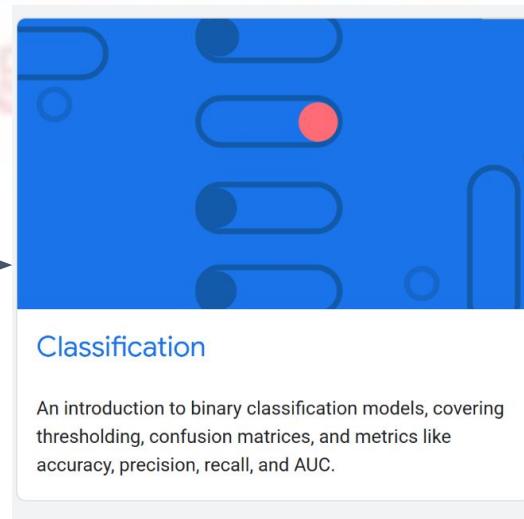


# TECHCRUSH ARTIFICIAL INTELLIGENCE BOOTCAMP

Facilitator: Hammed Obasekore  
September 12th, 2025

## Recap



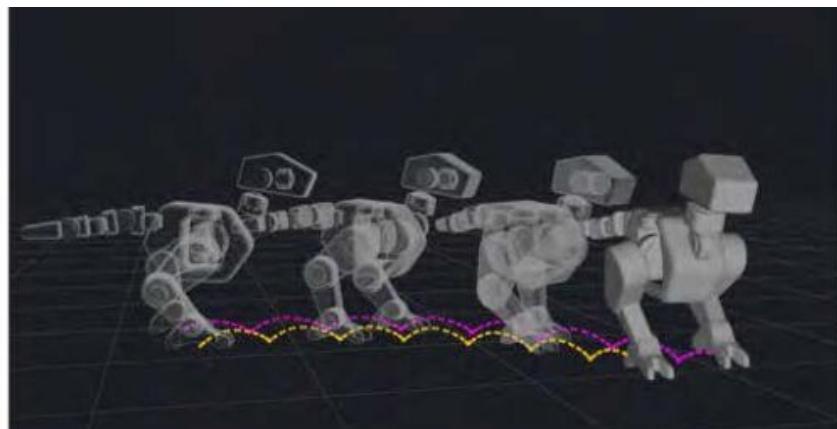
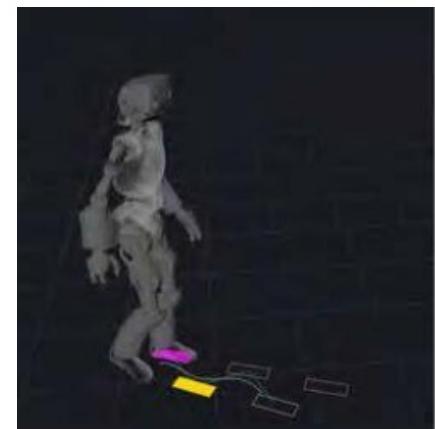
*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

## Essence of Consulting Research Outcome for AI

Importance of checking research to solve AI problem

Case Study: Disney

- [Disney Research](#)
- [Optimization Animatronics](#)



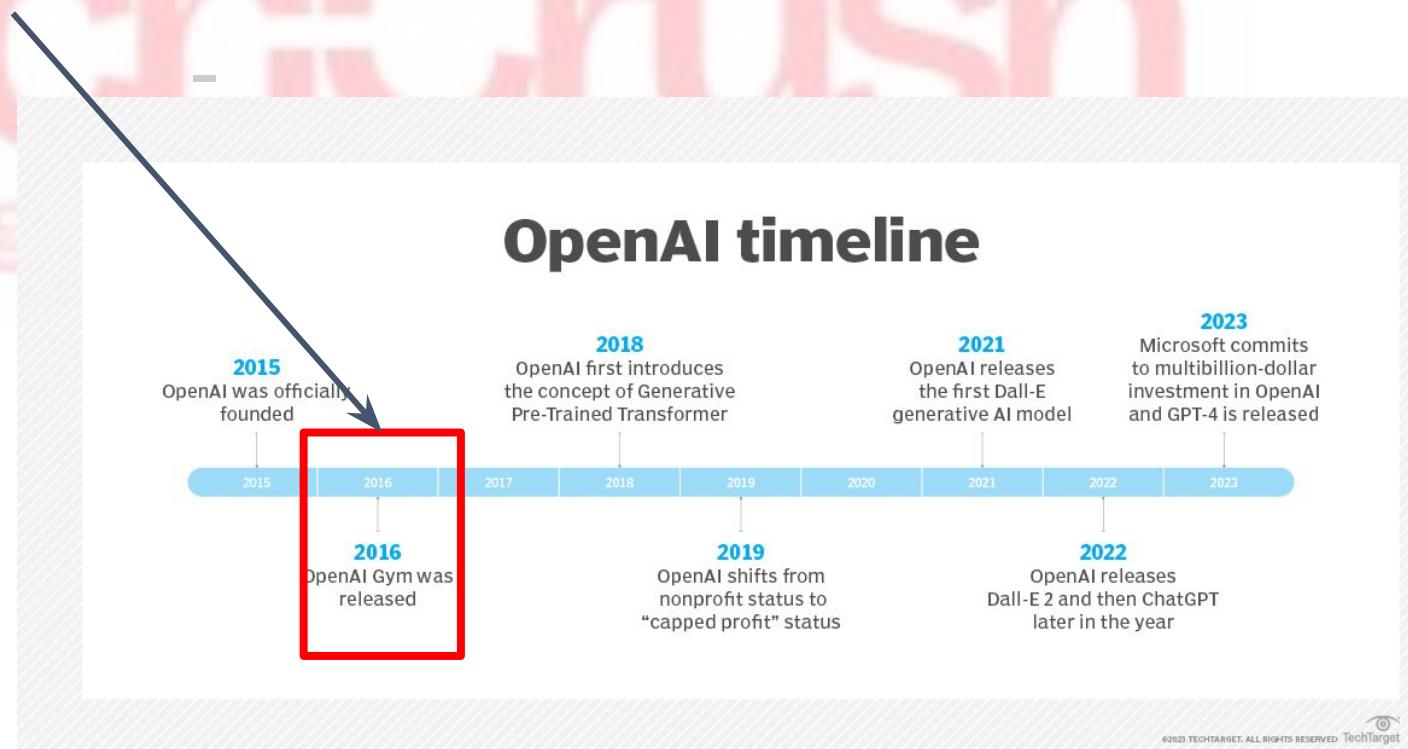
*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

