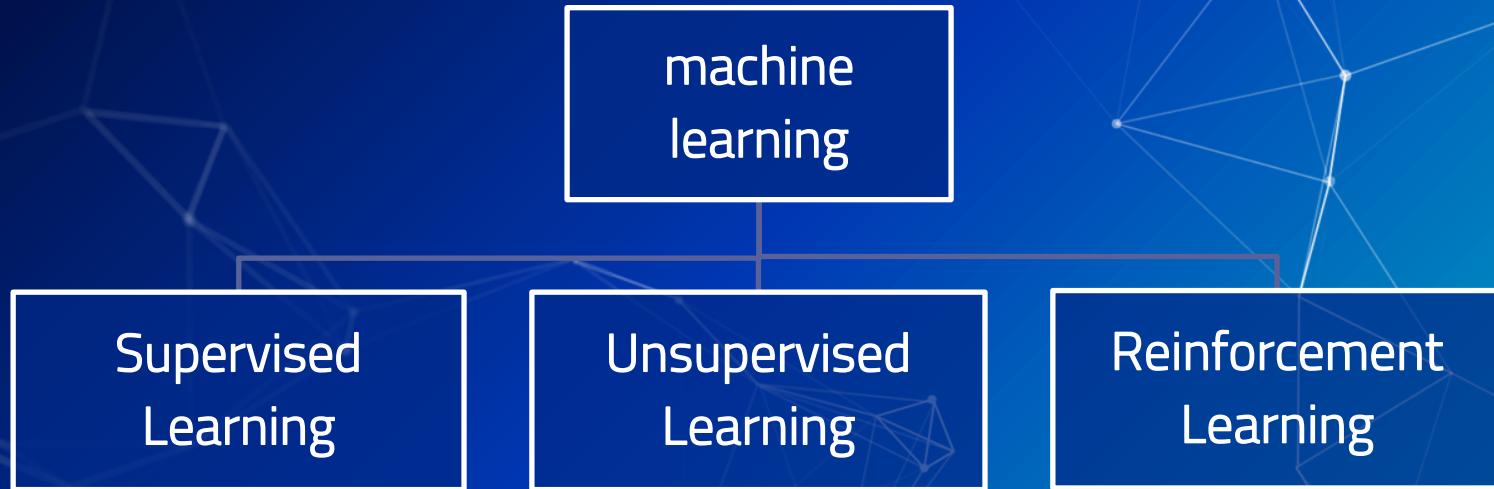
● Output Layer

Deep Learning Neural Network



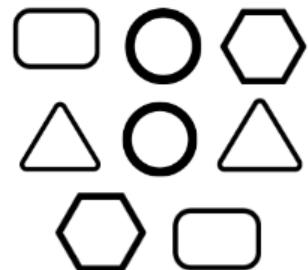


Types of Machine Learning

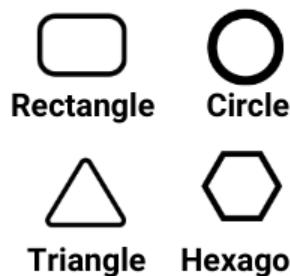


Supervised Learning

Labeled Data



Labels



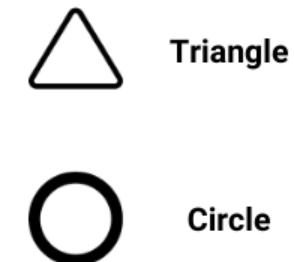
