Supervised learning

- Supervised learning is like teaching a computer how to do something by showing it examples.
- In supervised learning, the model learns from labeled training data.

Example: Email Spam Detection

Step1: Data Collection

Step2: Training

Step3: Testing

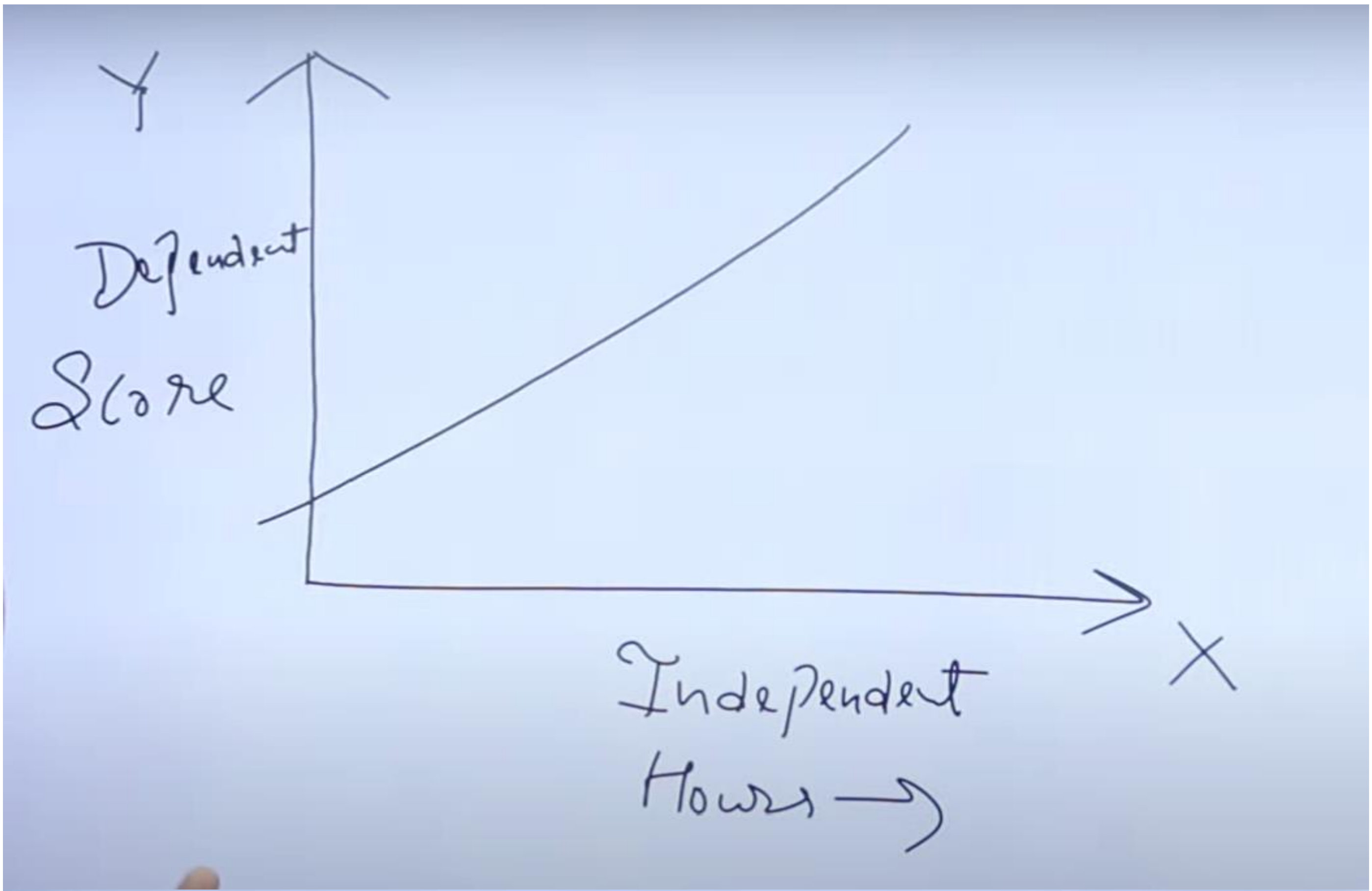
Algorithms

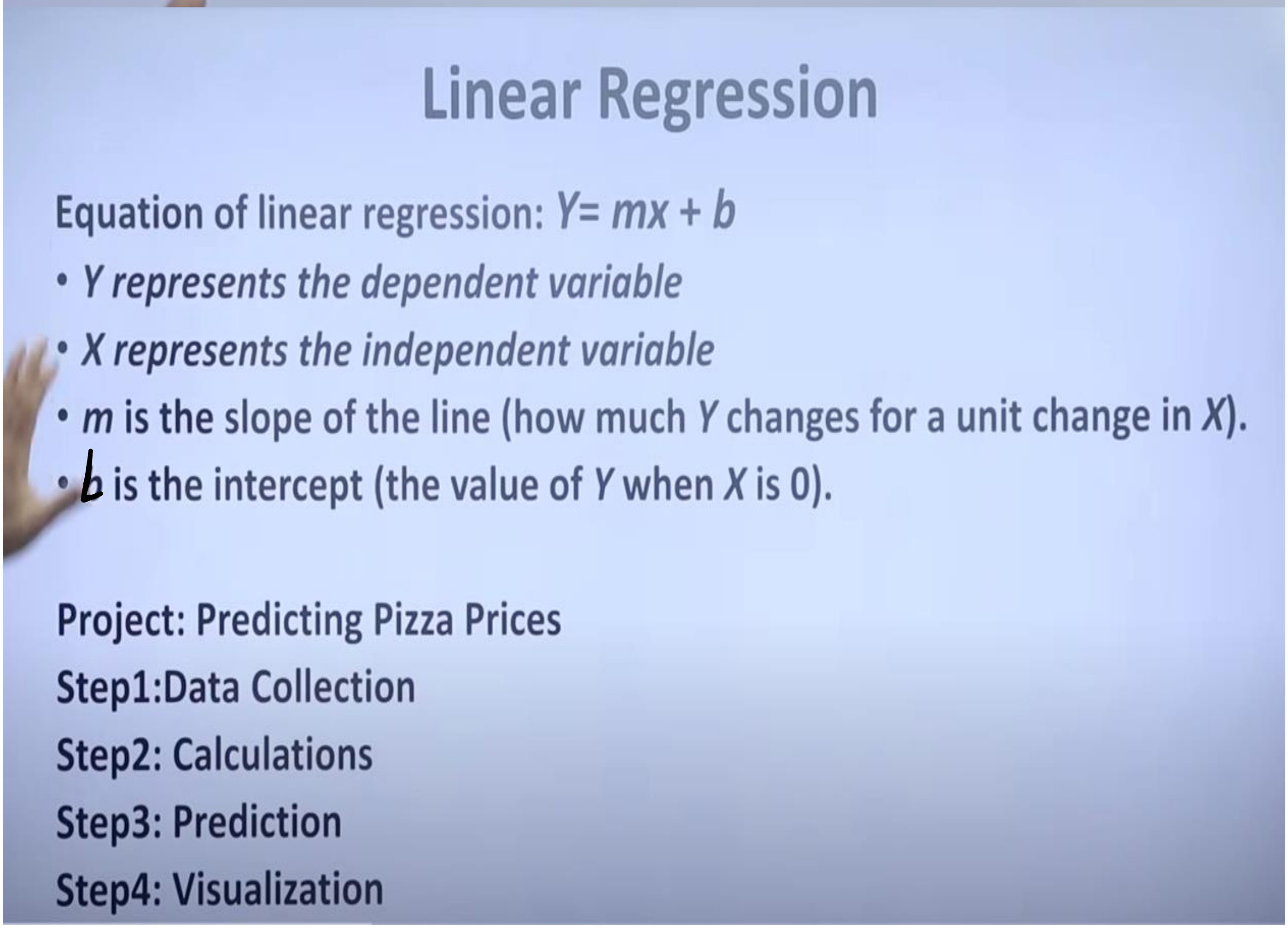
- **Classification** goal is to assign input data to predefined categories or classes. Algorithms like K-Nearest Neighbors (KNN), Logistic Regression, Decision Trees, Support Vector Machines, and Neural Networks are commonly used for classification tasks.
- **Regression** goal is to predict a numerical value based on input features. Algorithms like Linear Regression, Polynomial Regression, and various types of regression trees are used for regression tasks.