## Essence of Consulting Research Outcome for AI

Importance of checking research to solve AI problem

Case Study: OpenAI

- OpenAI gym: Became Standard for Reinforcement Learning



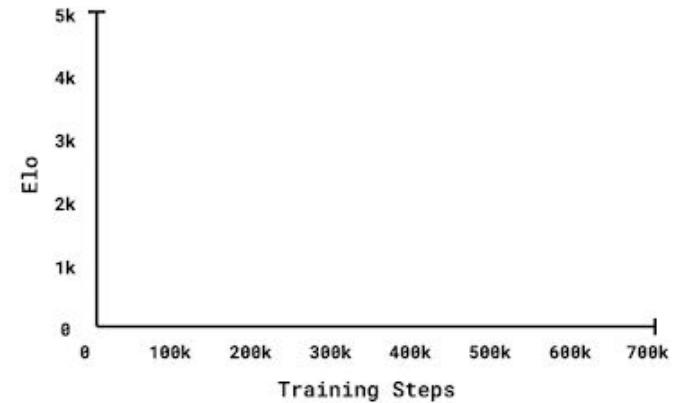
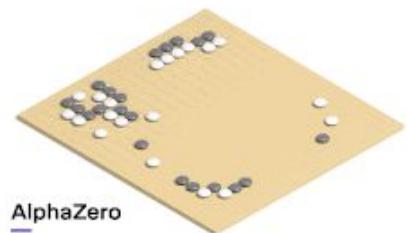
*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

## Essence of Consulting Research Outcome for AI

Importance of checking research to solve AI problem

Case Study: Google

- [Google Research](#)
- [AI Chess](#)
- 



## Essence of Consulting Research Outcome for AI

Importance of checking research to solve AI problem  
Case Study: Myself



Invitation to Interview for: Reinforcement Learning Expert Needed for dm\_control (mujoco) Stable-Baselines3 project   

donotreply@upwork.com   Fri, Sep 12, 7:27 AM (19 hours ago)    

**upwork**

Congrats! You have been invited to submit a proposal!

Read more about the job below and submit a proposal if you are interested.

Hello!

I'd like to invite you to take a look at the job I've posted. Please submit a proposal if you're available and interested.

[Reinforcement Learning Expert Needed for dm\\_control \(mujoco\) Stable-Baselines3 project](#)

I am seeking a skilled freelancer to implement reinforcement learning using Stable-Baselines3 and a dm\_control environment for a UAV platform based on... [more](#)

Payment Type  Fixed Price 

*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

## Essence of Consulting Research Outcome for AI

Importance of checking research to solve AI problem

If you can't check research

Wait for

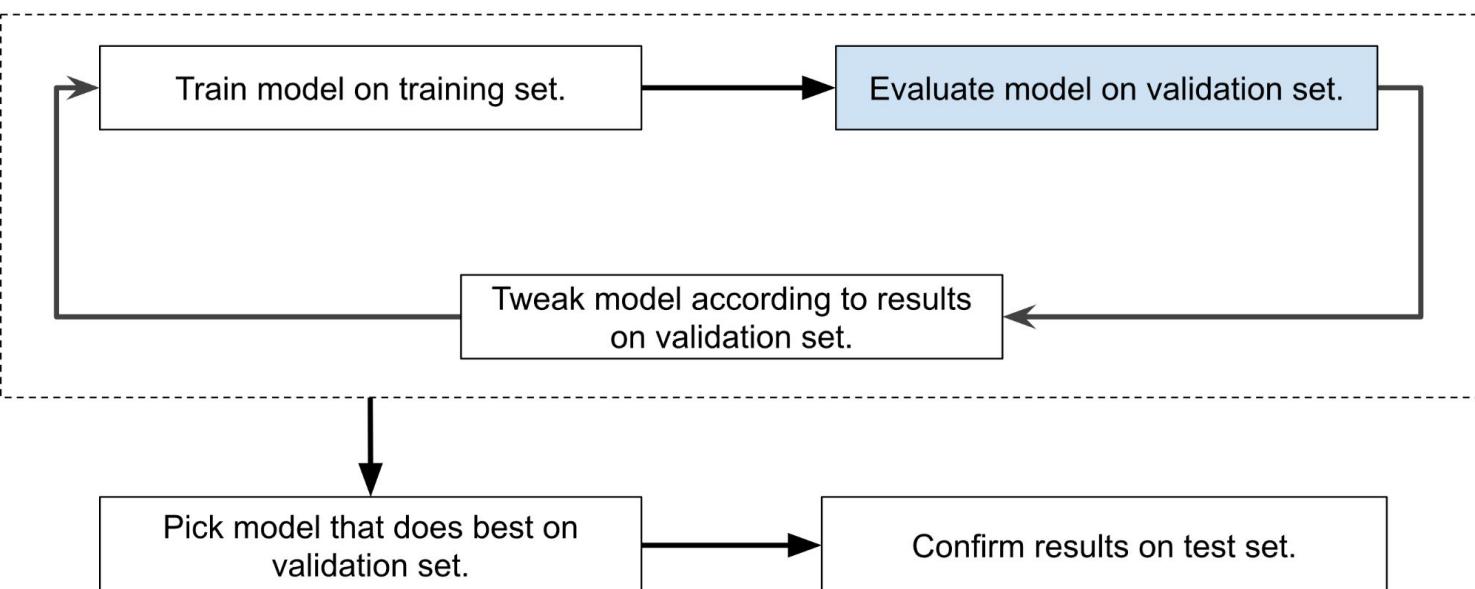
- Layman implementation
- Subscription based services
- Abstraction to Language
- Application Programming Interface (API)