Machine



ML Model



Predictions

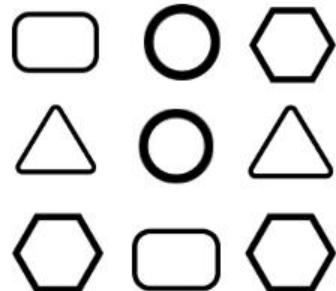


Test Data



Unsupervised Learning

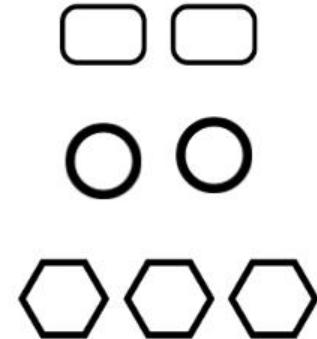
Unlabelled Data



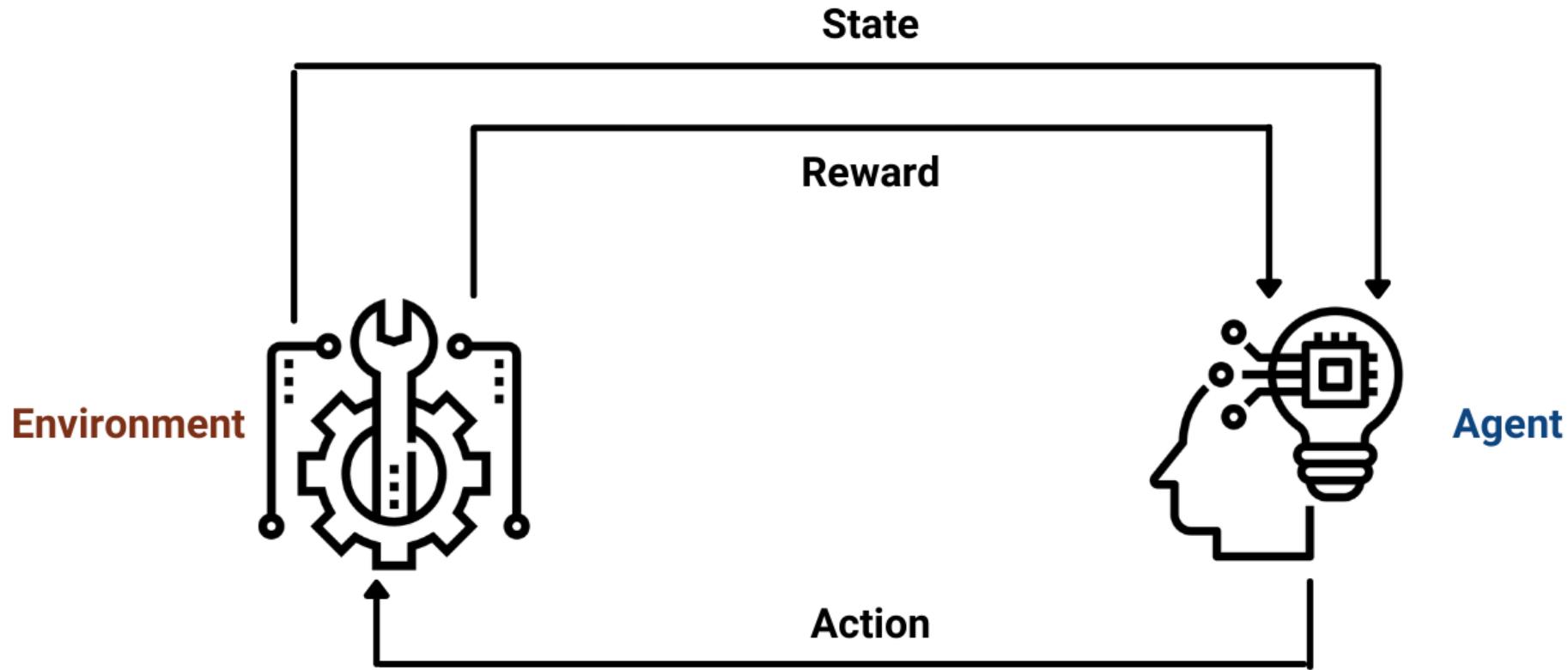
Machine



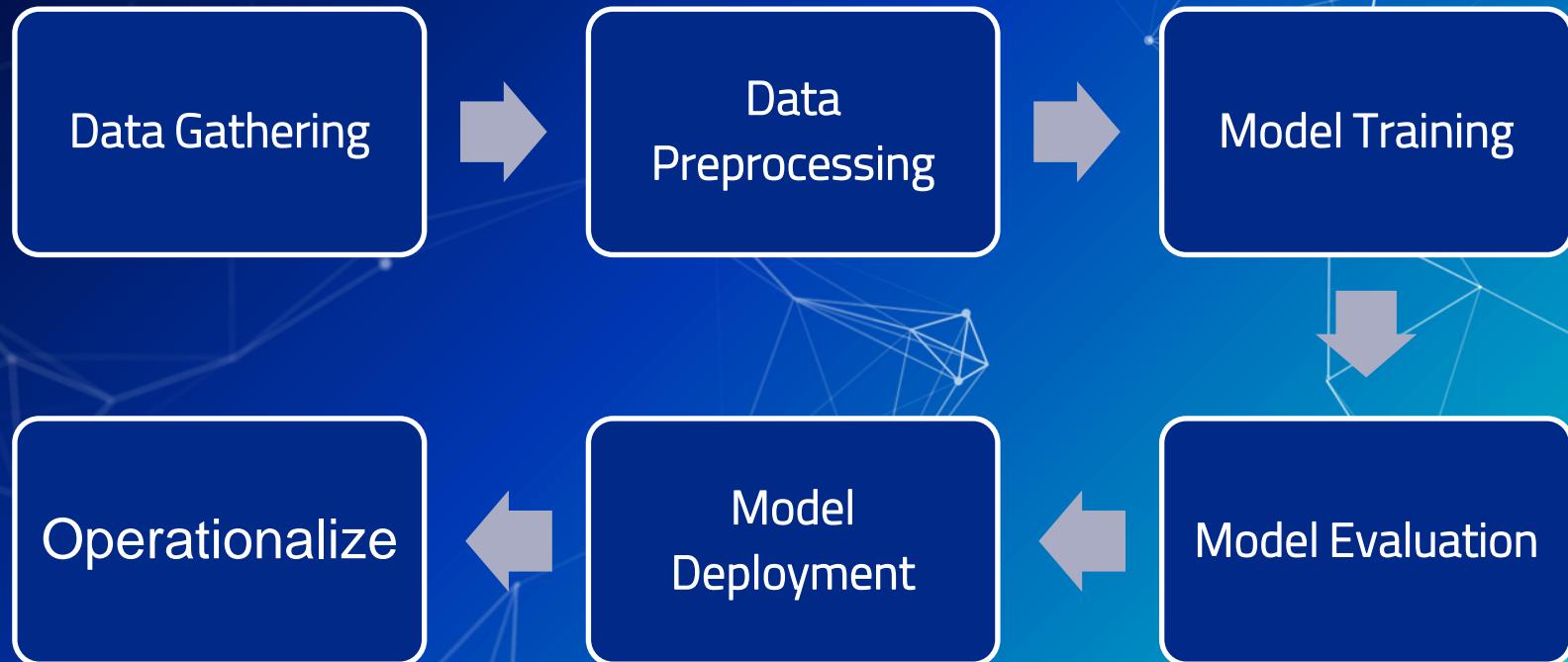