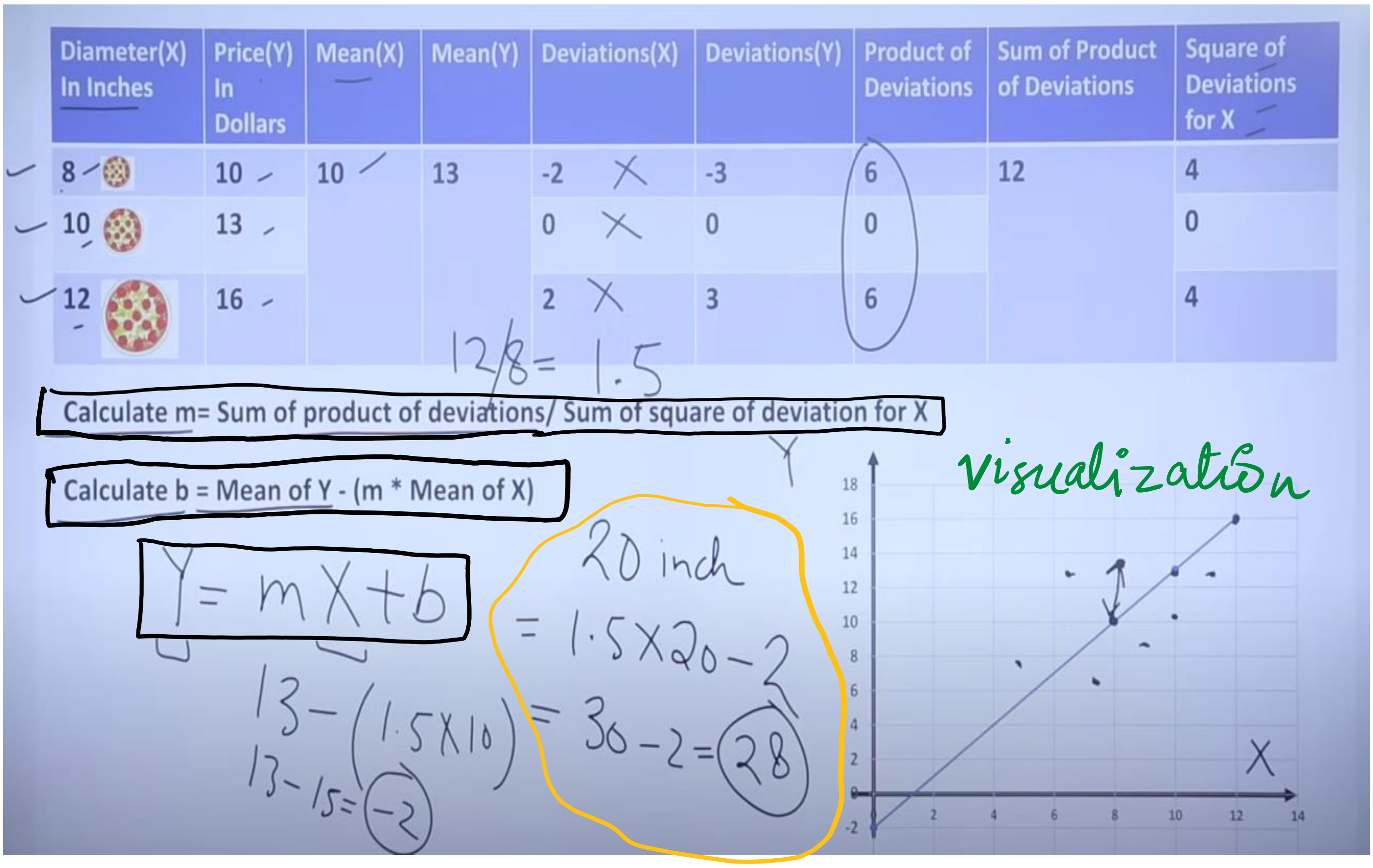
Regression

- Regression is a statistical method that helps us understand and predict the relationship between variables.
- Describes how one variable (dependent variable) changes as another variable (independent variable) changes.
- Dependent Variable: We are trying to predict or explain(Y).
- Independent Variable: that are used to predict or explain the changes in the dependent variable(X).

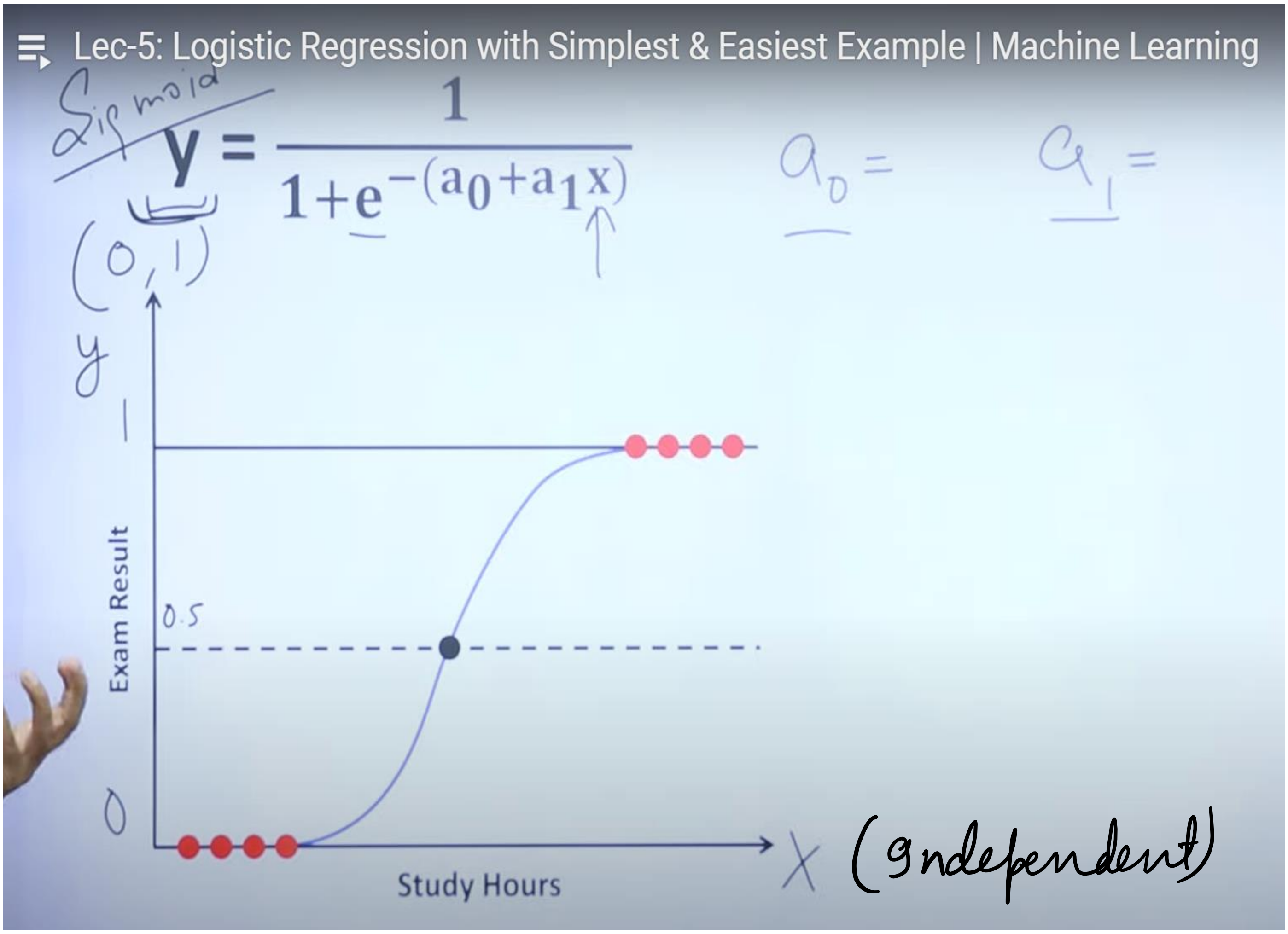
For example: predicting salary based on years of experience, predicting exam score based on study hours, predicting resale price based on vehicle age.

