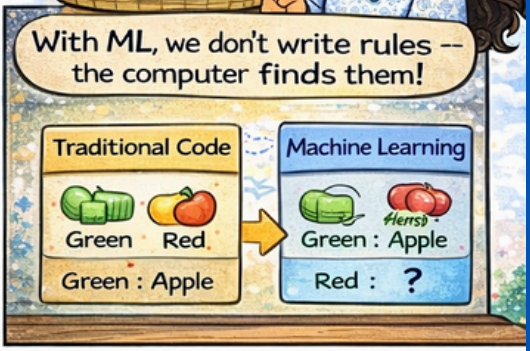
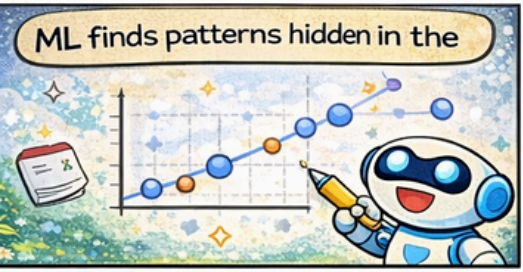
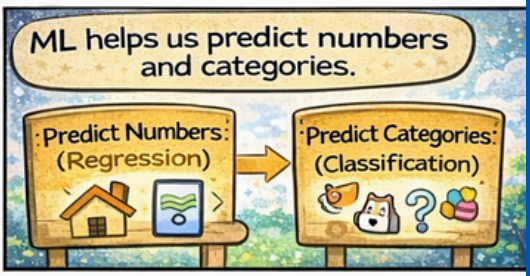

Understanding
MACHINE LEARNING
A Comic Beginner Series

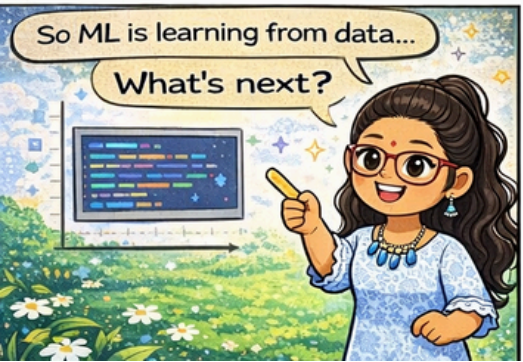
What is MACHINE LEARNING?

Episode 1



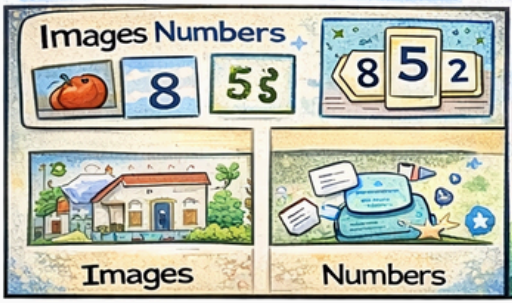


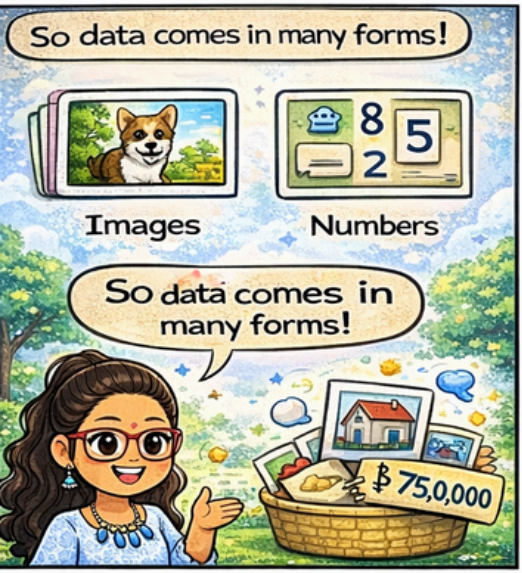


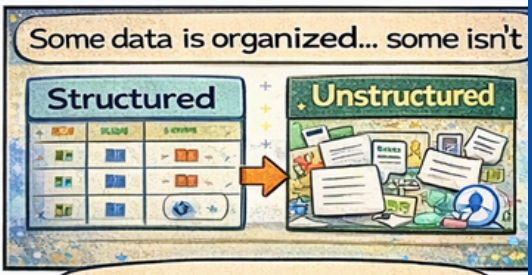


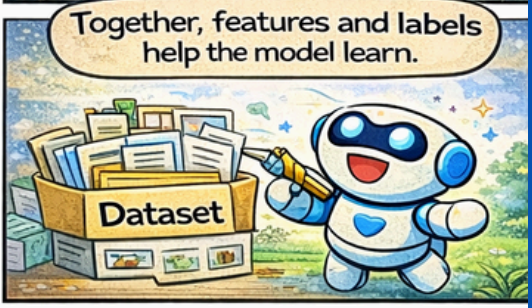
Data is the Fuel

Episode 2









Supervised vs Unsupervised Learning

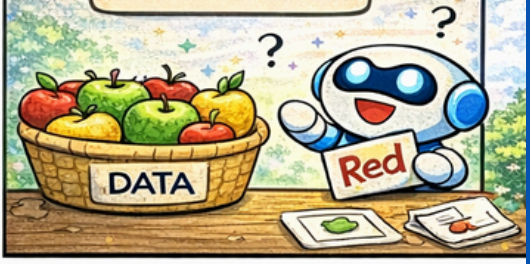
Episode 3

Supervised Learning Examples

In supervised learning, we give the machine examples with labels.

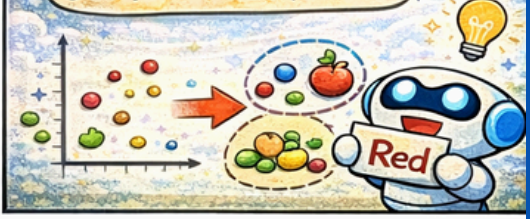
Unsupervised Learning

In unsupervised learning, the machine gets only the data... no labels.