Results



Reinforcement Learning

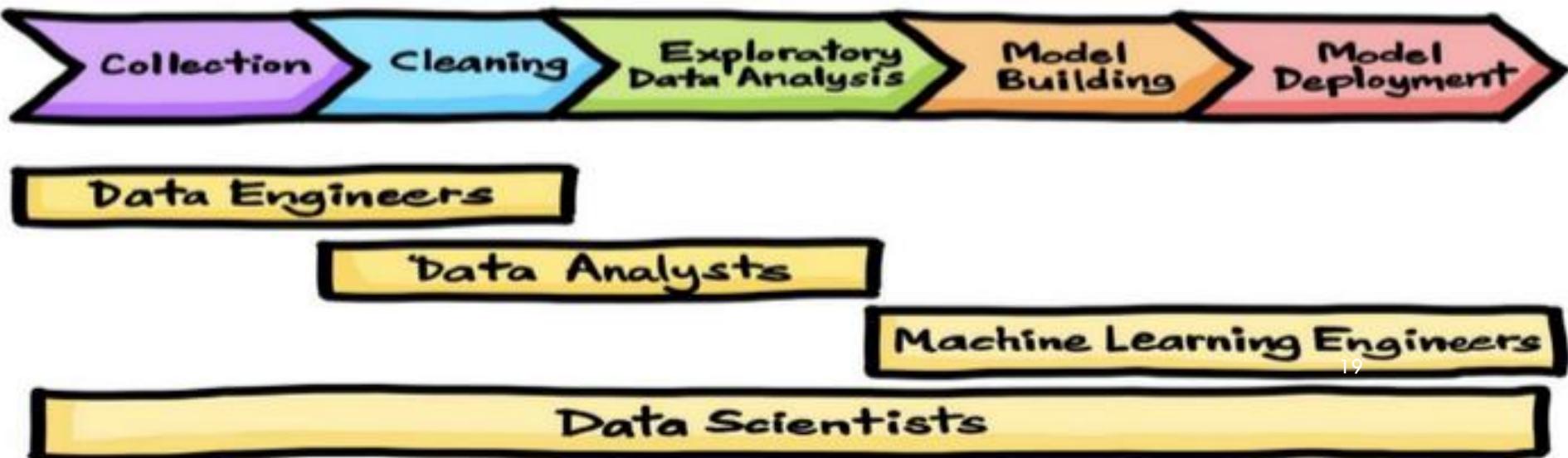


Machine Learning Process

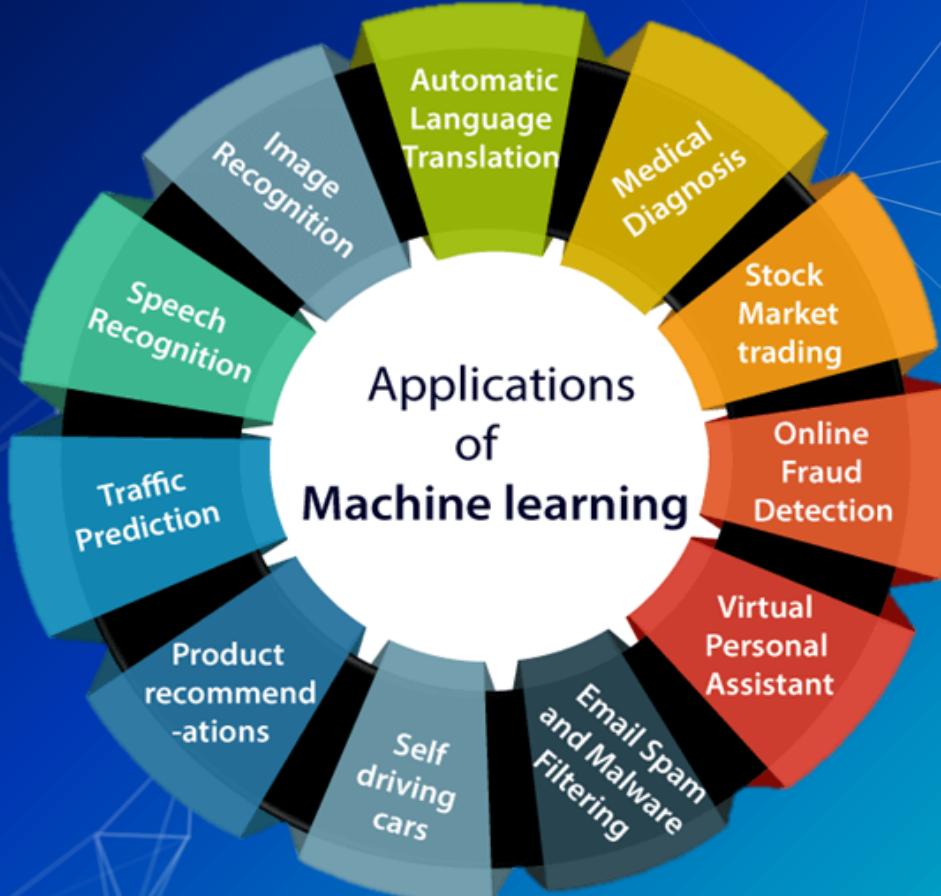


THE DATA SCIENCE PROCESS

Pipeline



Applications of Machine Learning



Course content

- Getting Started with Python
- Mathematical Library with NumPy (Linear Algebra)
- Data Preprocessing and Visualization with Pandas & Matplotlib
- Supervised Machine Learning (Classification & Regression)
- Unsupervised Machine Learning (Clustering)
- Advanced Machine Learning
- Introduction to Deep Learning