## Understanding Data

# Cross-validation



### 4-fold validation ( $k=4$ )



*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

## Classification: Practice - 2

# Binary Classification

## Understanding Dataset

### Rice Dataset Commeo and Osmancik

Rice Dataset: 2 Class Commeo and Osmancik Rice



Data Card    Code (15)    Discussion (2)    Suggestions (0)

### About Dataset

Usability  
9.38

## Understanding AI Report



International Journal of  
Intelligent Systems and Applications in Engineering

ISSN:2147-6799

[www.ijisae.org](http://www.ijisae.org)

Original Research Paper

### Classification of Rice Varieties Using Artificial Intelligence Methods

Ilkay CINAR<sup>1</sup>, Murat KOKLU<sup>2</sup>

Submitted: 24/08/2019    Accepted : 30/09/2019

**Abstract:** Rice being one of the most widely produced and consumed cereal crops in the world, is also the one of the main sustenance in our country because of its economical and nutritive nature. Rice, starting from farm to our table, goes through some manufacturing steps such as a cleaning process, color sorting and classification. If these stages are to be mentioned briefly, cleaning is the process of separating rice from foreign substances, classification is the process of separating broken ones with sturdy ones; color extraction is the process of separating the stained and striped ones except the whiteness on the rice surface. In this study, a computerized vision system was developed in order to distinguish between two proprietary rice species. A total of 3810 rice grain's images were taken for the two species, processed and feature inferences were made. 7 morphological features were obtained for each grain of rice. With these features, models were created using LR, MLP, SVM, DT, RF, NB and k-NN machine learning techniques and performance measurement values were obtained. Success rates in the classification were obtained 93.02% (LR), 92.86% (MLP), 92.83% (SVM), 92.49% (DT), 92.39% (RF), 91.71% (NB), 88.58% (k-NN). When we look at the results of the success rate of obtain, it is possible to say that the study achieved success.

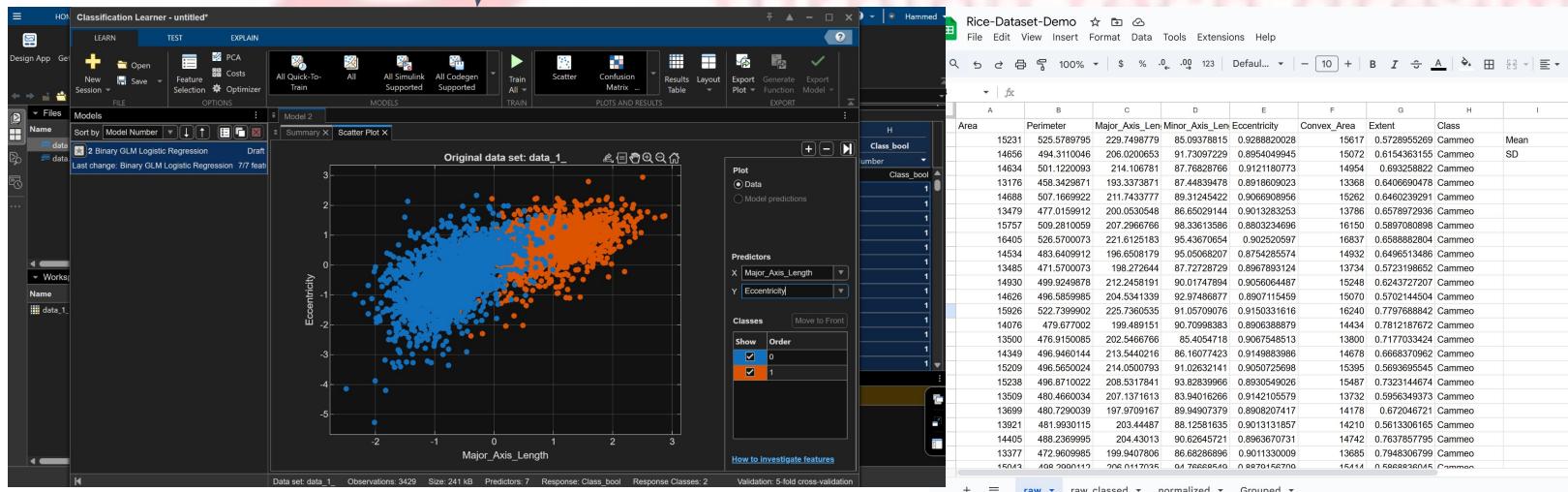
**Keywords:** Classification of rice, computer vision system, image processing, machine learning system