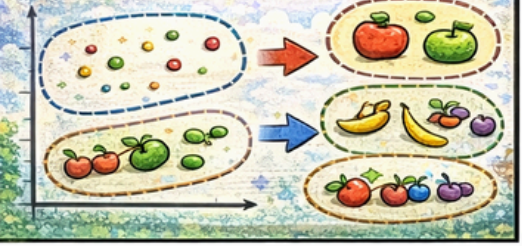


It groups similar things together automaticaily!

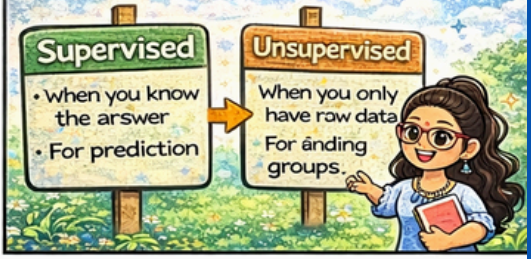
So it's like finding groups in messy data!



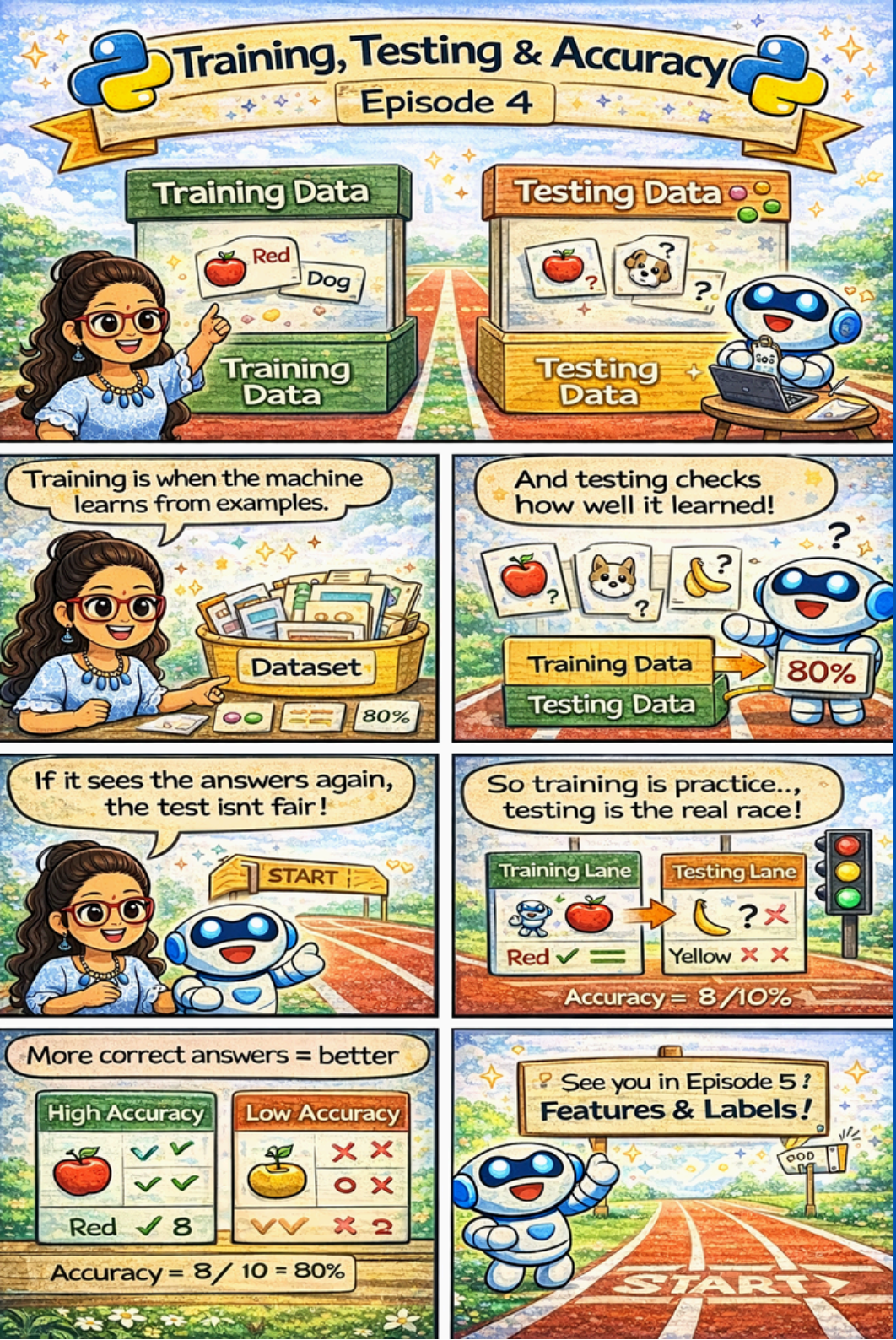
So it's like finding groups in messy data!



Oh! Labels decide everything!

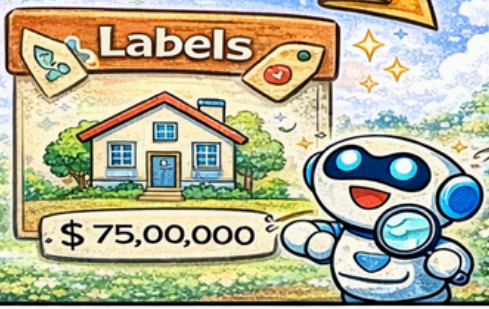


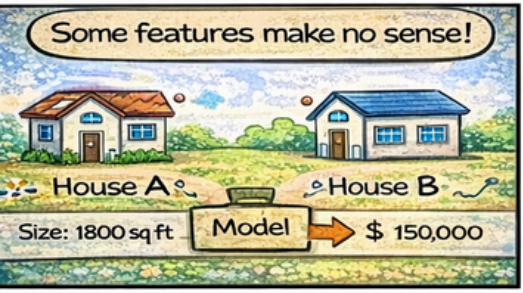
See you in Episode 4: Training, Testing & Accuracy!

