

# Machine Learning for Astronomy

**What made you  
think ML?  
Is it necessary?**

A decorative graphic at the top of the slide featuring a network of interconnected nodes and lines, resembling a molecular or neural network structure. The nodes are represented by circles of varying sizes, some solid and some dashed, connected by thin lines. A central node is highlighted with a larger dashed circle.

“

***Artificial Intelligence, deep learning,  
machine learning—whatever you’re  
doing if you don’t understand it—learn  
it. Because otherwise, you’re going to be  
a dinosaur within 3 years.***

~Mark Cuban



# Why Machine Learning?

# Data Science Job Roles



Data Scientist



Data Analyst



Data Architect



Data Engineer



Statistician



Database  
Administrator



Business Analyst



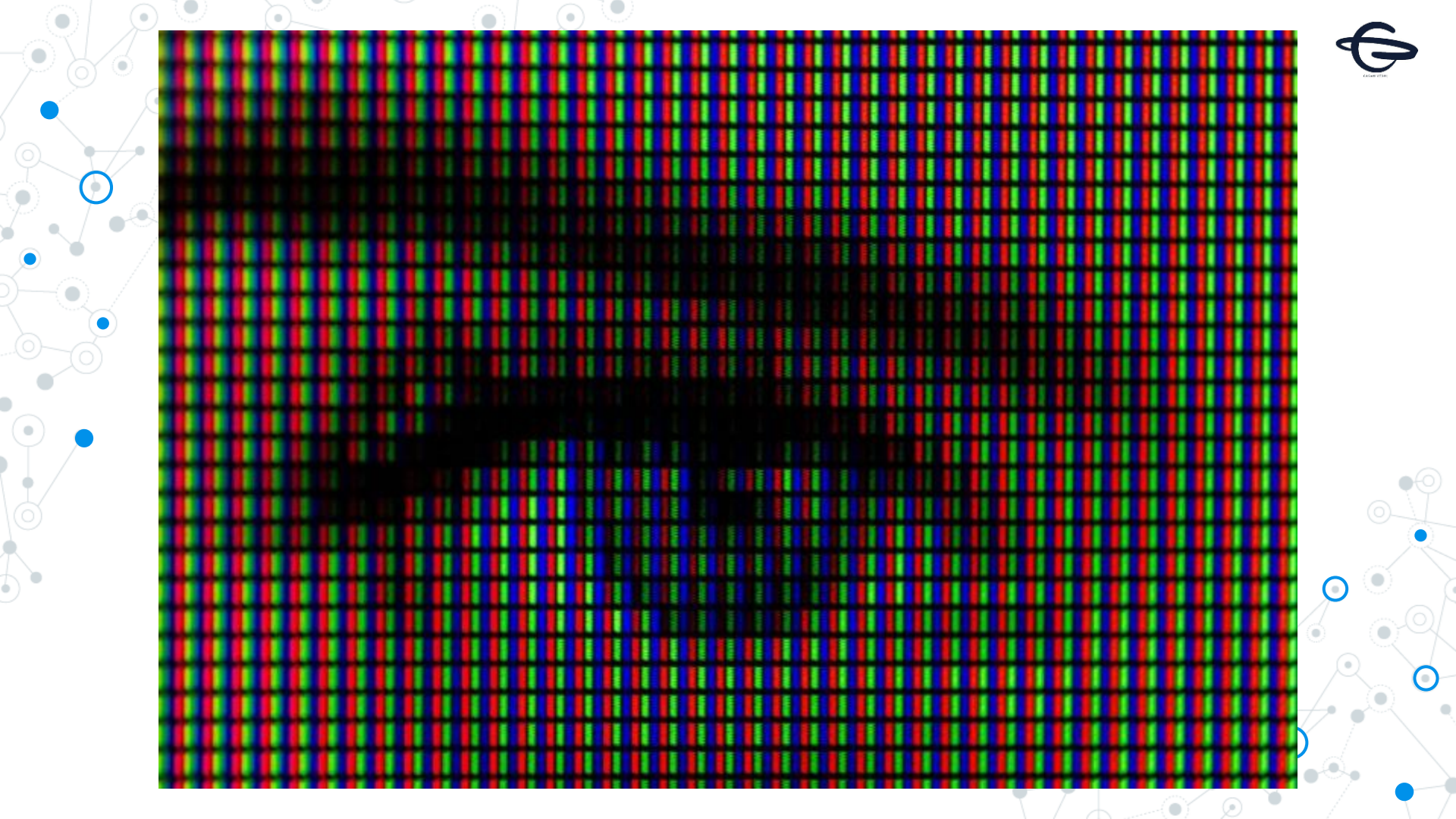
Data & Analytics  
Manager



# **Necessity is the mother of Invention**











08	02	22	97	38	15	00	40	00	75	04	05	07	78	52	12	50	77	87	28
49	49	99	40	17	81	18	57	60	87	17	40	98	43	69	45	04	56	62	00
81	49	31	73	55	79	14	29	93	71	40	67	55	85	30	03	49	13	36	65
52	70	95	23	04	60	11	42	65	84	65	56	01	32	56	71	37	02	36	91
22	31	16	71	51	62	05	89	41	92	36	54	22	40	40	28	66	33	13	80
24	47	33	60	99	03	45	02	44	75	33	53	78	36	84	20	35	17	12	50
32	98	81	28	64	23	67	10	26	38	40	67	59	54	70	66	18	38	64	70
67	26	20	68	02	62	12	20	95	63	94	39	63	08	40	91	66	49	94	21
24	55	58	05	66	73	99	26	97	17	78	78	96	83	14	88	34	89	63	72
21	36	23	09	75	00	76	44	20	45	35	14	00	61	33	97	34	31	33	95
78	17	53	28	22	75	31	67	15	94	03	80	04	62	16	14	09	53	56	92
16	39	05	42	96	35	31	47	55	58	88	24	00	17	54	24	36	29	85	57
86	56	00	48	35	71	89	07	05	44	44	37	44	60	21	58	51	54	17	58
19	80	81	68	05	94	47	69	28	73	92	13	86	52	17	77	04	89	55	40
04	52	08	83	97	35	99	16	07	97	57	32	16	26	26	79	33	27	95	66
13	36	68	87	57	62	20	72	03	46	33	67	46	55	12	32	63	93	53	69
04	42	16	73	35	85	39	11	24	94	72	18	08	46	29	32	40	62	76	36
20	69	36	41	72	30	23	88	34	65	98	69	82	67	59	85	74	04	36	16
20	73	35	29	78	31	90	01	74	31	49	71	48	85	81	16	23	57	05	54
01	70	54	71	83	51	54	69	16	92	33	48	61	43	52	01	89	19	62	48