*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

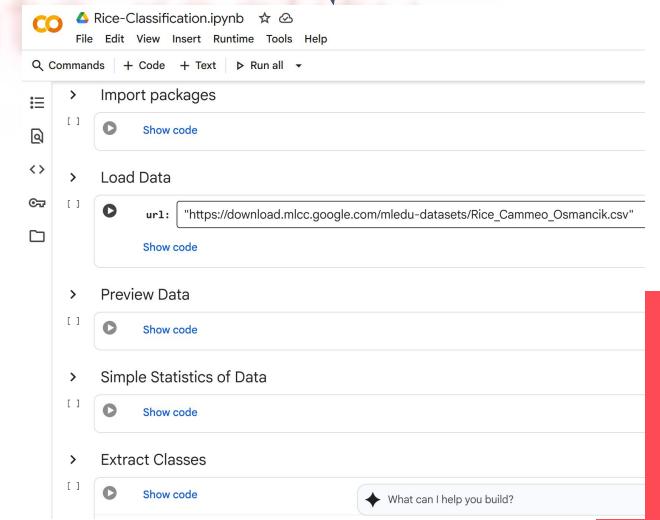
## Classification: Practice - 2

# Rice Dataset

Using MATLAB



Using Python



The figure shows a Jupyter Notebook interface with the file 'Rice-Classification.ipynb'. The code cell contains the following:

```

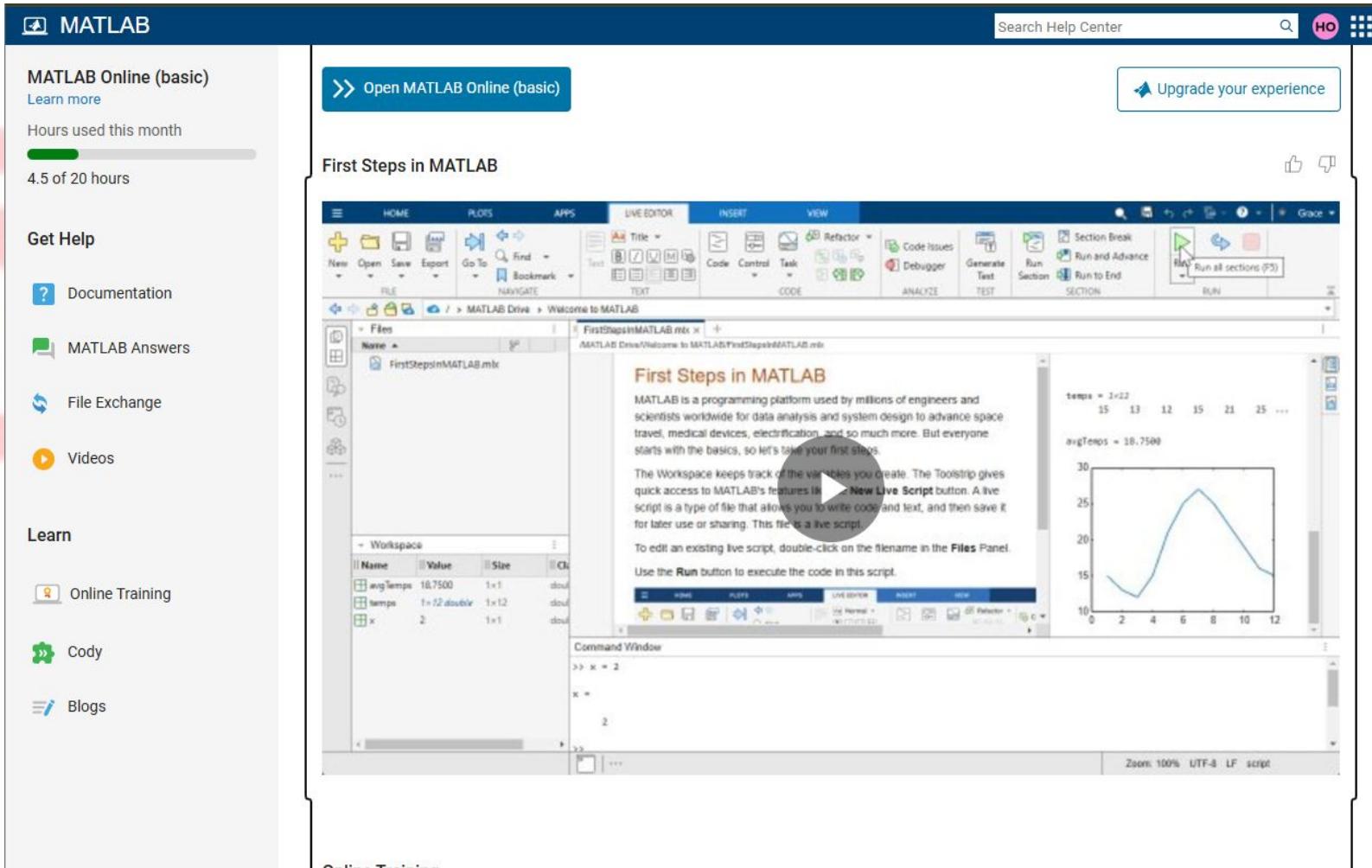
Rice-Classification.ipynb
File Edit View Insert Runtime Tools Help
Commands + Code + Text ▶ Run all ▾
> Import packages
Show code
> Load Data
urllib: "https://download.miccc.google.com/mledu-datasets/Rice_Cammeo_Osmancik.csv"
Show code
> Preview Data
Show code
> Simple Statistics of Data
Show code
> Extract Classes
Show code
What can I help you build?

```

*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

# Understanding Data

# MATLAB



The screenshot shows the MATLAB desktop environment. On the left, there's a sidebar with links for "MATLAB Online (basic)", "Get Help" (Documentation, MATLAB Answers, File Exchange, Videos), and "Learn" (Online Training, Cody, Blogs). The main workspace features the "First Steps in MATLAB" guide, which includes a live script titled "FirstStepsinMATLAB.mlx". The script displays a plot of temperatures over time and shows variable definitions for "temp" and "avgTemp". Below the script, the "Command Window" shows the command `>> x = 2`. The top navigation bar includes tabs for HOME, PLOTS, APPS, LIVE EDITOR (which is selected), INSERT, and VIEW.