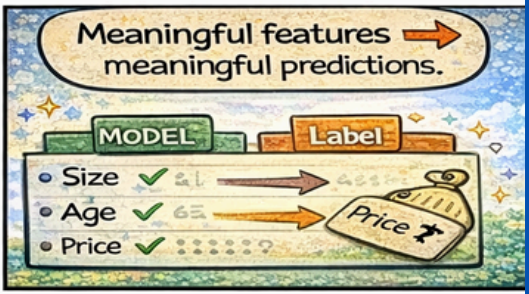

Training, Testing & Accuracy
Episode 4
Training Data
Red
Dog
Training Data
Testing Data
Testing Data
Training is when the machine learns from examples.
Dataset
80%
And testing checks how well it learned!
Training Data
Testing Data
80%
If it sees the answers again, the test isnt fair!
START
So training is practice.., testing is the real race!
Training Lane
Testing Lane
Red
Yellow
Accuracy = 8/10%
More correct answers = better
High Accuracy
Low Accuracy
Red 8
2
Accuracy = 8/ 10 = 80%
See you in Episode 5 !
Features & Labels!
START

Features & Labels

Episode 5



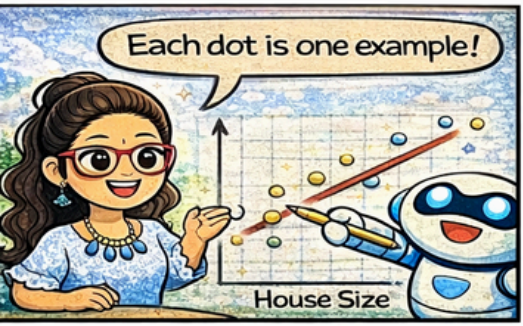


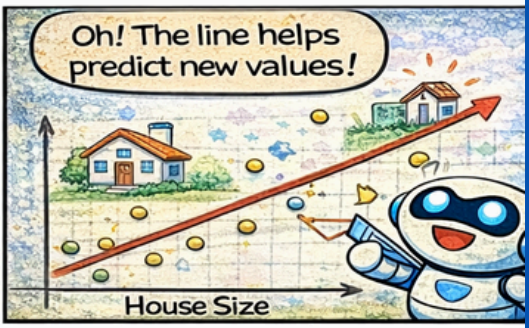


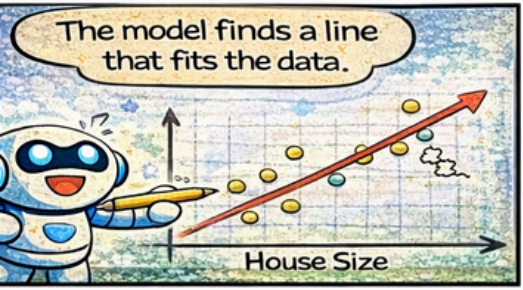
See you in Episode 6:
Regression – Predicting Numbers!