why use Python in machine learning

- Wide range of libraries
- Easy to use and learn
- Large community support
- Interoperability with other languages
- Scalability

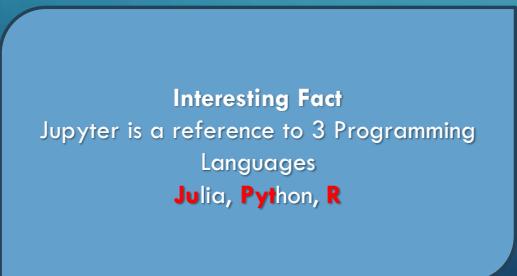


PYTHON (JUPYTER NOTEBOOK)

Most Preferred Programming Language in 2023

Recently used in most applications of the Computer Science

Jupyter Notebook is an open-source web app and the preferred IDE of Python since it organizes your code very well. Additionally, some similar platforms support it (Kaggle – Google Collab)



File Edit View Insert Cell Kernel Widgets Help

A large, stylized orange and grey Jupyter logo watermark is positioned in the center-right area of the interface.
File Run Cell Code Voila

In [13]: # importing libraries

```
from __future__ import print_function
from ipywidgets import interact, interactive, fixed, interact_manual
from IPython.core.display import display, HTML

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import plotly.express as px
import folium
import plotly.graph_objects as go
import seaborn as sns
import ipywidgets as widgets
```

In [14]: # loading data right from the source:

```
death_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_
confirmed_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_c
recovered_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_c
country_df = pd.read_csv('https://raw.githubusercontent.com/CSSEGISandData/COVID-19/web-data/data/cases_country.csv')
```

In [15]: confirmed_df.head()

In [16]: recovered_df.head()

In [17]: death_df.head()

In [18]: country_df.head()

1. Scientifics Computing Libraries



Pandas

(Data structures & tools)

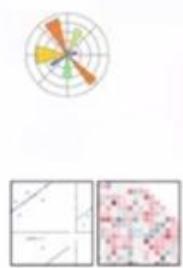
NumPy

(Arrays & matrices)

SciPy

(Integrals, solving differential equations, optimization)

2. Visualization Libraries



Matplotlib

(plots & graphs, most popular)

Seaborn

(plots : heat maps, time series, violin plots)

3. Algorithmic libraries



Scikit-learn

(Machine Learning : regression, classification,...)

Statsmodels

(Explore data, estimate statistical models, and perform statistical tests.)



TensorFlow



TENSORFLOW & PYTORCH



ANACONDA®

USING ANACONDA PLATFORM

- Download Anaconda From The Given Link
[\(\[anaconda.com\]\(http://anaconda.com\)\)](http://anaconda.com)

Unleash Your Innovation

The world's most popular platform to develop and deploy secure Python solutions, faster