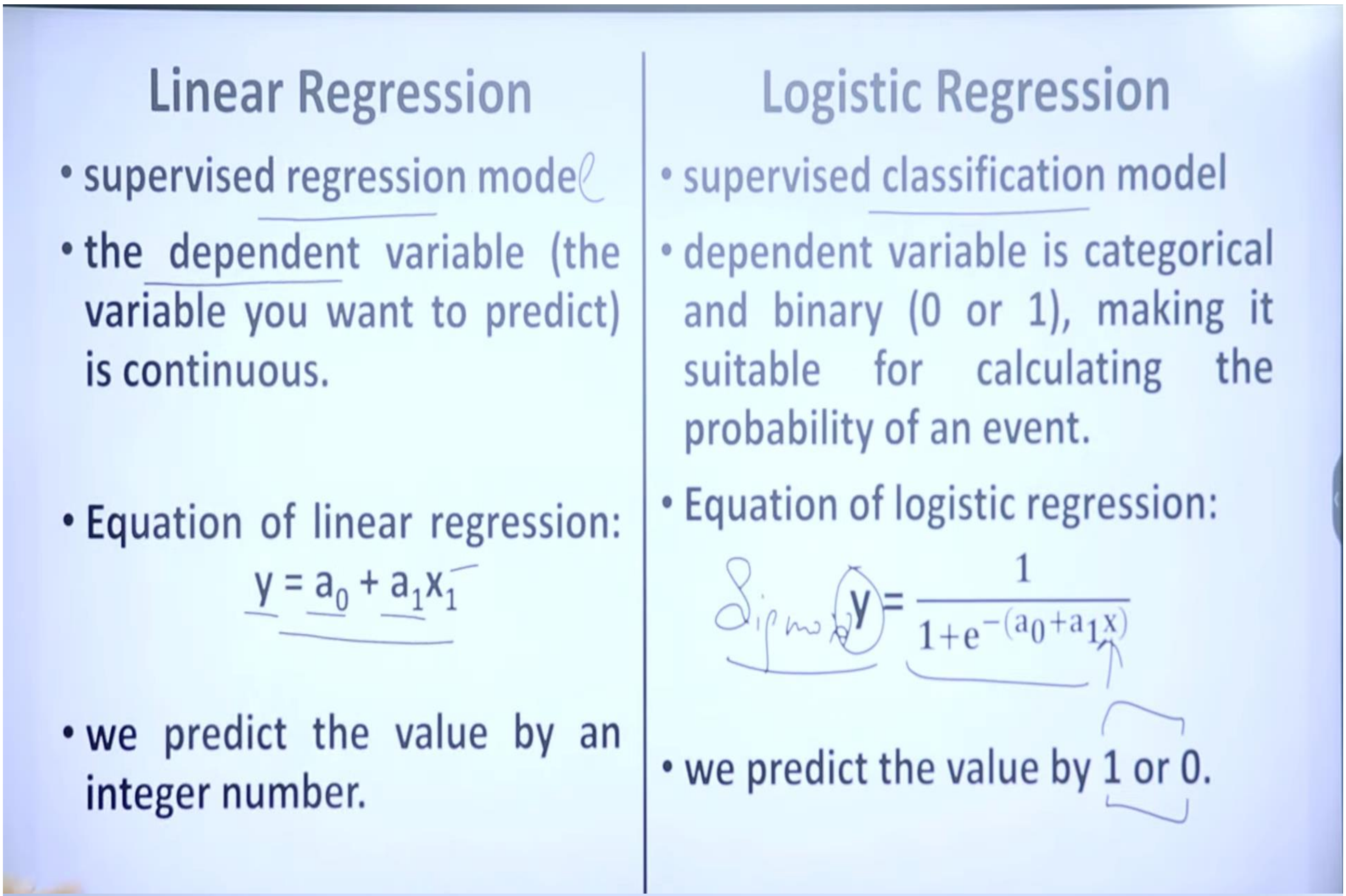

Y
Dependent
Score
Independent
Hours →
X

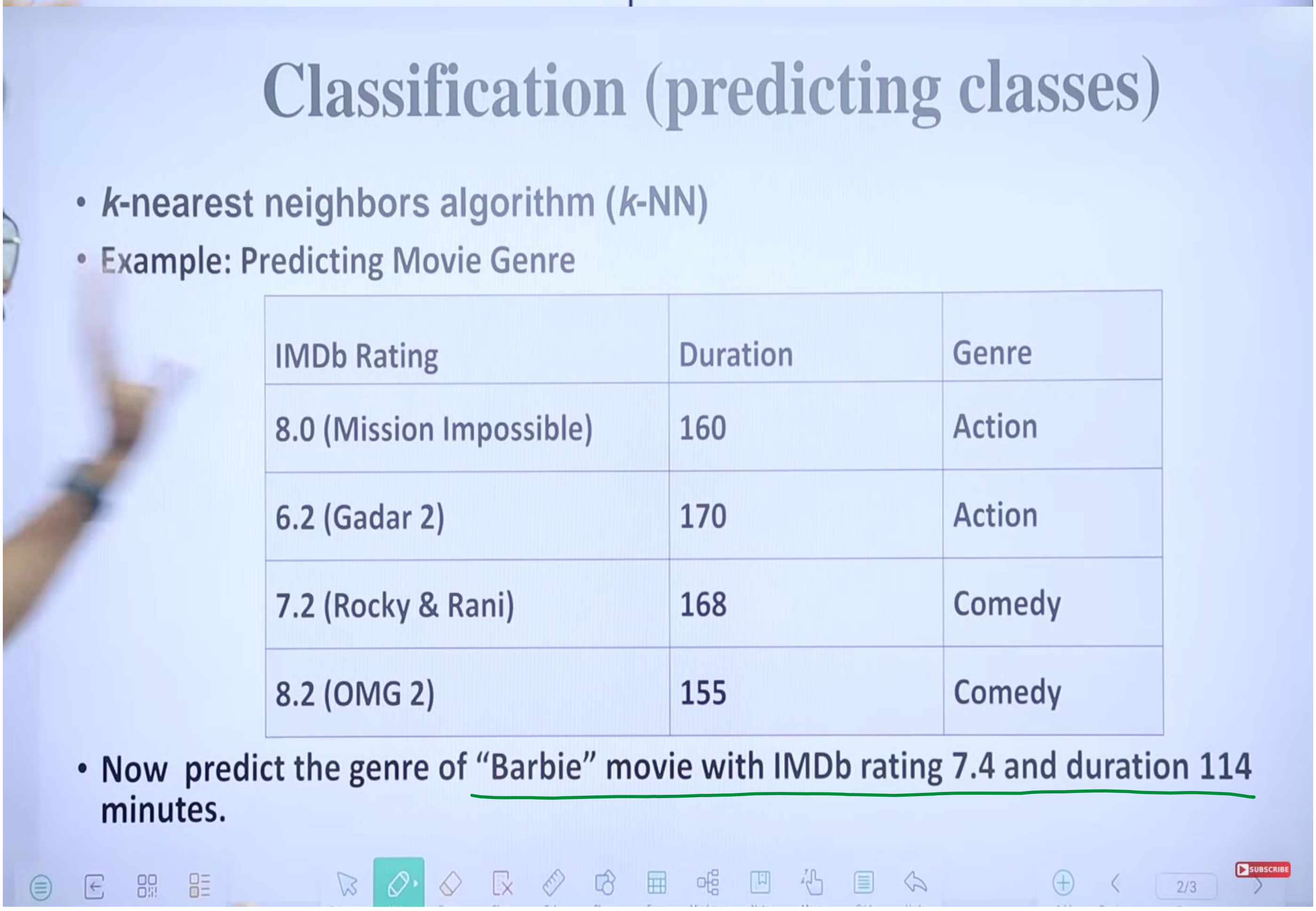

Linear Regression
Equation of linear regression: Y= mx + b
• Y represents the dependent variable
• X represents the independent variable
• m is the slope of the line (how much Y changes for a unit change in X).
• b is the intercept (the value of Y when X is 0).
Project: Predicting Pizza Prices
Step1:Data Collection
Step2: Calculations
Step3: Prediction
Step4: Visualization



Diameter(X) In Inches	Price(Y) In Dollars	Mean(X)	Mean(Y)	Deviations(X)	Deviations(Y)	Product of Deviations	Sum of Product of Deviations	Square of Deviations for X
8	10	10	13	-2	-3	6	12	4
10	13			0	0	0		0
12	16			2	3	6		4