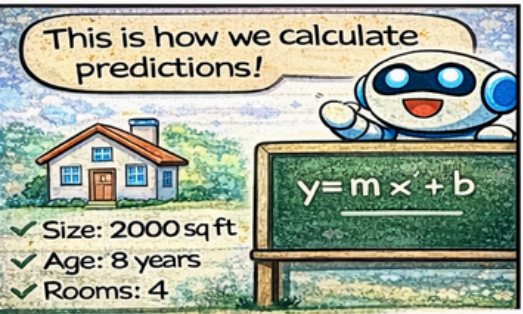
Regression – Predicting Numbers

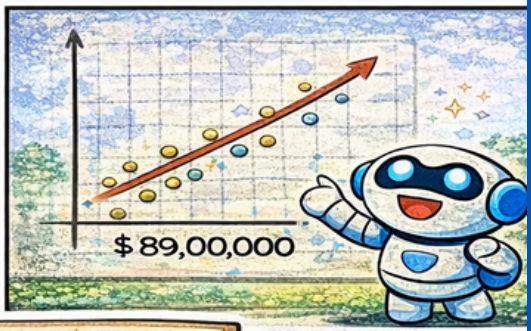
Episode 6







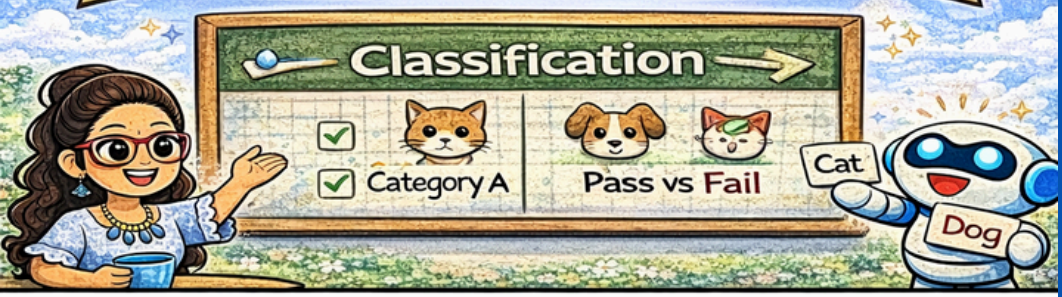


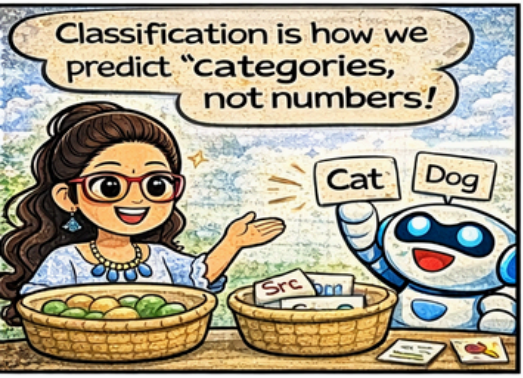


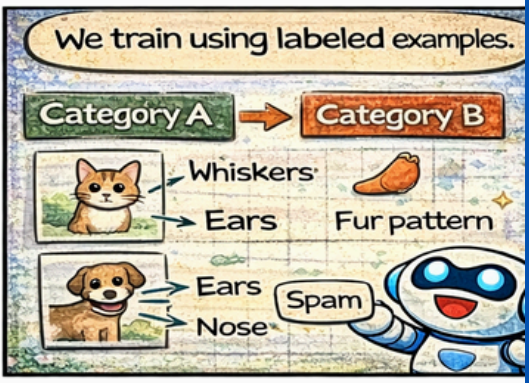
See you in Episode 7:
Classification – Predicting Categories!