*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

# Python & Libraries



pandas Library

scikit-learn

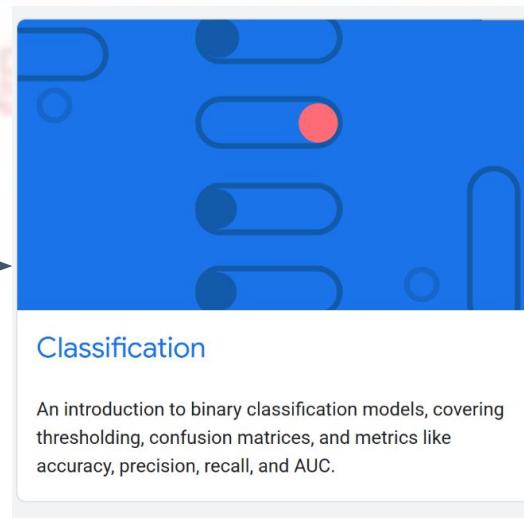


*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

# TECHCRUSH ARTIFICIAL INTELLIGENCE BOOTCAMP

Facilitator: Hammed Obasekore  
September 15th, 2025

## Recap



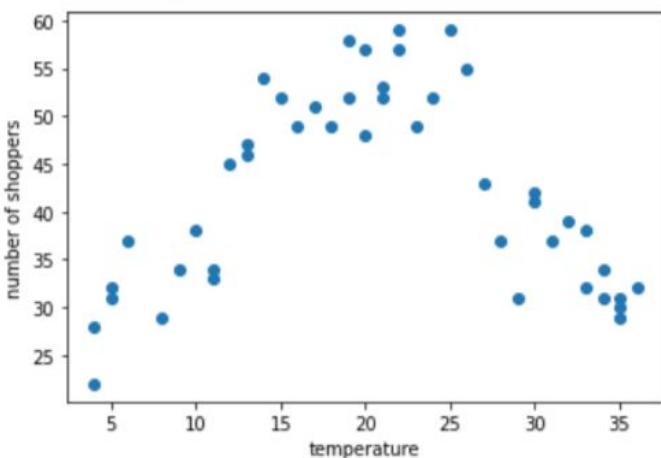
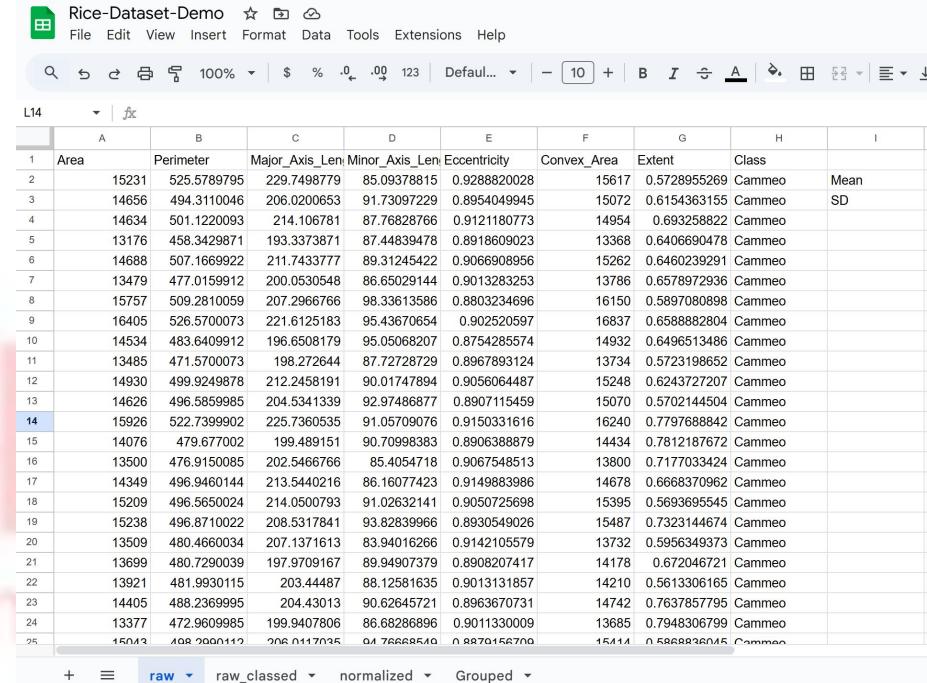
*Disclaimer: This training material belongs to techcrush and shouldn't be shared*

## Understanding Data

# Data Annotation

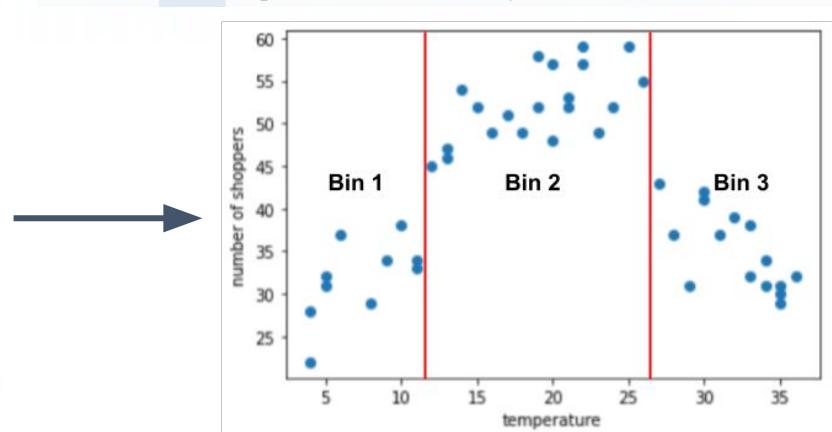
is the process of assigning true value of a data for supervised learning.