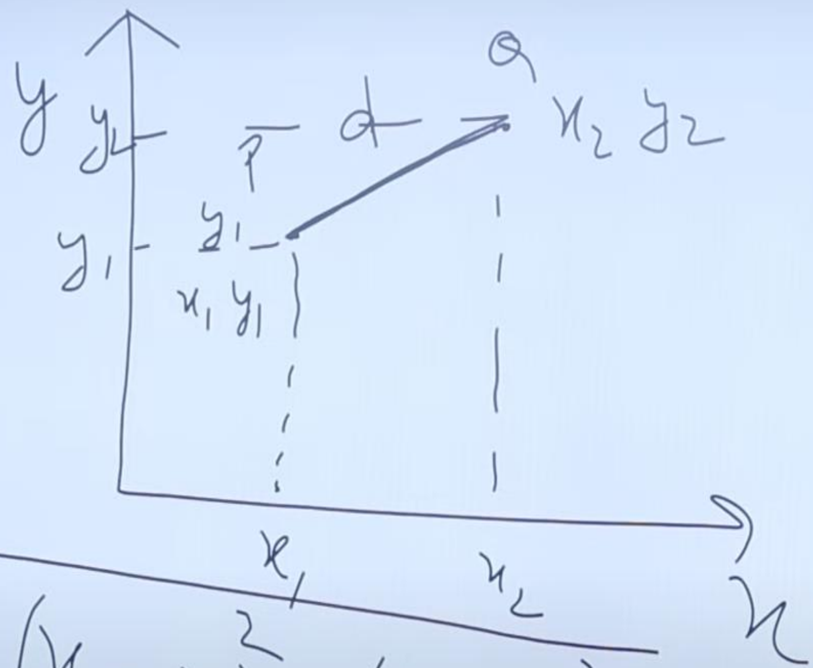

Linear Regression
• supervised regression model
• the dependent variable (the variable you want to predict) is continuous.
• Equation of linear regression:
$y = a_0 + a_1x_1$
• we predict the value by an integer number.
Logistic Regression
• supervised classification model
• dependent variable is categorical and binary (0 or 1), making it suitable for calculating the probability of an event.
• Equation of logistic regression:
Sigmoid $y = \frac{1}{1+e^{-(a_0+a_1x)}}$
• we predict the value by 1 or 0.


Classification (predicting classes)
• *k*-nearest neighbors algorithm (*k*-NN)
• Example: Predicting Movie Genre

IMDb Rating	Duration	Genre
8.0 (Mission Impossible)	160	Action
6.2 (Gadar 2)	170	Action
7.2 (Rocky & Rani)	168	Comedy
8.2 (OMG 2)	155	Comedy

• Now predict the genre of "Barbie" movie with IMDb rating 7.4 and duration 114 minutes.

Euclidean distance



$$d(p,q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Step 1: Calculate Distances x_1 7.4, 114 y_1

Calculate the Euclidean distance between the new movie and each movie in the dataset.

x_2, y_2

Distance to (8.0, 160) = √((7.4 – 8.0)^2 + (114 - 160)^2) = √(0.36 + 2116) ≈ 46.00

Distance to (6.2, 160) = √((7.4 – 6.2)^2 + (114 – 170)^2) = √(1.44 + 3136) ≈ 56.01

Distance to (7.2, 168) = √((7.4 – 7.2)^2 + (114 – 168)^2) = √(0.04 + 2916) ≈ 54.00

Distance to (8.2, 155) = √((7.4 – 8.2)^2 + (114 – 155)^2) = √(0.64 + 1681) ≈ 41.00

Lowest → K=1

$\boxed{K=3}$ ← take 3 value and choose majority one.

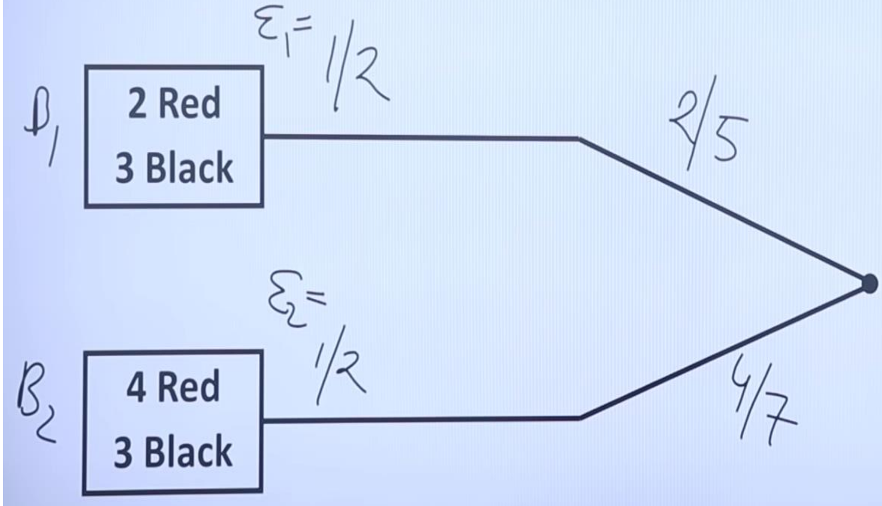
Step 2: Select K Nearest Neighbors

A, C, C

Step 3: Majority Voting (Classification)

Understand Bayes' Theorem With Example

- Bag 1 contain 2 Red and 3 Black balls. Bag 2 contain 3 Red and 4 Black Balls. One ball is drawn at random from one of these bags and its red. Find the probability that it is drawn from Bag 1.