What the computer sees

image classification

82% cat  
15% dog  
2% hat  
1% mug

Viewpoint variation



Scale variation



Deformation



Occlusion



Illumination conditions



Background clutter



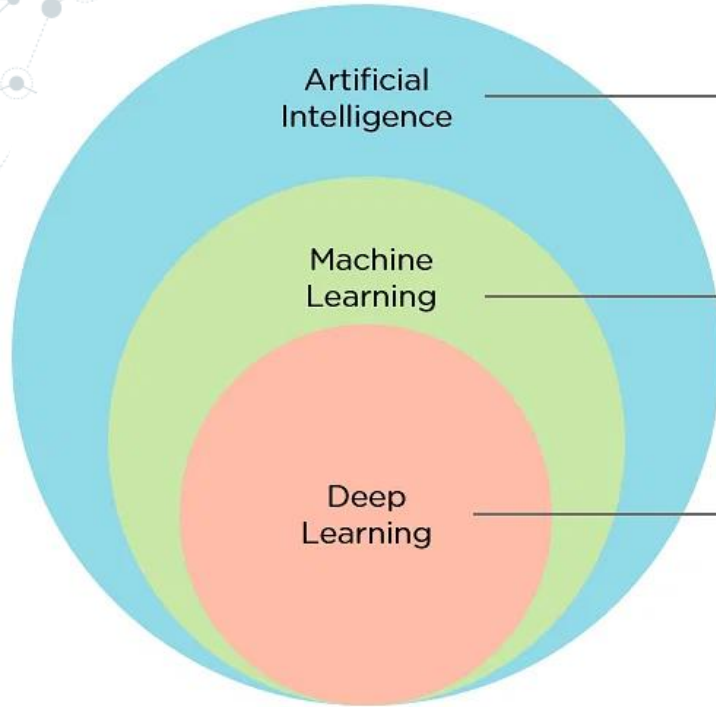
Intra-class variation





# Teach them the pattern!



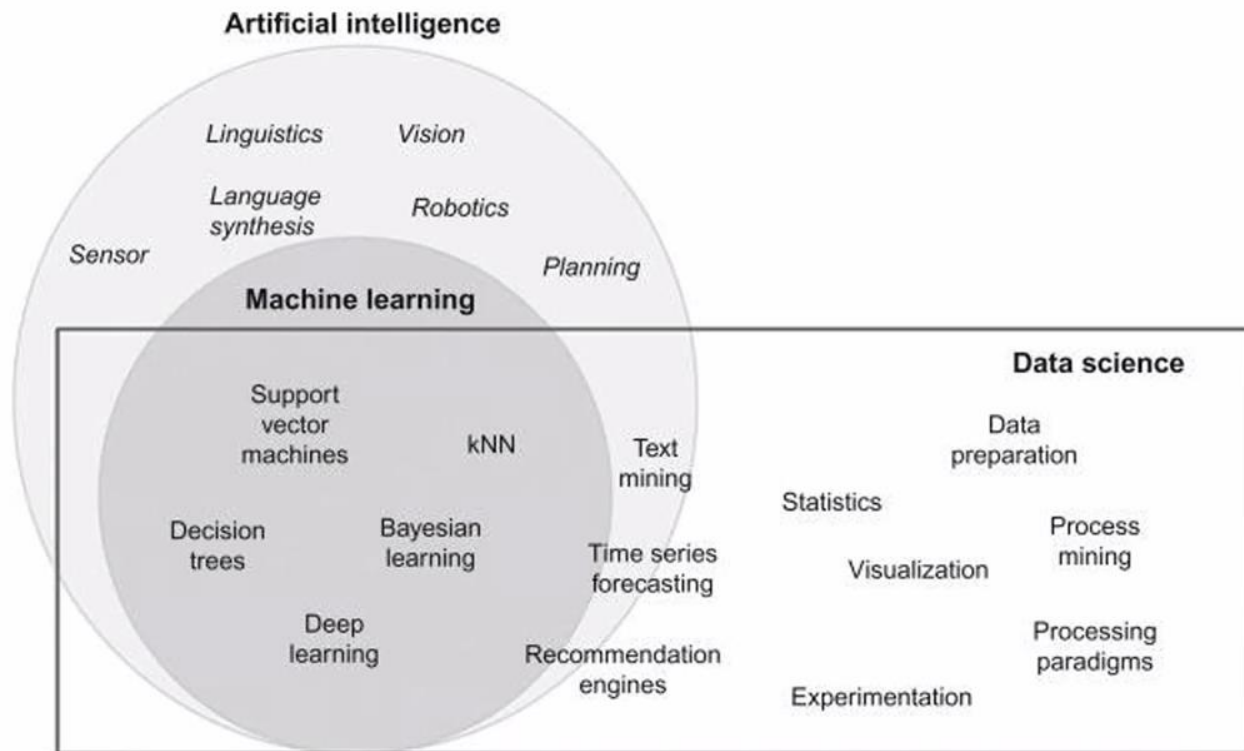


Ability of a machine to imitate intelligent human behavior

Application of AI that allows a system to automatically learn and improve from experience

Application of Machine Learning that uses complex algorithms and deep neural nets to train a model

# AI vs ML vs DS





# AI

## What is Artificial Intelligence?

- The process of **imparting data, information, and human intelligence** to machines.
- The main goal is to develop self-reliant machines that can think and act like humans.
- Mimic human behavior and perform tasks by learning and problem-solving.
- AI systems simulate natural intelligence to solve complex problems.



# AI

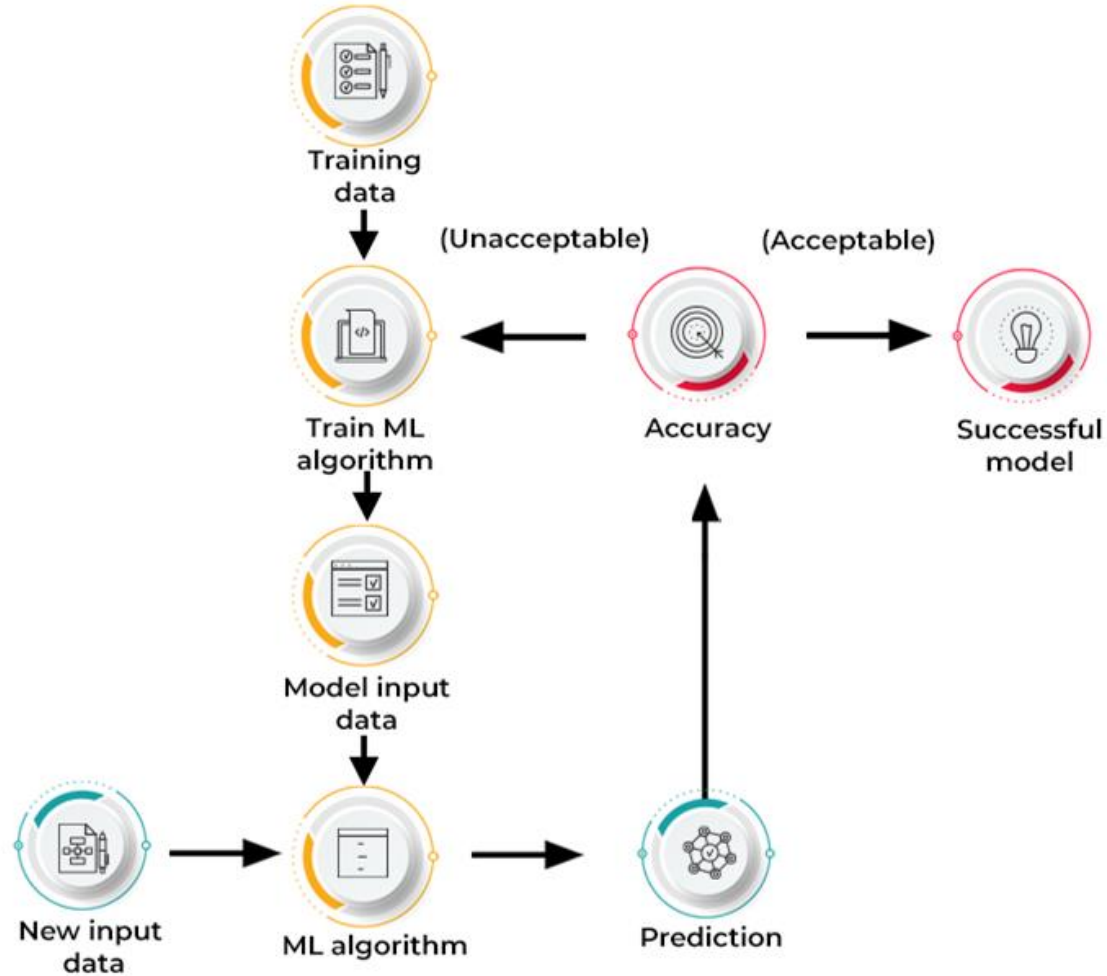
## Applications of Artificial Intelligence:

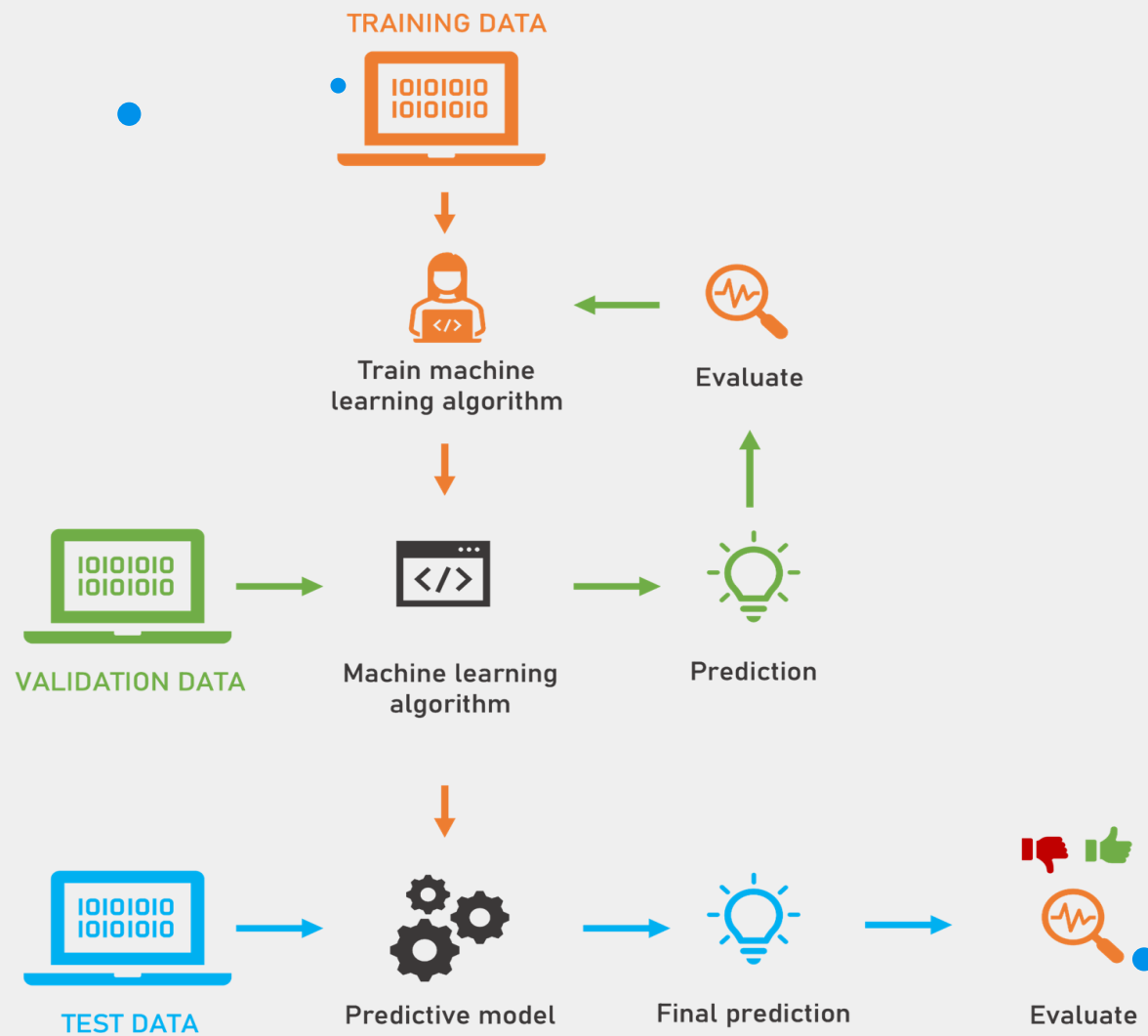
- Machine Translation such as Google Translate
- Self Driving Vehicles such as Google's Waymo
- AI Robots such as Sophia and Aibo
- Speech Recognition applications like Apple's Siri or OK Google