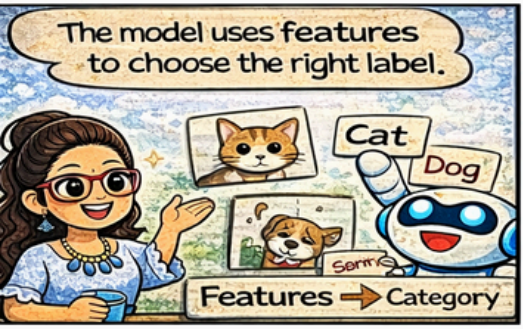
Classification– Predicting Categories

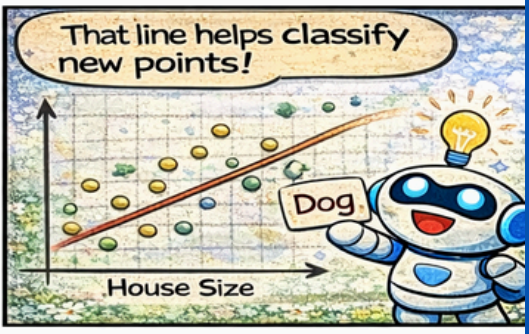
Episode 7





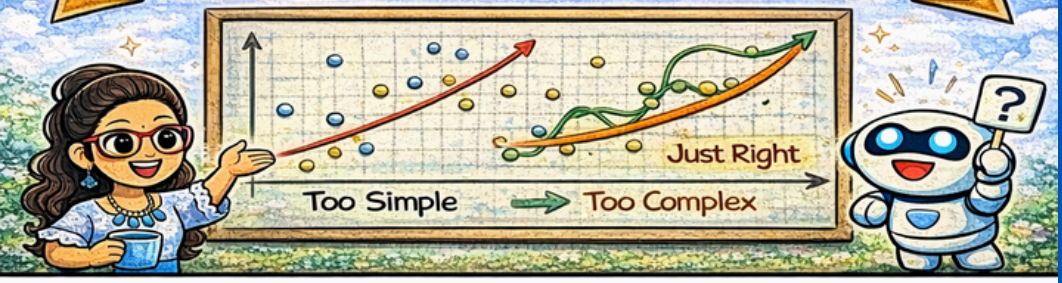


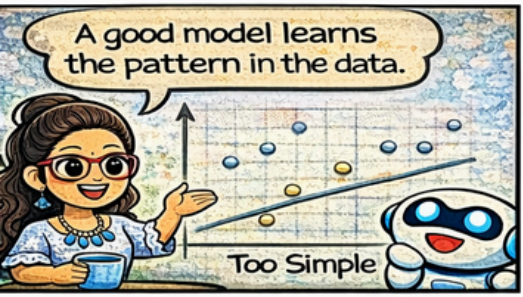


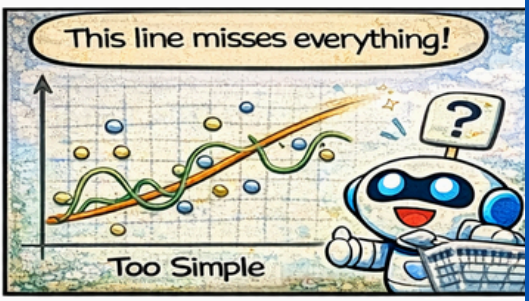


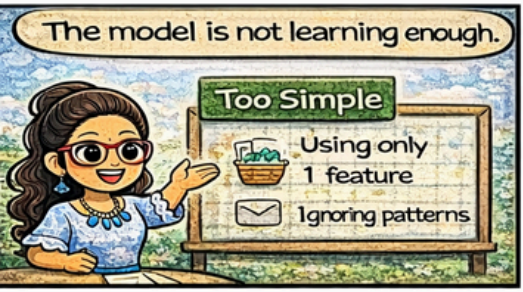
Underfitting & Overfitting

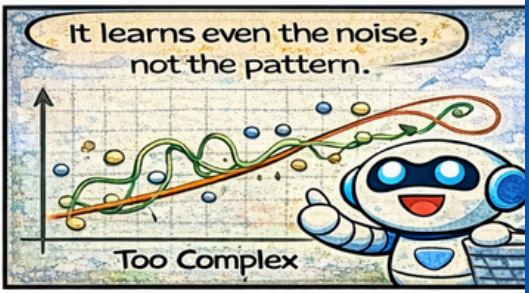
Episode 8

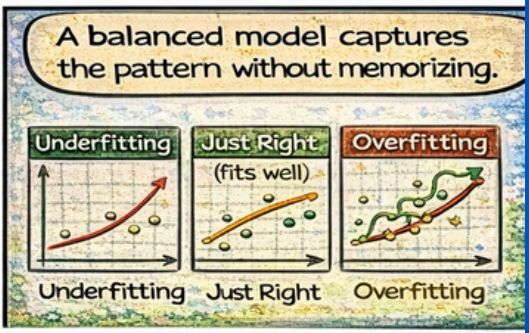






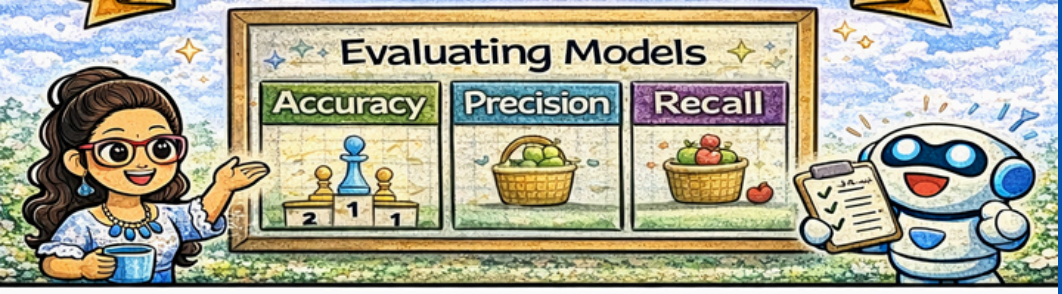


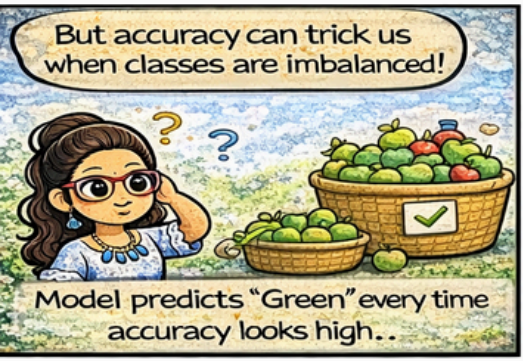


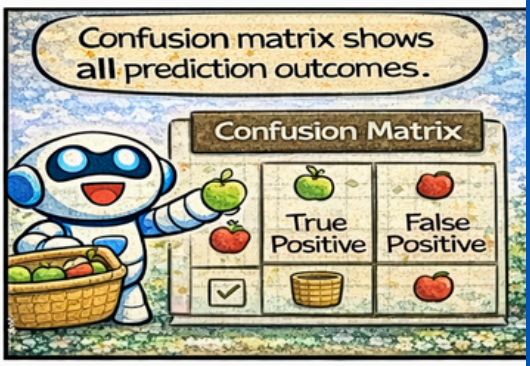


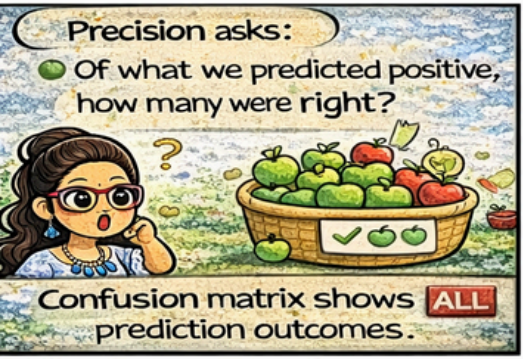
Evaluating Models