- By Column
- By Folder
- By Name
- Bining

Rice-Dataset-Demo

	A	B	C	D	E	F	G	H	I
1	Area	Perimeter	Major_Axis_Len	Minor_Axis_Len	Eccentricity	Convex_Area	Extent	Class	
2	15231	525.5789795	229.7498779	85.09378815	0.9288820028	15617	0.5728955269	Cammeo	Mean
3	14656	494.3110046	206.0200653	91.73097229	0.8954049945	15072	0.6154363155	Cammeo	SD
4	14634	501.1220093	214.106781	87.76828766	0.9121180773	14954	0.693258822	Cammeo	
5	13176	458.3429871	193.3373871	87.44839478	0.8918609023	13368	0.6406690478	Cammeo	
6	14688	507.1669922	211.7433777	89.31245422	0.9066908956	15262	0.646023921	Cammeo	
7	13479	477.0159912	200.0530548	86.65029144	0.9013283253	13786	0.6578972936	Cammeo	
8	15757	509.2810059	207.2966766	98.33613586	0.8803234696	16150	0.5897080898	Cammeo	
9	16405	526.5700073	221.6125183	95.43670654	0.9025205097	16837	0.6588828204	Cammeo	
10	14534	483.6409012	196.6508179	95.05068207	0.8754285574	14932	0.6496513486	Cammeo	
11	13485	471.5700073	198.272644	87.72728729	0.8967893124	13734	0.5723198652	Cammeo	
12	14930	499.9249878	212.2458191	90.01747894	0.9056064487	15248	0.6243727207	Cammeo	
13	14626	496.5859985	204.5341339	92.97468677	0.8907115459	15070	0.5702144504	Cammeo	
14	15926	522.7399902	225.7360535	91.05709076	0.9150331616	16240	0.7797688842	Cammeo	
15	14076	479.677002	199.489151	90.70998383	0.8906388879	14434	0.7812187672	Cammeo	
16	13500	476.9150085	202.5466766	85.4054718	0.9067548513	13800	0.7177033424	Cammeo	
17	14349	496.9460144	213.5440216	86.16077423	0.9149883986	14678	0.6668370962	Cammeo	
18	15209	496.5650024	214.0500793	91.02632141	0.9050725698	15395	0.5693695545	Cammeo	
19	15238	496.8710022	208.5317841	93.82839966	0.8930549026	15487	0.7323144674	Cammeo	
20	13509	480.4660034	207.1371613	83.94016266	0.9142105579	13732	0.5956349373	Cammeo	
21	13699	480.7290039	197.709167	89.94907379	0.8908207417	14178	0.672046721	Cammeo	
22	13921	481.9930115	203.44487	88.12581635	0.9013131857	14210	0.5613306165	Cammeo	
23	14405	488.2369995	204.43013	90.62645721	0.8963670731	14742	0.7637857795	Cammeo	
24	13377	472.9609985	199.9407086	86.68286896	0.9011330009	13685	0.7948306799	Cammeo	
25	15043	488.2900112	208.0117035	94.76868510	0.8870156700	15114	0.5868826045	Cammeo	

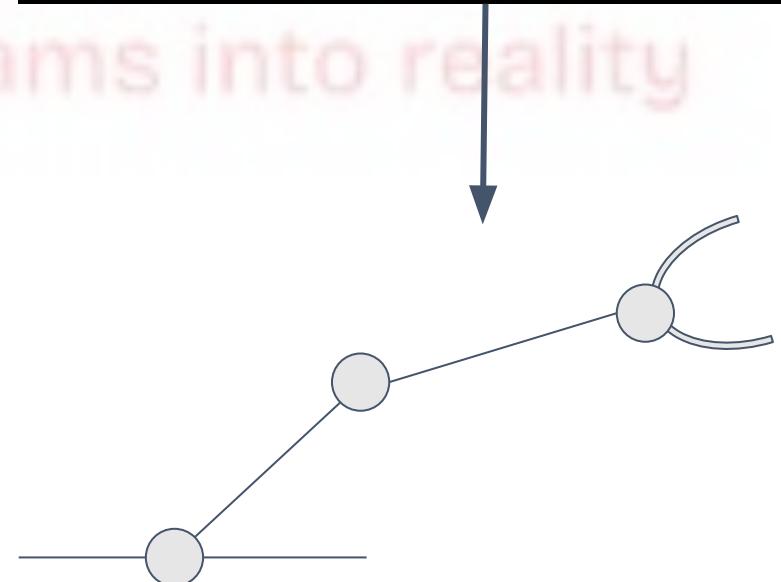
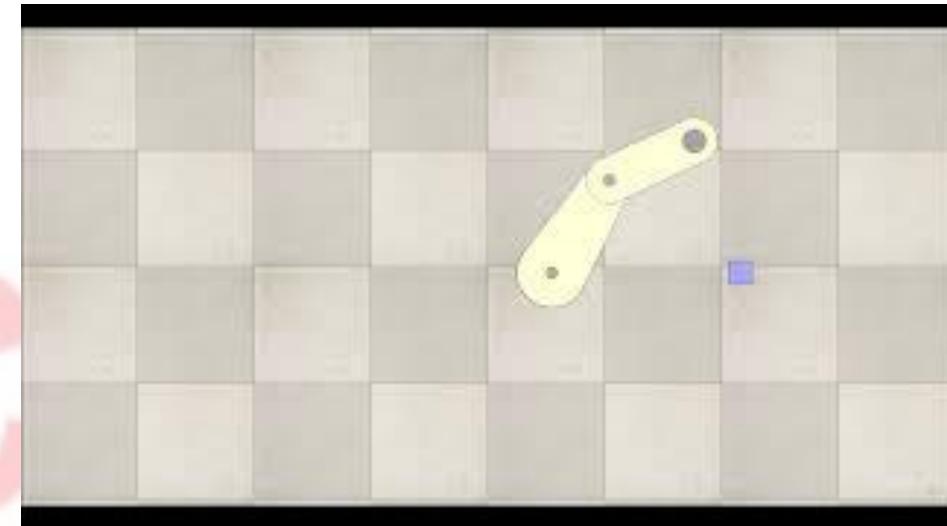


## Understanding Data

# Data Annotation

### Case Study

- AI to get the skeleton of a robot
  - Research the target AI model
- Annotation
  - [roboflow](#)
  - [Coco Annotator](#)



## Understanding Data

Normalization is to manipulate features so that they span a similar range, perhaps 0 to 1 or -1 to +1.

- help model train more effectively.
- Helps models converge more quickly during training.
- Helps the model learn appropriate weights for each feature.

NB: If you normalize a feature during training, you must also normalize that feature when making predictions.

```
/usr/local/lib/python3.12/dist-packages/sklearn/linear_model/_logistic.py:465: ConvergenceWarning: lbfgs failed to converge (status=1):  
STOP: TOTAL NO. OF ITERATIONS REACHED LIMIT.  
  
Increase the number of iterations (max_iter) or scale the data as shown in:  
    https://scikit-learn.org/stable/modules/preprocessing.html  
Please also refer to the documentation for alternative solver options:  
    https://scikit-learn.org/stable/modules/linear\_model.html#logistic-regression  
n_iter_i = _check_optimize_result(  
[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 0.0s finished  
+ LogisticRegression ① ?  
LogisticRegression(verbose=1)
```

## Understanding Data

Linear scaling normalization means converting floating-point values from their natural range into a standard range—usually 0 to 1 or -1 to +1.