# Machine Learning



# HOW DOES MACHINE LEARNING WORK?







# ML

## Some Technical Terms:

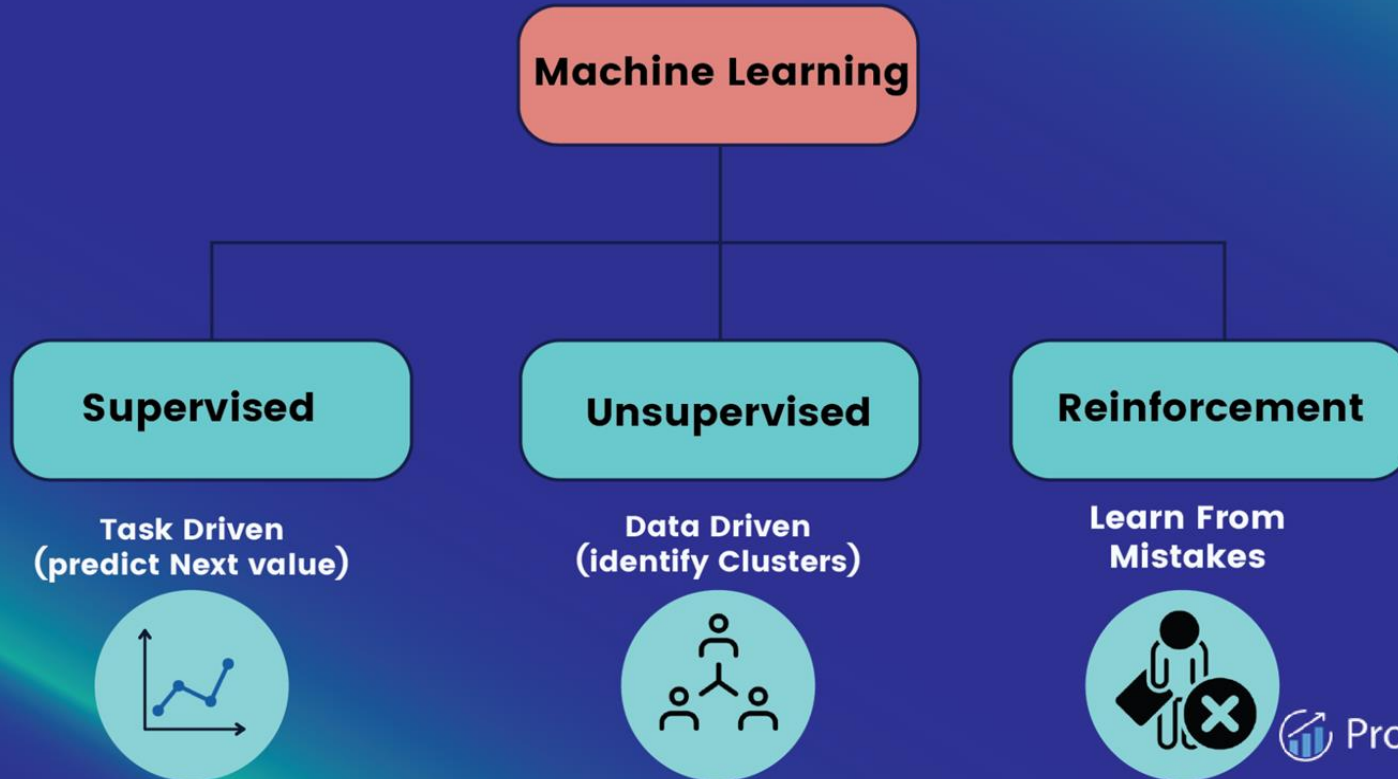
- Model - a magical blackbox that learns and predicts.
- Learning - The process of teaching the model.
- Performance/Accuracy - Marks obtained by the model in exams.
- Loss - A mathematical value so that model regrets his bad performance.
- Backpropagation - The action of model is improving his mistakes.

# ML

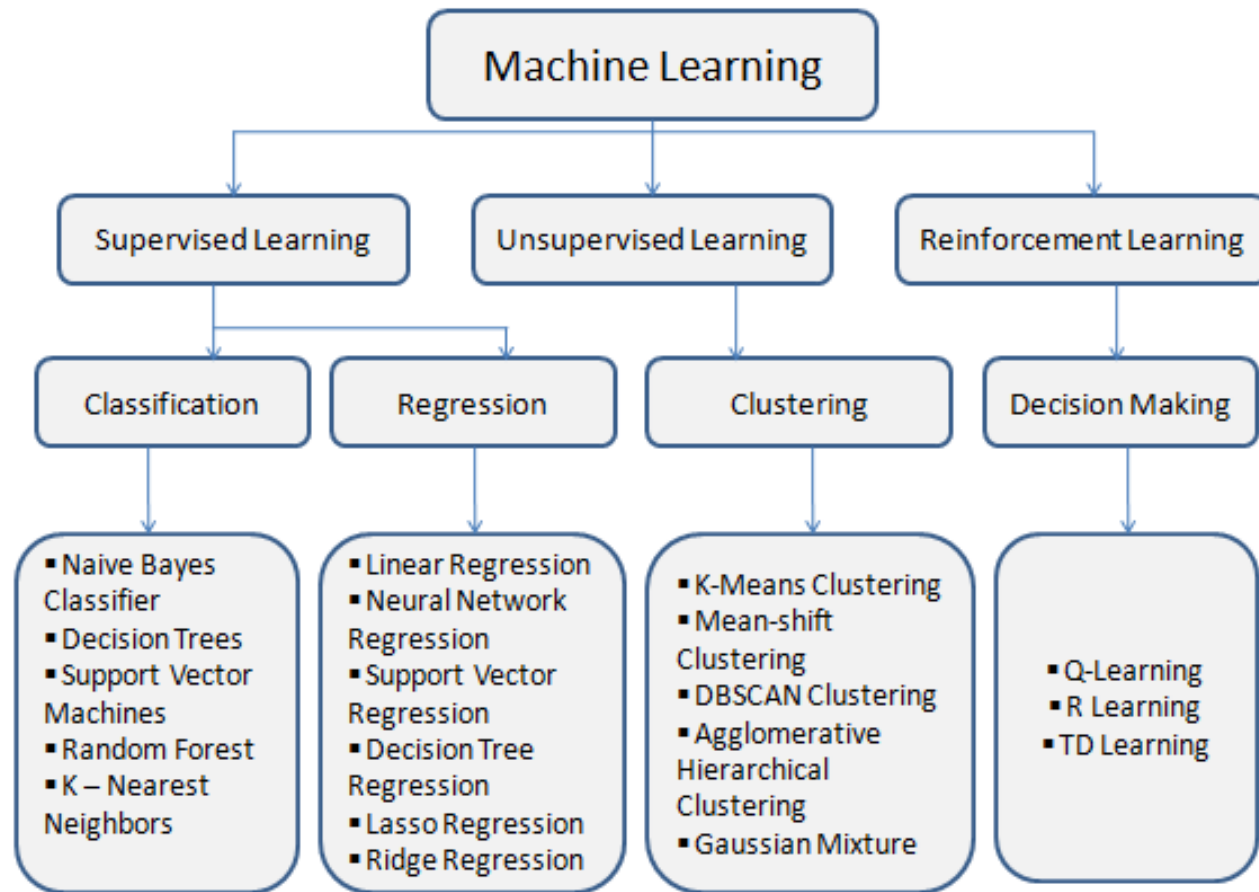
## Demystifying Technical Terms:

- Model - a function - ranging from equation of line to DNN
- Hyperparameters - some variable constants used in the function
- Learning - The process of tuning the parameters for good performance
- Performance - (A Number) Closeness of the prediction with true value
- Loss - ( " ) Inverse or Negative of Performance (some exaggeration intended)
- Backpropagation - Changing the parameters of model based on the current loss

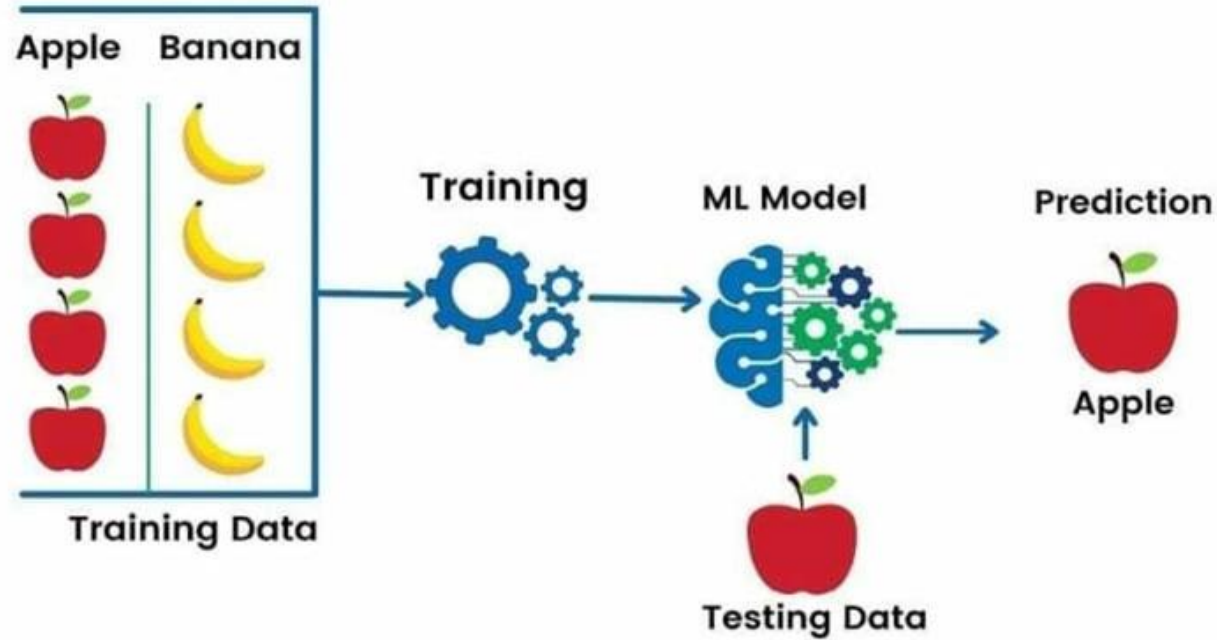
## Types of Machine Learning



# Types of ML



# Supervised ML





# Supervised ML



## Regression



What will be the temperature tomorrow?

84°



Fahrenheit

## Classification



Will it be hot or cold tomorrow?

COLD

HOT



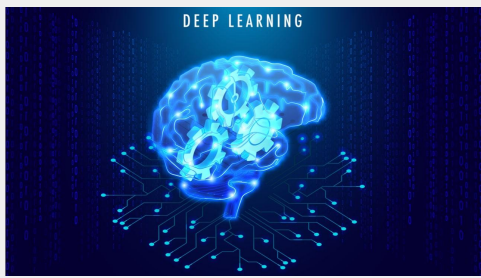
Fahrenheit

# Supervised ML

	Regression	Classification
<b>Description</b>	A regression model seeks to predict a continuous quantity.	A classification model seeks to predict some class label.
<b>Type of algorithm</b>	Supervised learning algorithm	Supervised learning algorithm
<b>Type of response variable</b>	Continuous	Categorical
<b>How to assess model fit</b>	Root mean squared error	Percentage of correct classifications

# DEEP LEARNING





The most powerful thing in this sector.  
What can it do ?





The most powerful thing in this sector.  
What can it do ?

This:

As I enter the gates of IIT, a new chapter begins,  
But soon I realize, it's not all sunshine and grins,  
Countless assignments and deadlines, oh what a feat,  
Determined to succeed, I refuse to accept defeat

Every day is a struggle, with workloads so high,  
Frustration and stress, make me want to cry,  
Gone are the days of leisure, now it's do or die,  
However, amidst the chaos, small joys do lie.

In the company of friends, we laugh and we share,  
Jokes and stories, without a single care,  
Kindred spirits, united in this journey so rare,  
Life at IIT, is a rollercoaster affair.

In the company of friends, we laugh and we share,  
Jokes and stories, without a single care,  
Kindred spirits, united in this journey so rare,  
Life at IIT, is a rollercoaster affair.

Memories are made, as we learn and we grow,  
Navigating challenges, with each passing flow,  
Opportunities abound, as we reap what we sow,  
Perseverance and hard work, are the seeds that we throw.

Quiet moments of reflection, as we look to the sky,  
Realizing our potential, as we aim to fly high,  
Success may be elusive, but we continue to try,  
Taking each day as it comes, with a hopeful sigh.

Underneath the pressure, we discover our strength,  
Venturing forth with courage, to go any length,  
With determination and grit, we'll succeed at any expense,  
X marks the spot, as we reach for excellence.

Yes, life at IIT may be tough and intense,  
Zero regrets though - it all makes perfect sense.





The most powerful thing in this sector.  
What can it do ?  
This:

Astronomy and ML, two fields so vast,  
But when they come together, the results are a blast,  
Countless stars and galaxies, waiting to be classed,  
Data-driven insights, from the first to the last.

Every observation, a piece of the puzzle,  
Fitting together, with no need for muscle,  
Gone are the days of manual toil and hustle,  
However, the algorithms still need some subtle.

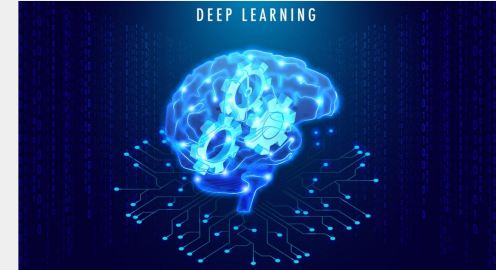
In the realm of space, we explore and we learn,  
Journeys to distant worlds, at every turn,  
Knowledge gained through data, as we discern,  
Life beyond Earth, a possibility to yearn.

Machine learning models, trained on the data we  
collect,  
Navigating the cosmos, with precision and intellect,  
Opportunities abound, as we detect and select,  
Planets and stars, with characteristics we expect.

Quietly working behind the scenes,  
Realizing our dreams and fulfilling our means,  
Success in astronomy and ML, it seems,  
Taking us closer to understanding the universe's schemes.

Underneath it all, lies a world of complexity,  
Venturing forth with curiosity and dexterity,  
With data and algorithms, we unlock the galaxy's perplexity,  
Xploring the unknown, with astronomy and ML's connectivity.

Yes, the journey may be long and intense,  
Zeroing in on the mysteries of the universe makes perfect  
sense.





The most powerful thing in this sector.  
What can it do ?  
This: GANS :)

**A**stronomy and ML, two fields so vast,  
**B**ut when they come together, the results are a blast,  
**C**ountless stars and galaxies, waiting to be classed,  
**D**ata-driven insights, from the first to the last.

**E**very observation, a piece of the puzzle,  
**F**itting together, with no need for muscle,  
**G**one are the days of manual toil and hustle,  
**H**owever, the algorithms still need some subtle.

**I**n the realm of space, we explore and we learn,  
**J**ourneys to distant worlds, at every turn,  
**K**nowledge gained through data, as we discern,  
**L**ife beyond Earth, a possibility to yearn.

**M**achine learning models, trained on the data we  
collect,  
**N**avigating the cosmos, with precision and intellect,  
**O**pportunities abound, as we detect and select,  
**P**lanets and stars, with characteristics we expect.

**Q**uietly working behind the scenes,  
**R**ealizing our dreams and fulfilling our means,  
**S**uccess in astronomy and ML, it seems,  
**T**aking us closer to understanding the universe's schemes.

**U**nderneath it all, lies a world of complexity,  
**V**enturing forth with curiosity and dexterity,  
**W**ith data and algorithms, we unlock the galaxy's perplexity,  
**X**ploring the unknown, with astronomy and ML's connectivity.