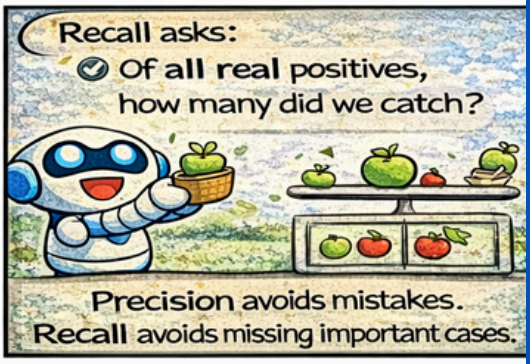
Episode 9

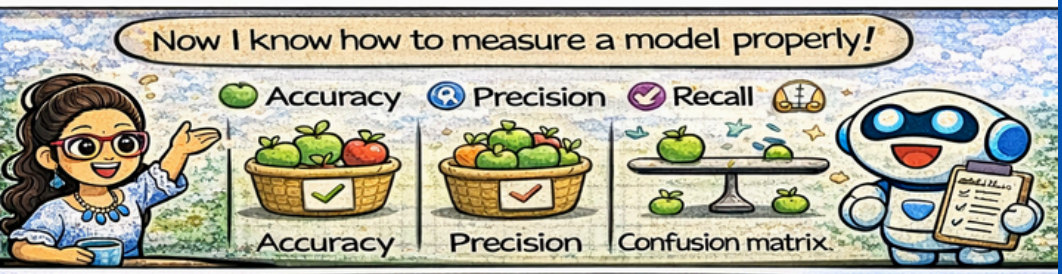


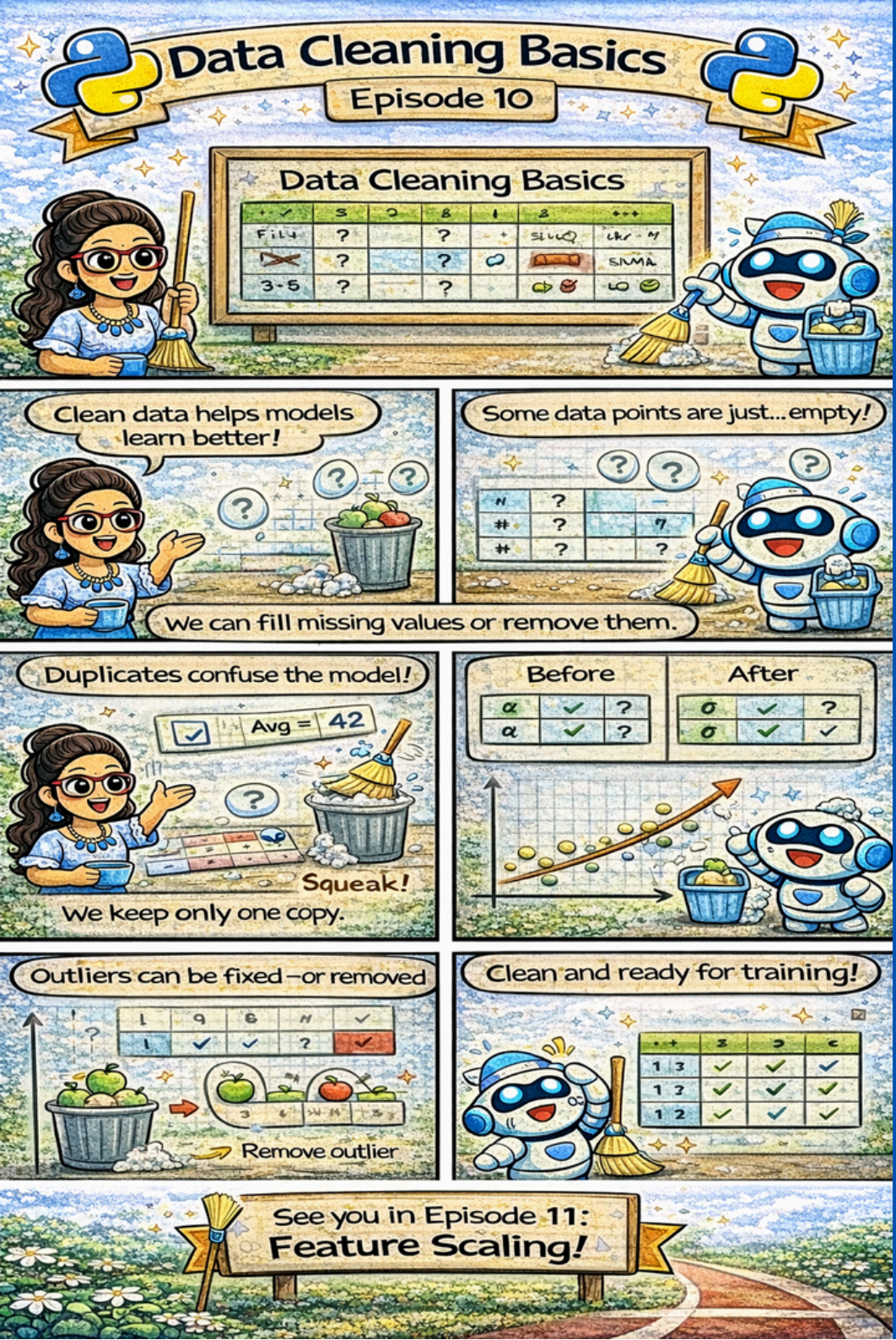


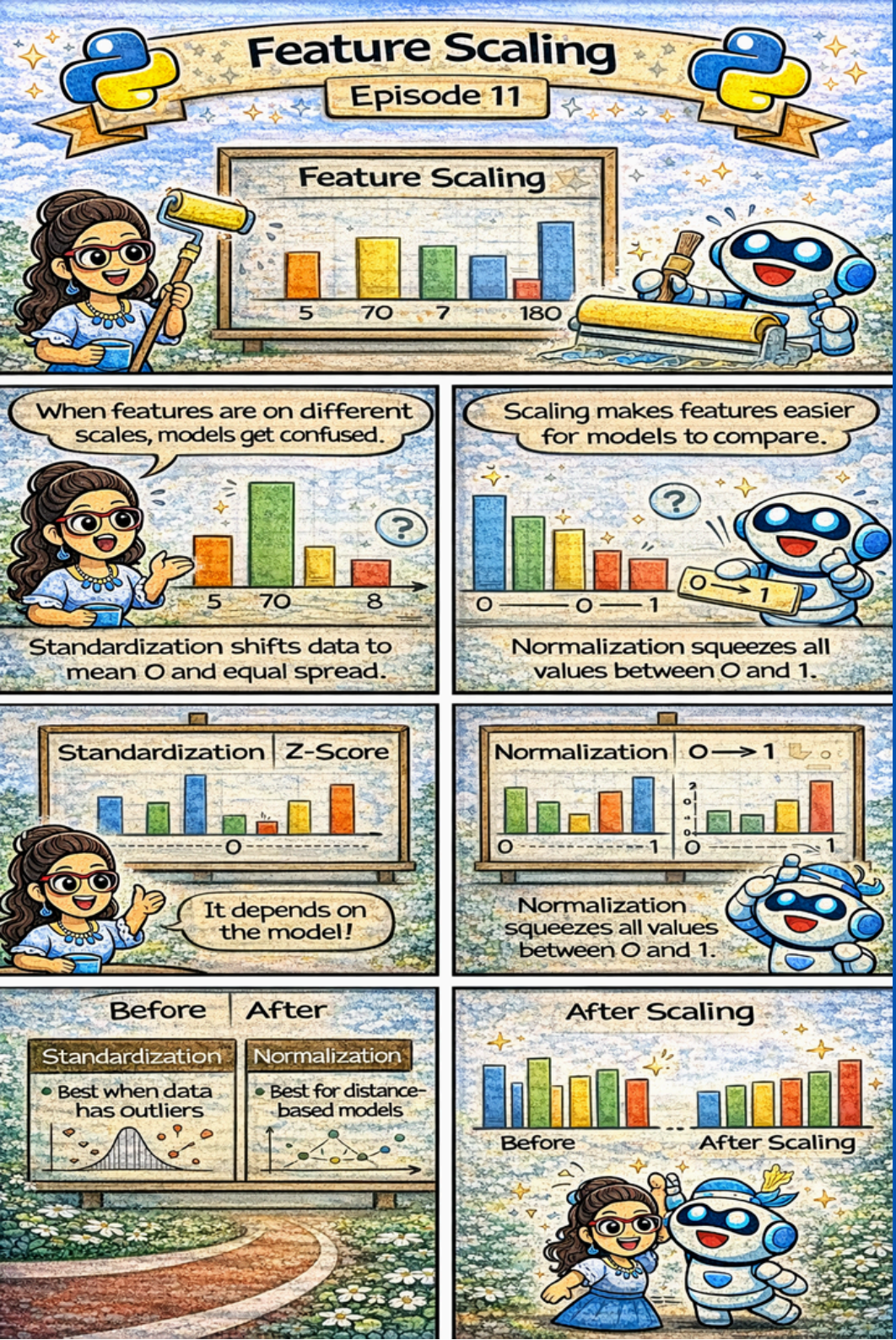






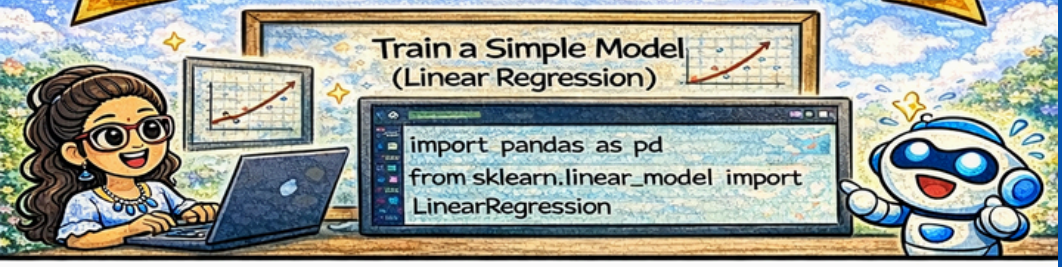


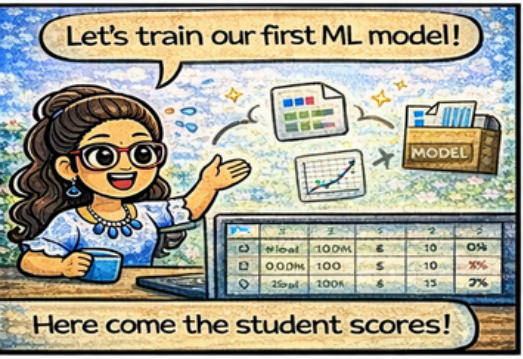

Data Cleaning Basics
Episode 10
Data Cleaning Basics
Clean data helps models learn better!
Some data points are just...empty!
We can fill missing values or remove them.
Duplicates confuse the model!
Avg = 42
Squeak!
We keep only one copy.
Before
After
Outliers can be fixed—or removed
Remove outlier
Clean and ready for training!
See you in Episode 11:
Feature Scaling!

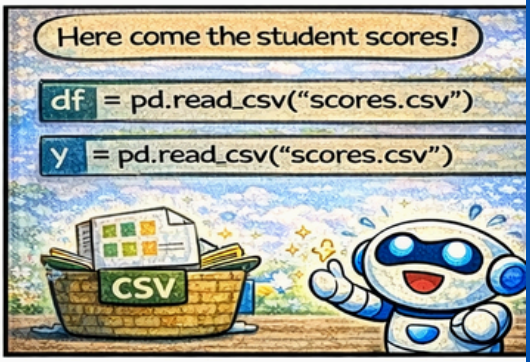

Feature Scaling
Episode 11
Feature Scaling
5
70
7
180
When features are on different scales, models get confused.
5
70
8
?
Standardization shifts data to mean 0 and equal spread.
Scaling makes features easier for models to compare.
?
0
0
1
0 → 1
Normalization squeezes all values between 0 and 1.
Standardization
Z-Score
0
It depends on the model!
Normalization
0 → 1
0
1
0
1
Normalization squeezes all values between 0 and 1.
Before
After
Standardization
Best when data has outliers
Normalization
Best for distance-based models
After Scaling
Before
After Scaling