- The lower and upper bounds of your data don't change much over time.
- The feature contains few or no outliers, and those outliers aren't extreme.
- The feature is approximately uniformly distributed across its range. That is, a histogram would show roughly even bars for most values.

Use the following formula to scale to the standard range 0 to 1, inclusive:

$$x' = (x - x_{min}) / (x_{max} - x_{min})$$

where:

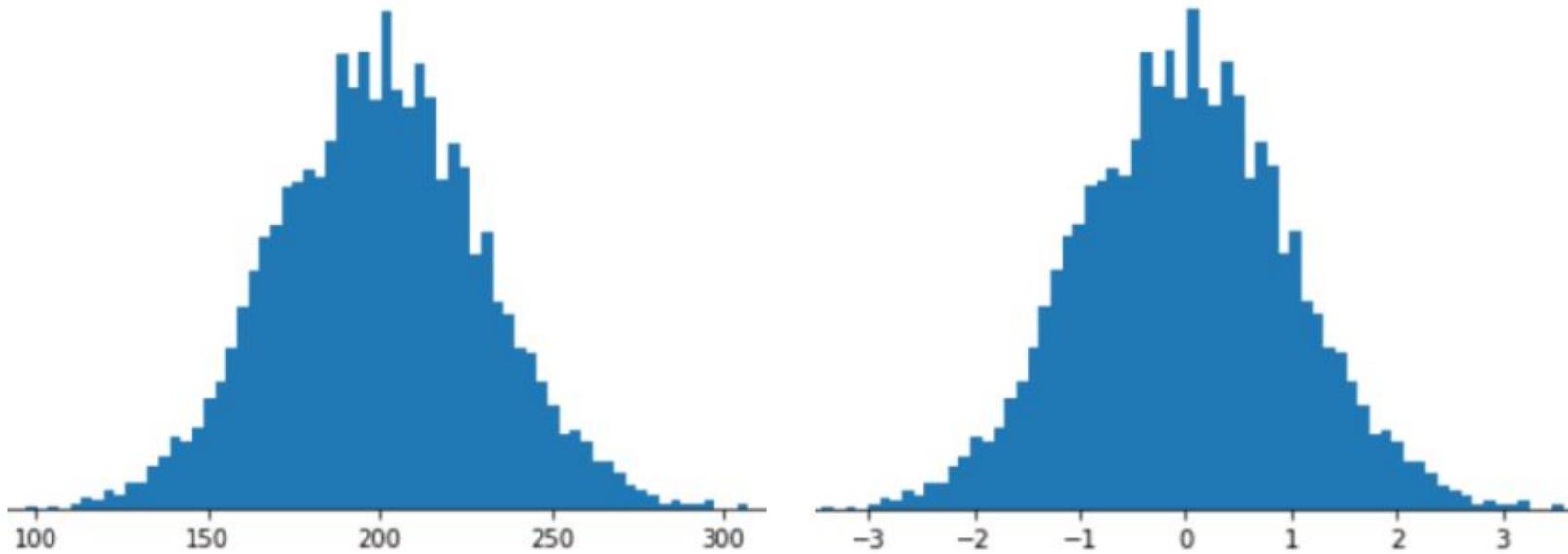
- $x'$  is the scaled value.
- $x$  is the original value.
- $x_{min}$  is the lowest value in the dataset of this feature.
- $x_{max}$  is the highest value in the dataset of this feature.



**Uniform distribution**  
(equal spread,  
no peaks)

## Understanding Data

Z-score normalization or Standardization is a good choice when the data follows a normal distribution or a distribution somewhat like a normal distribution.



Use the following formula to normalize a value,  $x$ , to its Z-score:

$$x' = (x - \mu)/\sigma$$

where:

- $x'$  is the Z-score.
- $x$  is the raw value; that is,  $x$  is the value you are normalizing.
- $\mu$  is the mean.
- $\sigma$  is the standard deviation.

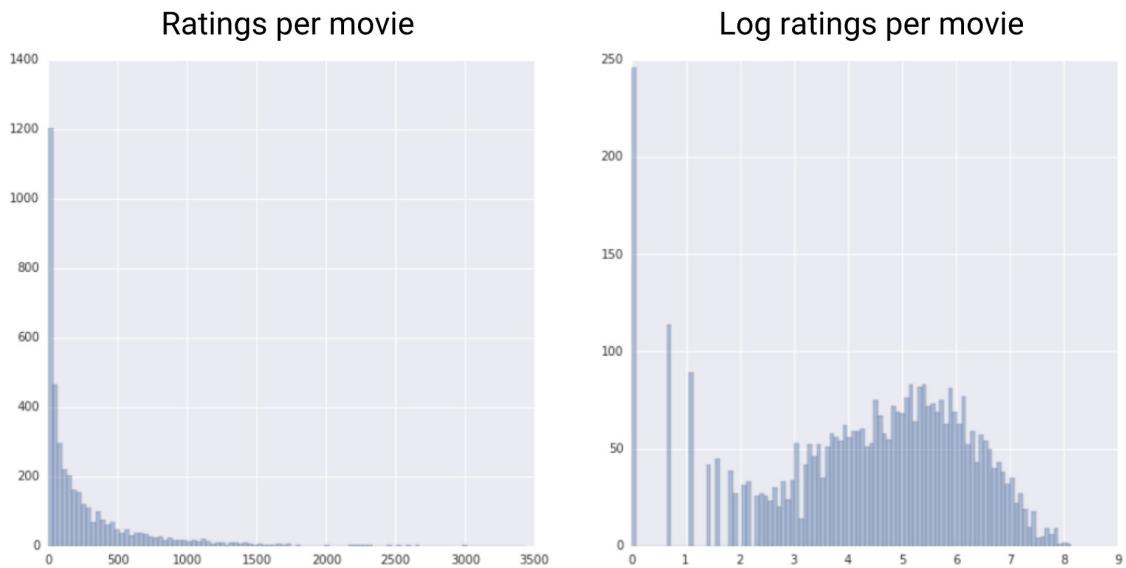
## Understanding Data

# Log scaling normalization

Log scaling computes the logarithm of the raw value.