**Y**es, the journey may be long and intense,  
**Z**eroing in on the mysteries of the universe makes perfect  
sense.





**DATA**



**ALGORITHM**

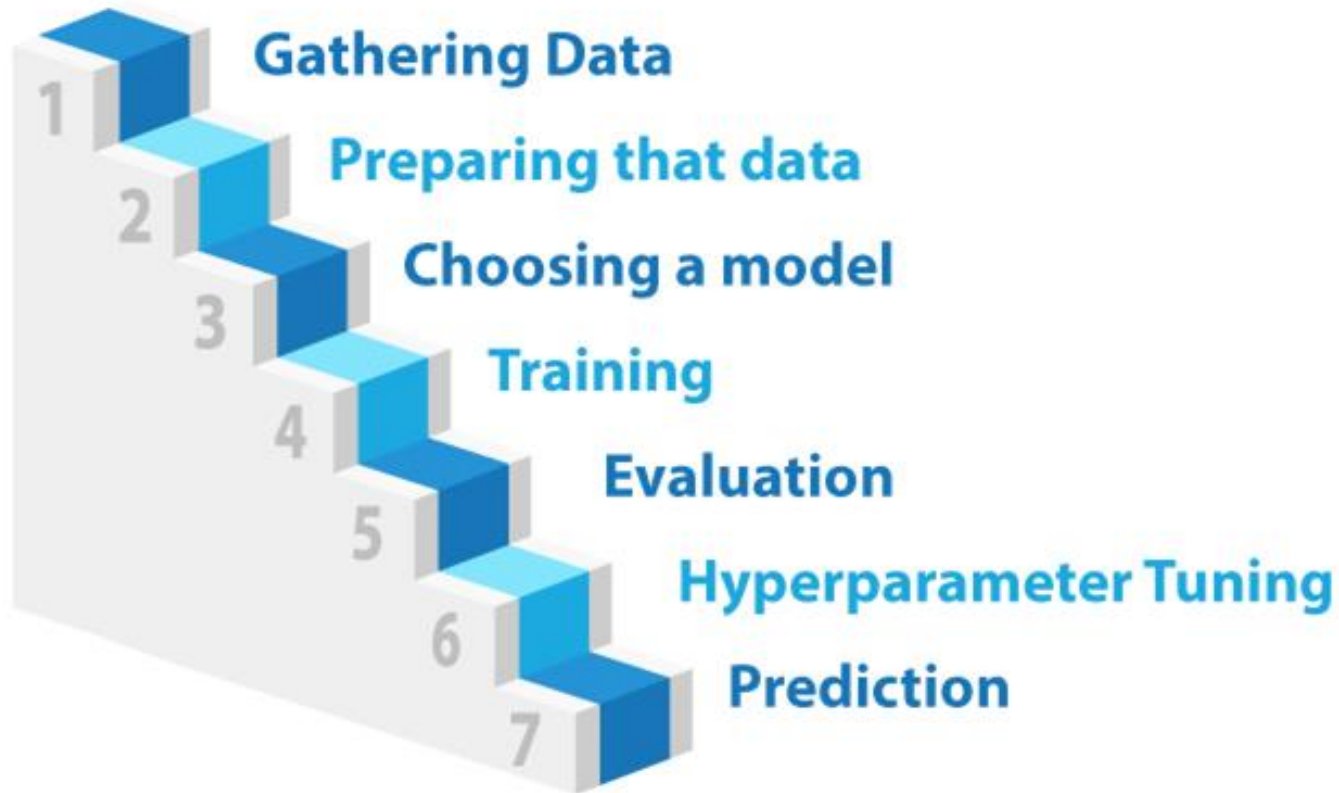


**COOL STUFF**




**LEARNING**

# 7 steps of Machine Learning



# **Some Libraries used in ML and Data Analysis**

A network diagram in the top-left corner consisting of a web of interconnected nodes. Some nodes are solid blue circles, while others are white circles with blue outlines. They are connected by thin grey lines.

Who are you?



"Hello there, I'm Numpy! I'm a powerful library for numerical computing in Python. I specialize in handling large multi-dimensional arrays and matrices, and offer a wide range of mathematical functions to work with them. I'm used in a variety of fields, from machine learning and data analysis to scientific computing and engineering. With me, you can easily perform complex mathematical operations and data manipulations with lightning-fast speed!"



Numerical  
Python





Hi there, I'm Pandas! I'm a popular library for data manipulation and analysis in Python. I'm built on top of Numpy, so I can handle large datasets with ease. I provide an intuitive and flexible data structure called the DataFrame, which lets you easily manipulate and analyze tabular data. With me, you can clean, transform, and explore data quickly and efficiently. I'm used in a variety of industries, from finance and marketing to healthcare and academia. With my powerful tools, you'll be able to make data-driven decisions and insights in no time!

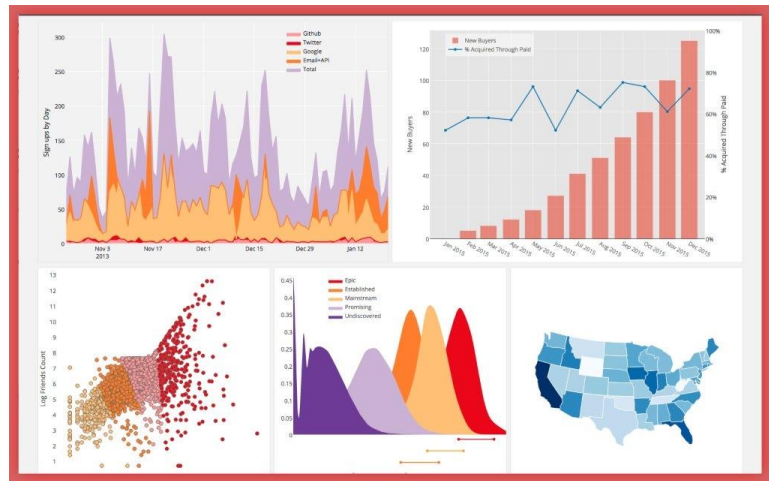
# Pandas



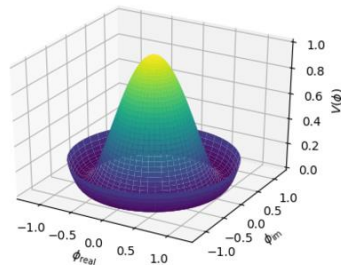
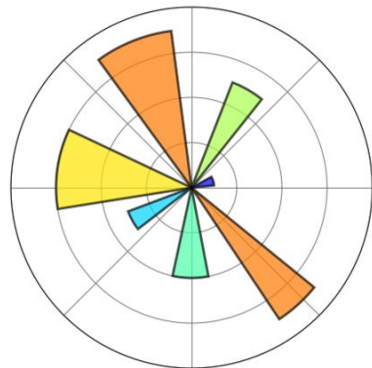
"Hello there, I'm Matplotlib! I'm a powerful library for creating static, animated, and interactive visualizations in Python. I offer a wide range of plotting functions and customization options to help you create beautiful and informative visualizations. I'm used in a variety of fields, from data analysis and machine learning to scientific research and engineering. With me, you can easily turn your data into stunning visual stories!"

# matplotlib

Maths +  
Plotting +  
Library

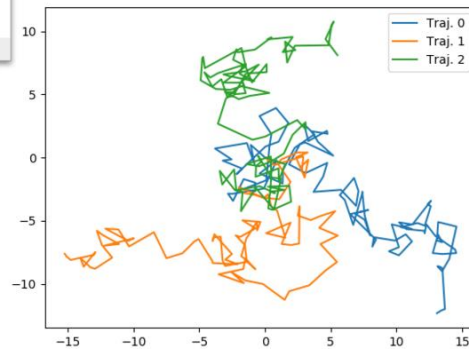


"He  
libr  
inte  
wic  
cus  
bea  
use  
and  
you



```
1 import matplotlib.pyplot as plt
2 import numpy as np
3
4 from matplotlib.cm import jet as colormap
5 from matplotlib.ticker import NullFormatter, MultipleLocator
6
7 t, w, r = zip((0.1, 0.4, 1), (0.9, 0.3, 5), (1.7, 0.5, 7), (2.7, 0.6, 6),
8             (3.5, 0.3, 3), (4.5, 0.4, 4), (5.5, 0.3, 7))
9
10 fig, ax = plt.subplots(subplot_kw={'polar': True})
11 bars = ax.bar(t, r, width=w, bottom=0.0, lw=2, edgecolor='black', zorder=2)
12
13 for r, bar in zip(r, bars):
14     bar.set_facecolor(colormap(r / 9.0))
15     bar.set_alpha(0.7)
16
17 ax.yaxis.set_major_locator(MultipleLocator(2))
18
19 for axis in (ax.xaxis, ax.yaxis):
20     axis.set_major_formatter(NullFormatter()) # no tick labels
21
22 ax.set_ylim([0, 8])
23 ax.grid(True)
24
25 plt.show()
```