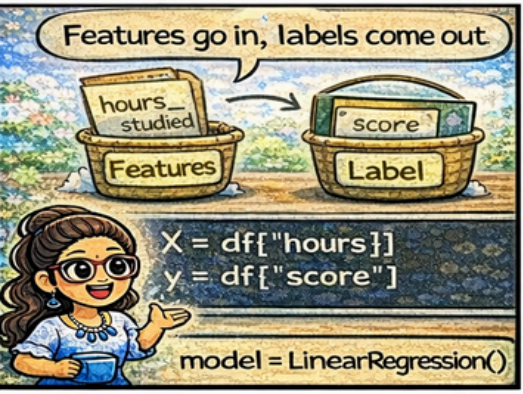
Train a Simple Model

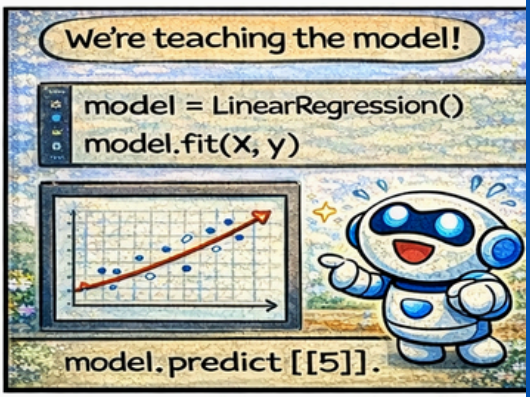
Episode 12

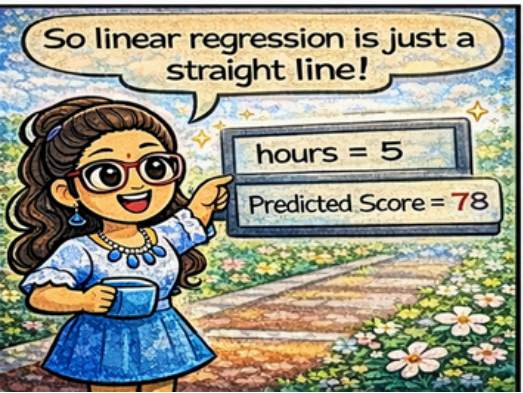


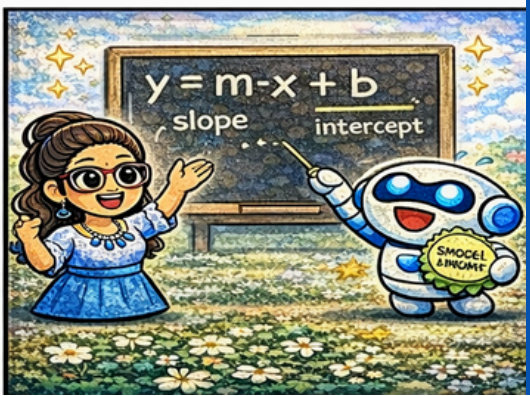








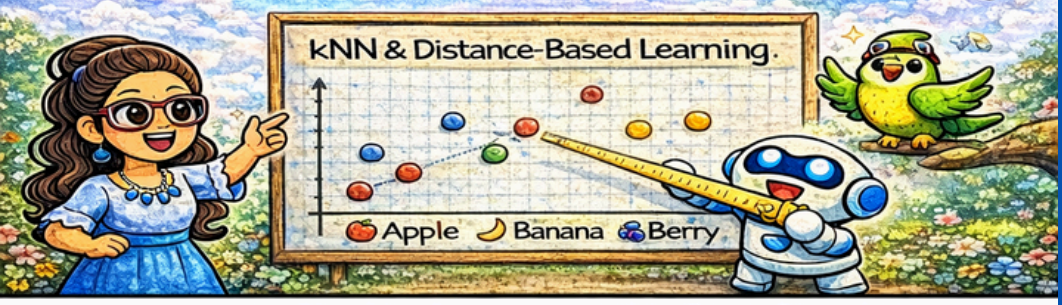


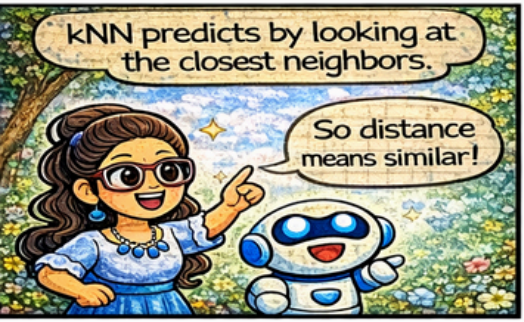


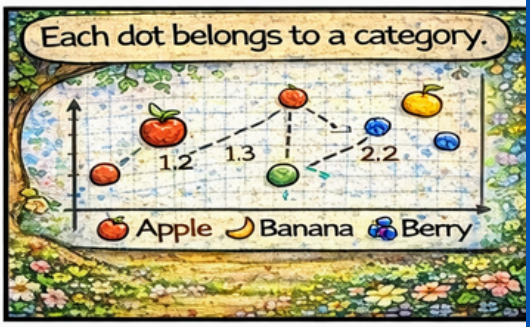

Decision Trees Intuition
Episode 13
A decision tree decides by asking simple questions.
?
?
?
?
Yes
No
Oh! It's like a quiz with steps!
Is it furry?
No → Dog?
Each question splits the data.
Color?
Oes
Size?
Shape?
Red Fruits
Green
Grapes
We keep splitting until choices get clear!
?
?
We make decisions like this every day.
Is it raining?
Go out
So leaves are the final predictions!
Is it raining?
Go out
Stay home
Go out
?
Apple
Banana
Beach
Shapping

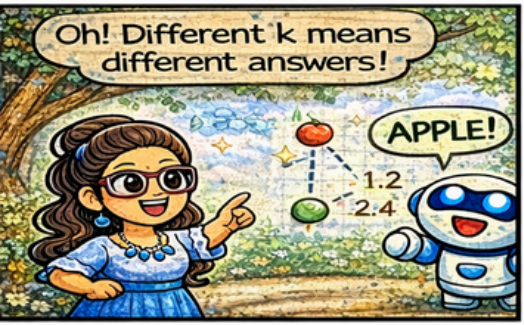
kNN & Distance-Based Learning

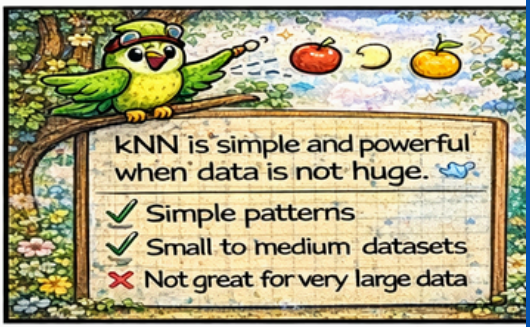
Episode 14












Mini Project: Predict Student Scores!
Episode 15
Student Score Predictor
Student | Hours | Score
1 | 40 | 40
3 | 77 | ?
2 | ? | ?
scores.csv
First, load the raw data.
Student | Hours | Score
1 | 40 | 40
3 | 77 | ?
? | ? | ?
Let's clean the messy parts!
Duplicates
Outairs
Train with one part, test with another:
Feature
Hours Studied
Missing
Duplicates
We'll use hours as the input, Score as the output.
moael fit(X train, y..train
Train with one part.
Dataset
1 | 3 | 2
2 | 6 | 2
Let's predict new student scores!
Input ➡ hours = 5
Model ➡ = 78
Predictions ➡ 78
SEVENTY, EIGHT!
Mini Project Complete –
See you in Part 2!


A Sriju Comic
Curiosity Lives Here!
Stories that smile back at you