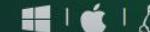


Start Coding Now



Download

Get Additional Installers



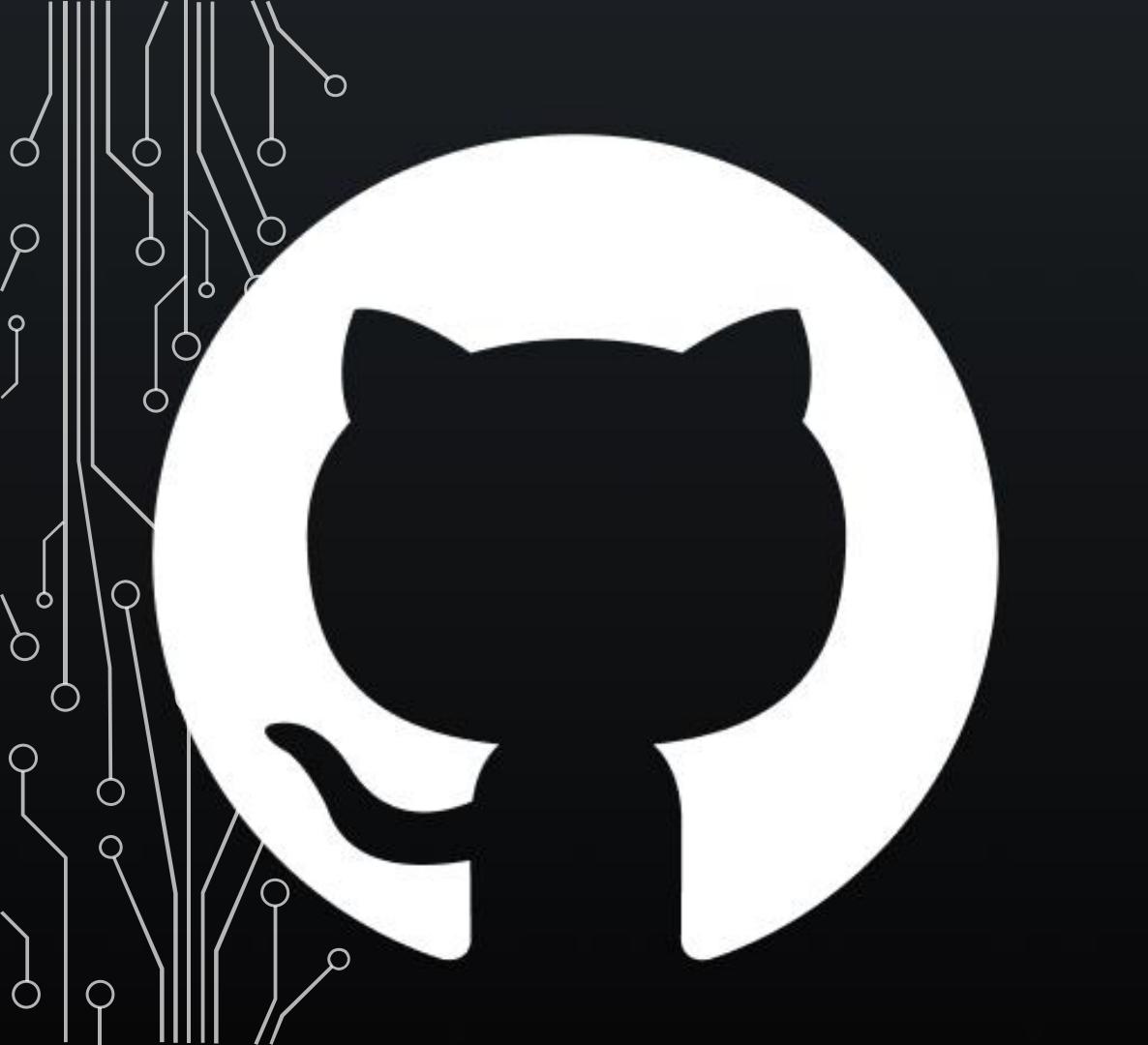
SETUP YOUR KAGGLE ACCOUNT

The Biggest Data science and Machine Learning Platform in the world

It provide a **HUGE** number of Datasets freely to be used in the projects

There are some free courses for short hours of some topics such as (Data Visualization, Data Analysis, Machine Learning, Advanced Machine Learning ... etc)

The Kaggle logo, consisting of the word "kaggle" in a lowercase, sans-serif font.



PREPARE
YOUR
GITHUB

Important Links

- [Home - UCI Machine Learning Repository](#)
- [Numpy_Cheat_Sheet.pdf \(datacamp.com\)](#)
- [Pandas_Cheat_Sheet.pdf \(datacamp.com\)](#)
- [Matplotlib_Cheat_Sheet.pdf \(datacamp.com\)](#)



Time for Practicality

THANKS!

Any questions?