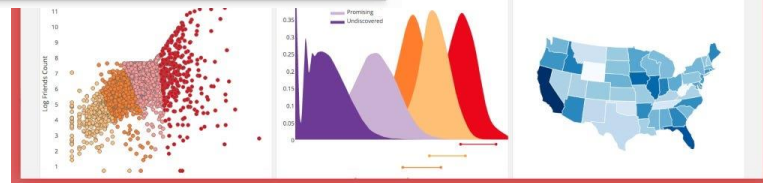
Python Large des tabulations : 4 Lig 24, Col 1 INS

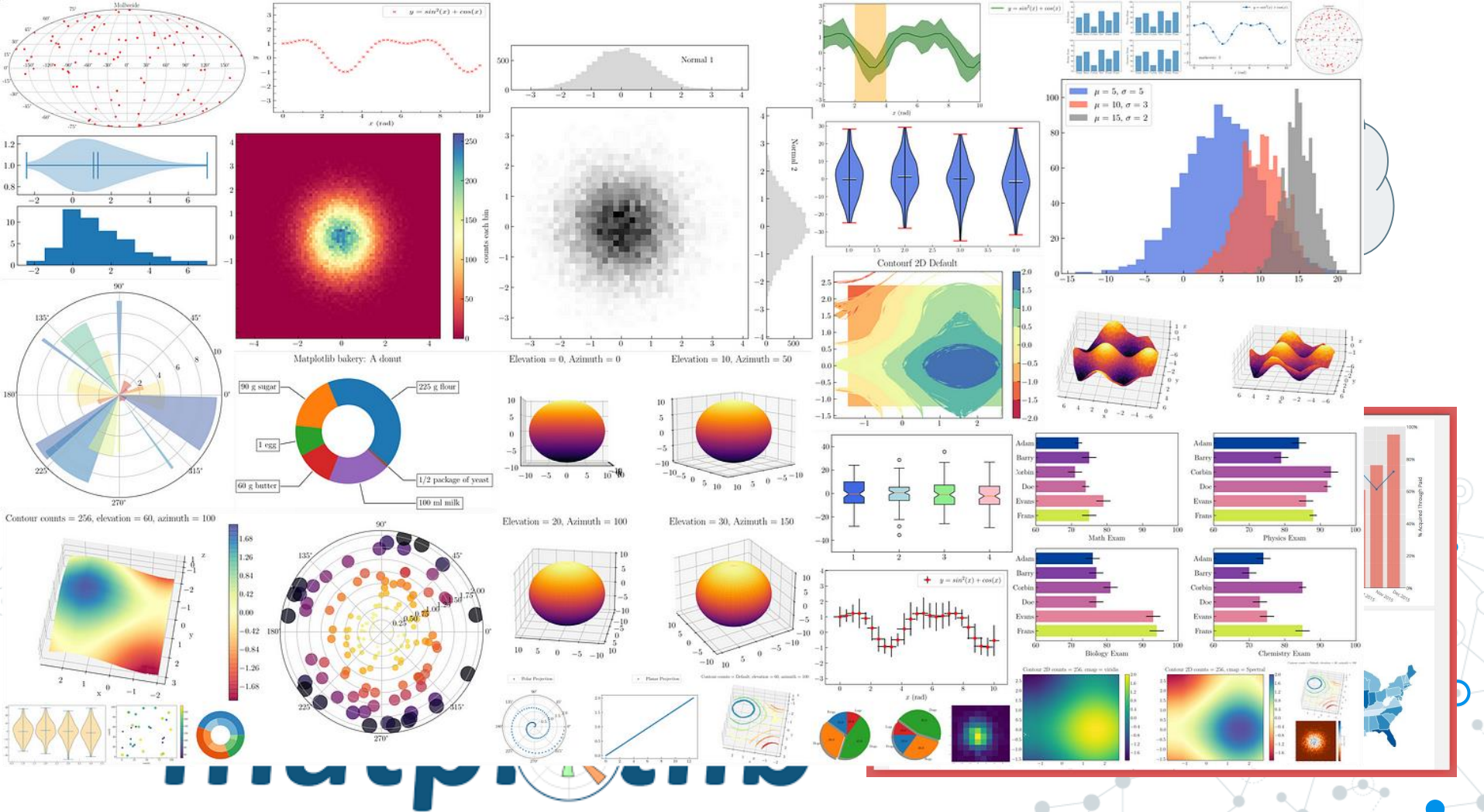


Maths +  
Plotting +  
Library



# matplotlib





# Les'Go NumPy



## Hands-On





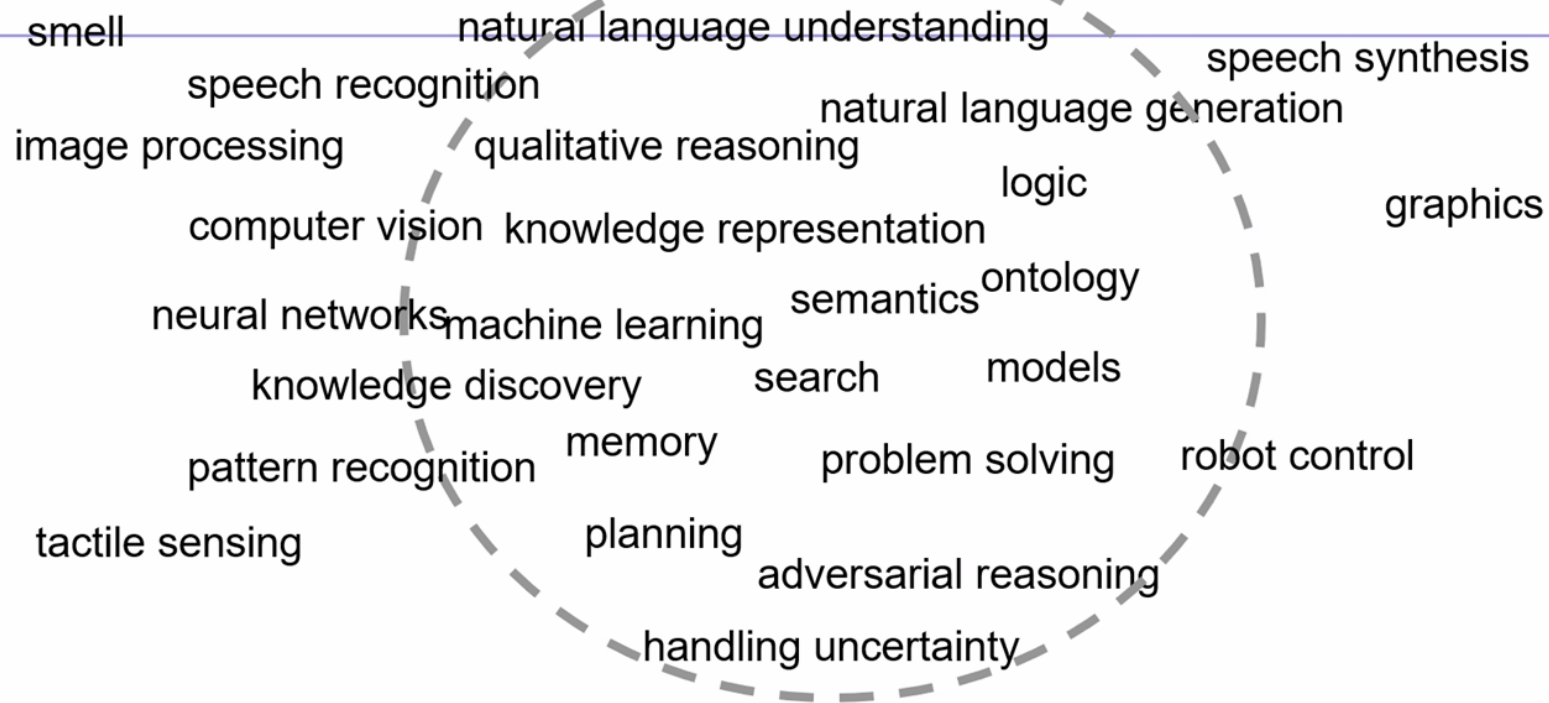
# **Basic Python?**

# **Python Installed?**

# **Anaconda ?**



# Topics in AI



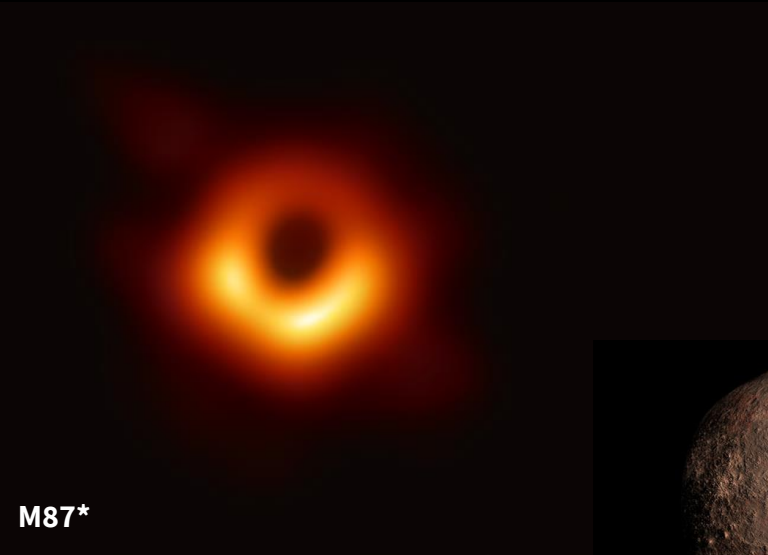
Sense → Deliberate → Act

Source: Deepak Khemani, A First Course in Artificial Intelligence

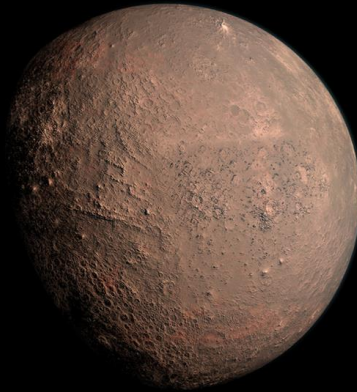




# ML in Astronomy and Astrophysics



M87\*



GJ 357 b



Seagull Nebula

# Stellar Classification Dataset - SDSS DR-17

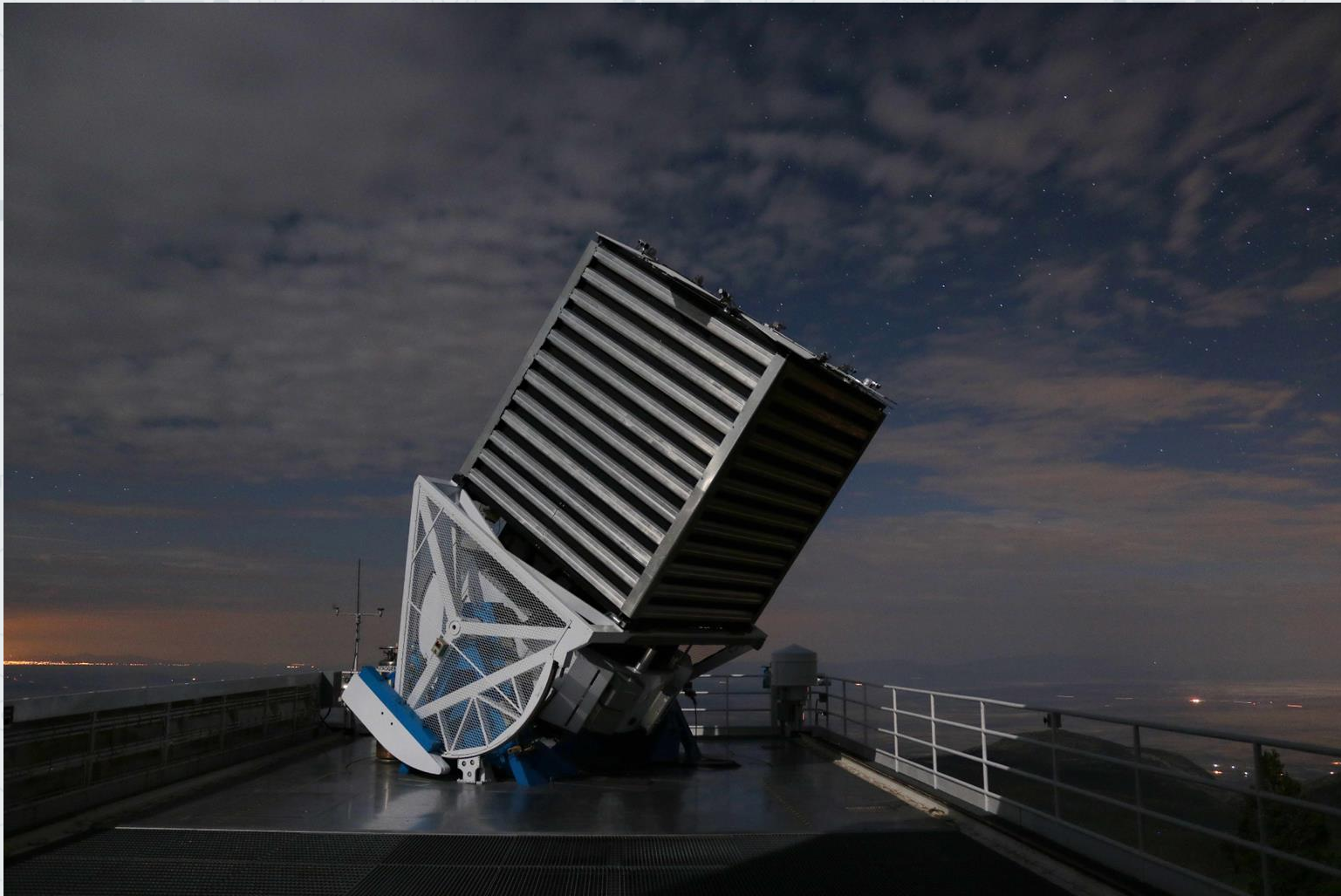
- ◎ Data consists of 100,000 observations of space taken by the SDSS.
- ◎ Each obs. described by 17 feature columns and 1 class column.
- ◎ Aimed at classifying stars, galaxies, and quasars based on their spectral characteristics.

# 17 features and 1 column for class

```