2R 3B $\frac{2}{5}$

$\frac{1}{2} \times \frac{2}{5} + \frac{1}{2} \times \frac{4}{7}$ = Total probability

Baye's theorem

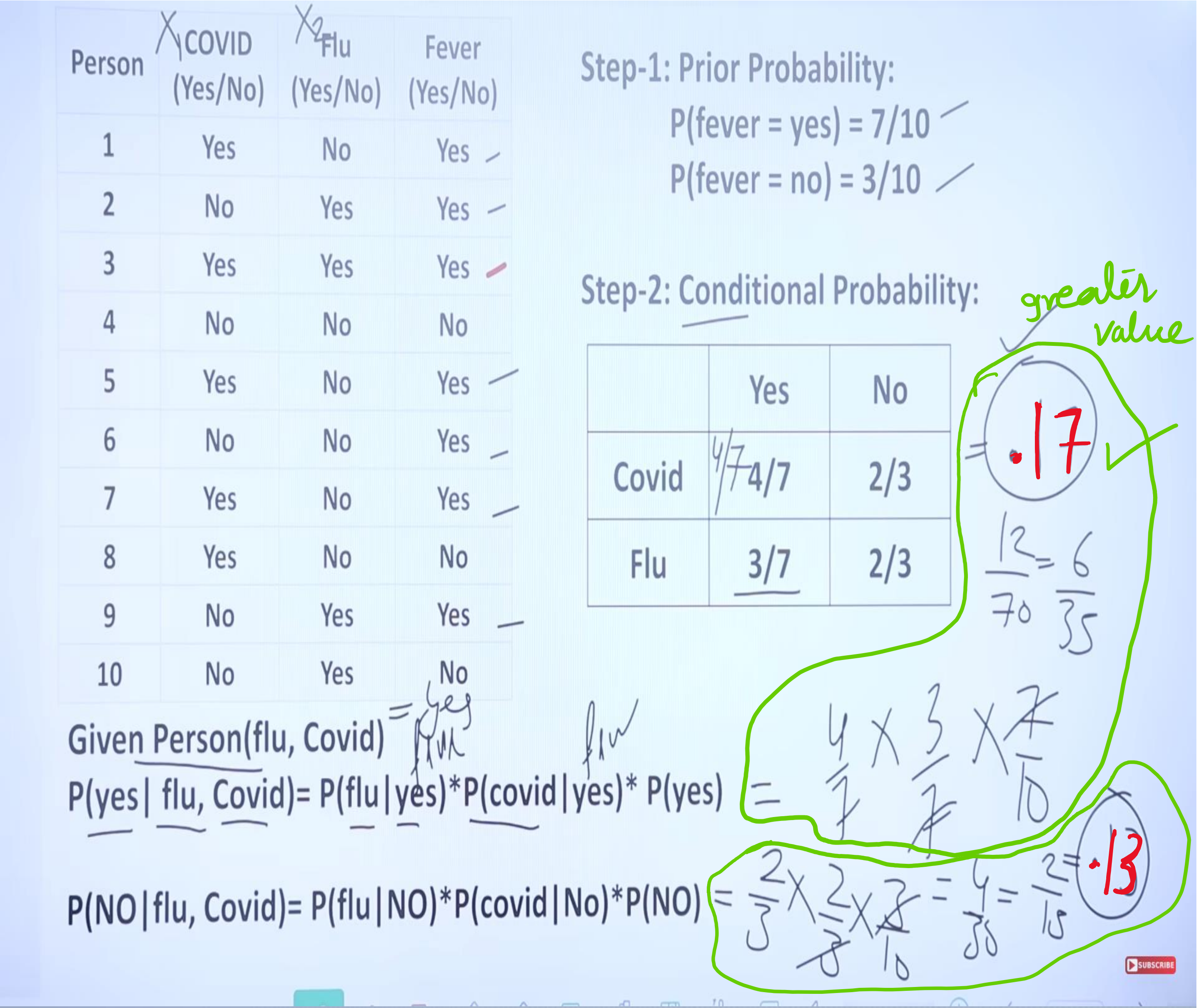
$$P(Y/X) = \frac{P(X|Y) * P(Y)}{P(X)} = \frac{\frac{2}{5} \times \frac{1}{2}}{}$$

$$P(Y|X_1, X_2, \ldots\ldots X_n) = \frac{P(X_1|Y)*P(X_2|Y)*P(X_3|Y)\ldots\ldots P(X_n|Y)*P(Y)}{P(X_1)*P(X_2)*P(X_3)\ldots\ldots P(X_n)}$$

$P(Y/X)$ { Finding probability of Y if value of X is given

$$P(N/X) = \frac{P(X|N) * P(N)}{P(X)}$$

$$P(N|X_1, X_2, \ldots\ldots X_n) = \frac{P(X_1|N)*P(X_2|N)*P(X_3|N)\ldots\ldots P(X_n|N)*P(N)}{P(X_1)*P(X_2)*P(X_3)\ldots\ldots P(X_n)}$$