This is helpful when the data conforms to a power law distribution.

As the values of X increase, the values of Y quickly decrease



Use the following formula to normalize a value,  $x$ , to its log:

$$x' = \ln(x)$$

where:

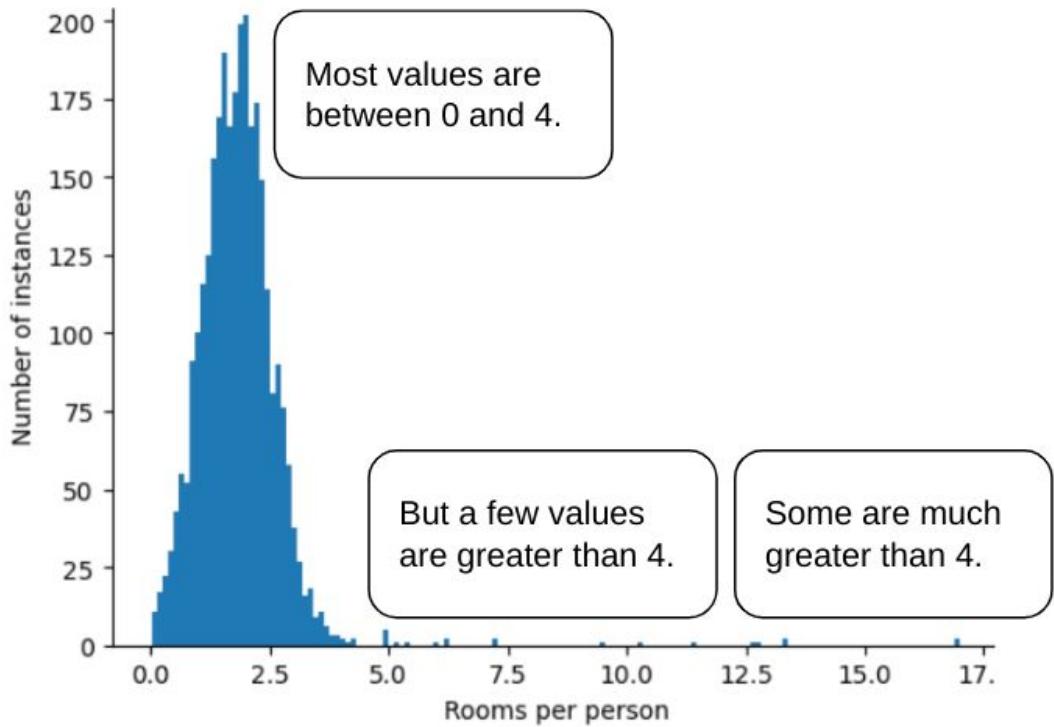
- $x'$  is the natural logarithm of  $x$ .
- original value = 54.598

Therefore, the log of the original value is about 4.0:

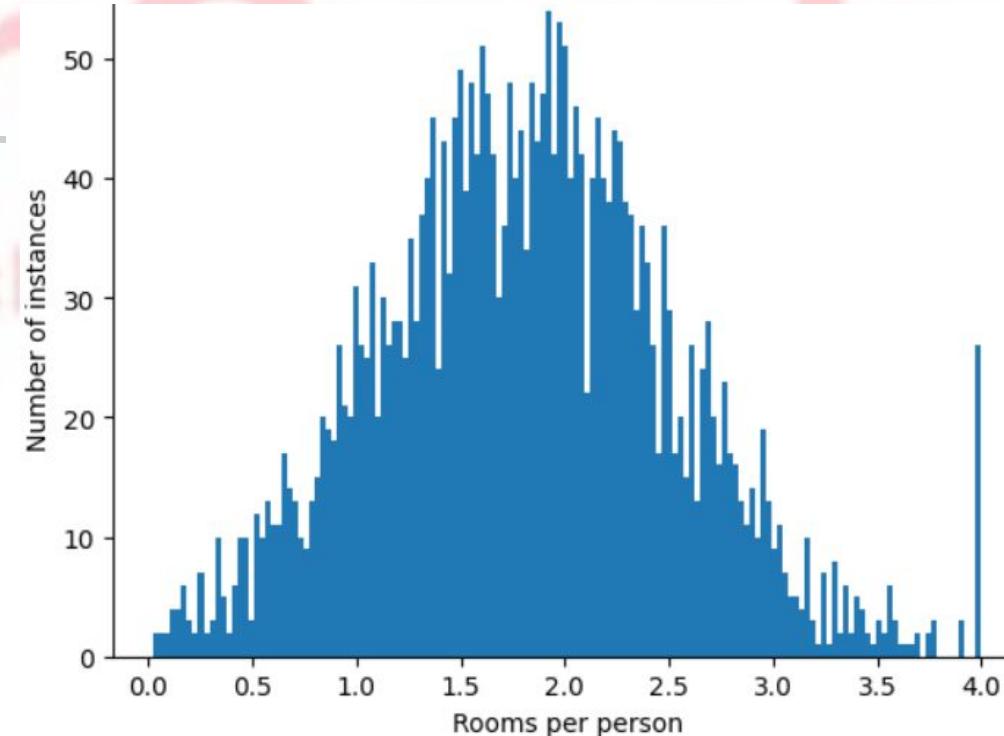
$$4.0 = \ln(54.598)$$

## Understanding Data

Clipping is a technique to minimize the influence of extreme outliers.



Before Clipping



After Clipping



*Disclaimer: This training material belongs to techcrush and shouldn't be shared*