1 # Now we will read the data from the csv file
2 df = pd.read_csv('star_classification.csv')
3
4 # Now we will print the first 5 rows of the data
5 df.head()
```

	obj_ID	alpha	delta	u	g	r	i	z	run_ID	rerun_ID	cam_col	field_ID	spec_obj_ID	class	redshift	plate	MJD	fiber_ID
0	1.237661e+18	135.689107	32.494632	23.87882	22.27530	20.39501	19.16573	18.79371	3606	301	2	79	6.543777e+18	GALAXY	0.634794	5812.0	56354.0	171.0
1	1.237665e+18	144.826101	31.274185	24.77759	22.83188	22.58444	21.16812	21.61427	4518	301	5	119	1.176014e+19	GALAXY	0.779136	10445.0	58158.0	427.0
2	1.237661e+18	142.188790	35.582444	25.26307	22.66389	20.60976	19.34857	18.94827	3606	301	2	120	5.152200e+18	GALAXY	0.644195	4576.0	55592.0	299.0
3	1.237663e+18	338.741038	-0.402828	22.13682	23.77656	21.61162	20.50454	19.25010	4192	301	3	214	1.030107e+19	GALAXY	0.932346	9149.0	58039.0	775.0
4	1.237680e+18	345.282593	21.183866	19.43718	17.58028	16.49747	15.97711	15.54461	8102	301	3	137	6.891865e+18	GALAXY	0.116123	6121.0	56187.0	842.0

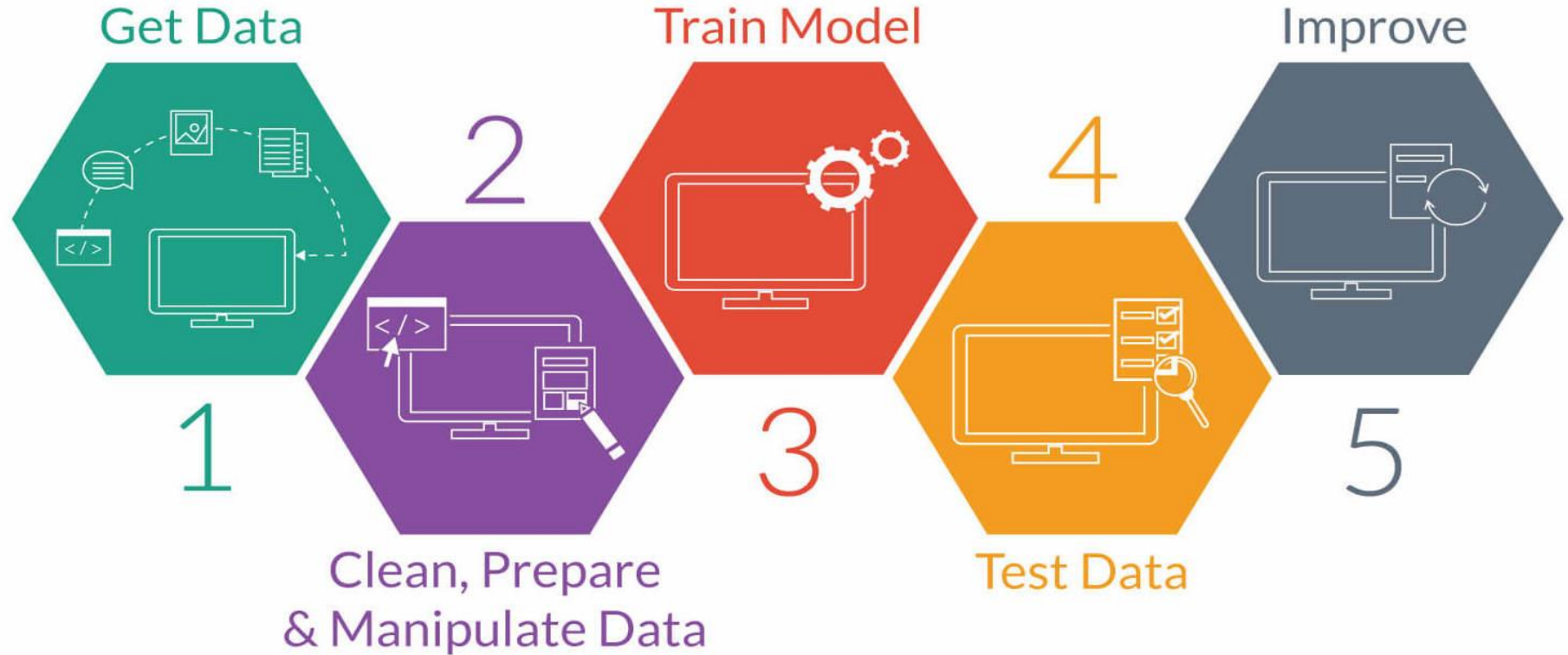
# SDSS telescope at night



[https://www.sdss4.org/wp-content/uploads/2016/07/sds\\_s\\_gaulme1.jpg](https://www.sdss4.org/wp-content/uploads/2016/07/sds_s_gaulme1.jpg)

Image Credit: Patrick Gaulme

# ML Workflow





# Logistic Regression Model



@dataaspirant.com

## Logistic Regression

- ◎ Instead of predicting exactly 0 or 1, logistic regression generates a probability—a value between 0 and 1. Depending on the **‘threshold’** we classify it.
- ◎ In Logistic regression, we **fit** an "S" shaped logistic function, which predicts two values (0 or 1).



```
from sklearn.linear_model import LogisticRegression  
  
model=LogisticRegression()
```

**That's it !?**

**What've you been telling then?**



# fit() and predict()

## **fit()**

Used to train the model on the given data and parameters.

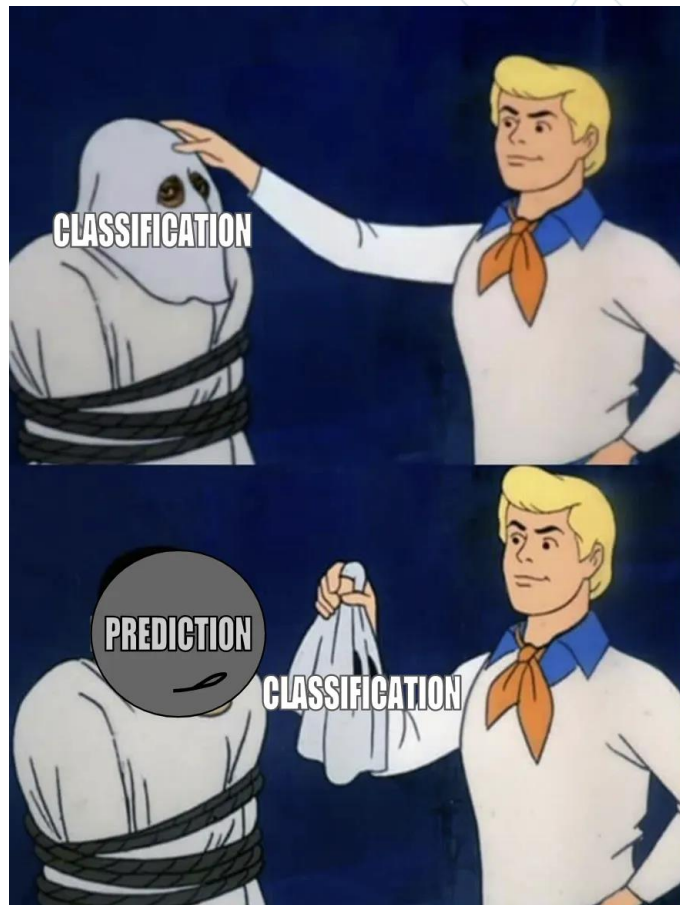
## **predict()**

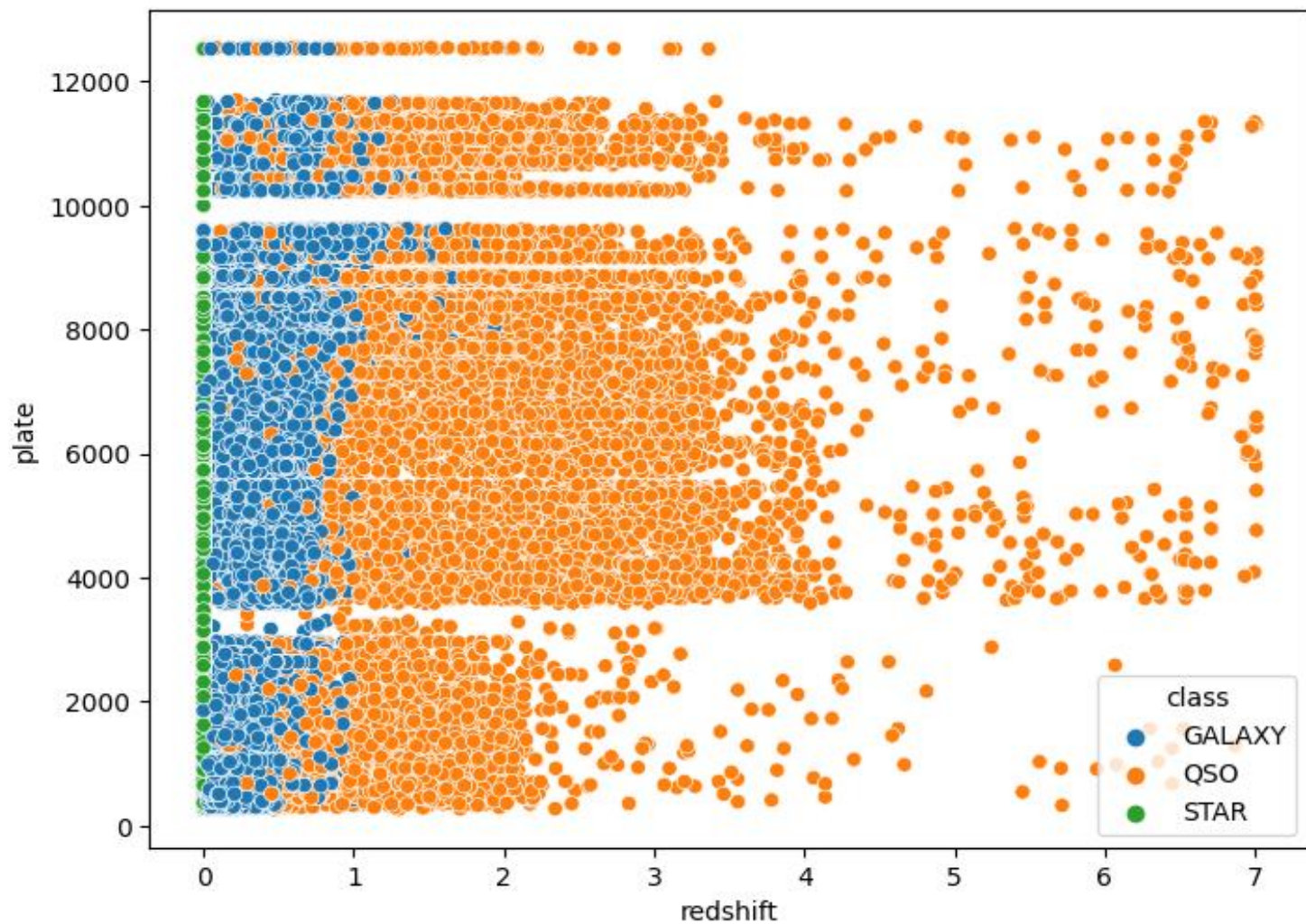
Used to make predictions on new data using the trained model

# From prediction to classification

$$P(y = 1 \mid \mathbf{x}, \mathbf{w}, b) = \sigma(\mathbf{w}^T \mathbf{x} + b)$$

$$y_{\text{pred}} = \text{round}(P(y=1 \mid \mathbf{x}, \mathbf{w}, b))$$





hue: class

# Intuition behind these results

How is just one feature  
working so well in place of  
using all 17?



# Approximate redshift values:

## Quasars

0.1 - 7 or higher

## Galaxies

0.01 - 0.1

## Stars

-0.001 - 0.001 or  
less



# Thanks!

## Any questions?

You can reach us at:

[cs20b029@iittp.ac.in](mailto:cs20b029@iittp.ac.in) &

[cs20b009@iittp.ac.in](mailto:cs20b009@iittp.ac.in)



Please let us know your feedback 😊





# A small announcement

## Club Team Form

...was released recently.

Will be reopened till EOD !