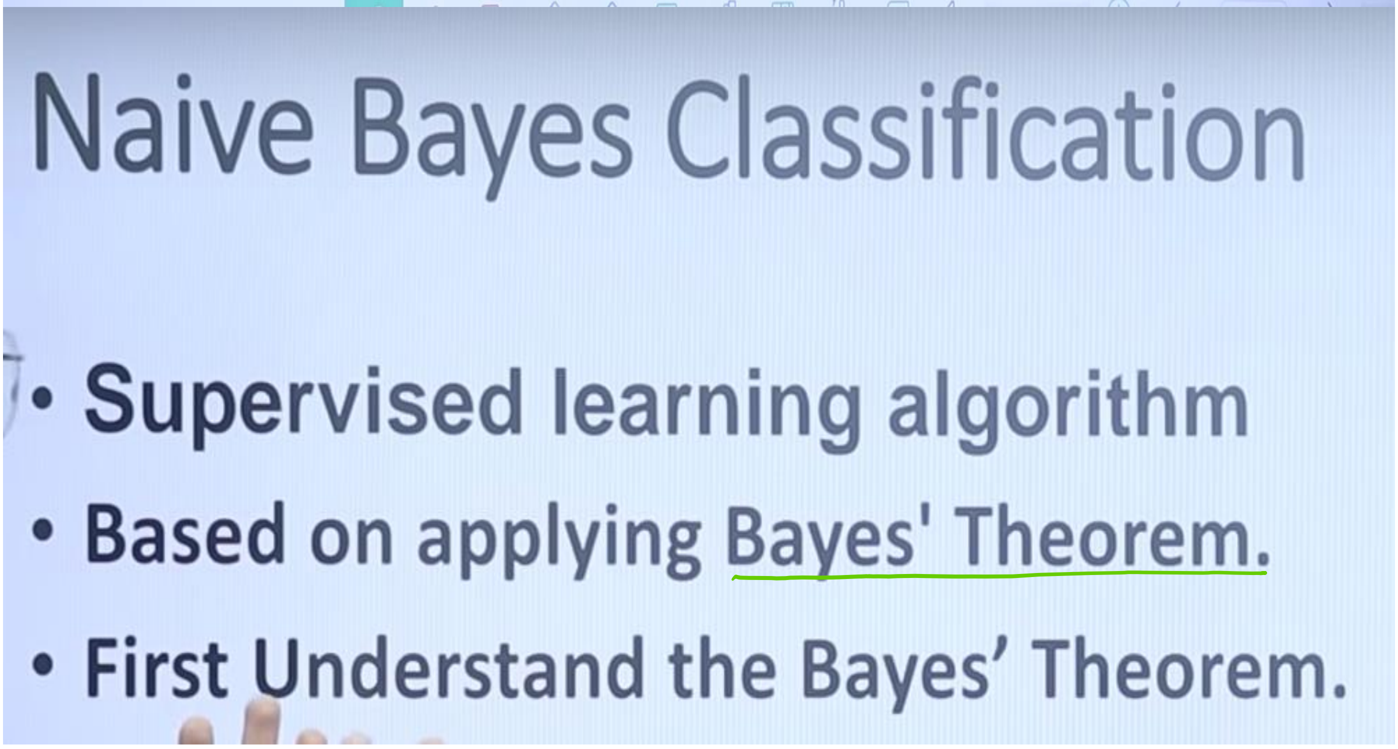


Person	X_1 COVID (Yes/No)	X_2 Flu (Yes/No)	Fever (Yes/No)
1	Yes	No	Yes
2	No	Yes	Yes
3	Yes	Yes	Yes
4	No	No	No
5	Yes	No	Yes
6	No	No	Yes
7	Yes	No	Yes
8	Yes	No	No
9	No	Yes	Yes
10	No	Yes	No

Step-1: Prior Probability:
P(fever = yes) = 7/10
P(fever = no) = 3/10
Step-2: Conditional Probability:

	Yes	No
Covid	4/7	2/3
Flu	3/7	2/3

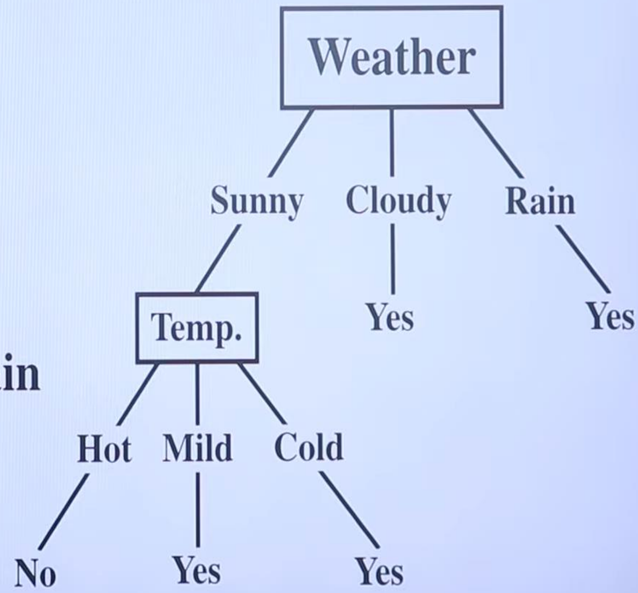
greater value
= .17
12/70 = 6/35
Given Person(flu, Covid) = yes
P(yes | flu, Covid)= P(flu | yes)*P(covid | yes)* P(yes) = 4/7 x 3/7 x 7/10
P(NO | flu, Covid)= P(flu | NO)*P(covid | No)*P(NO) = 2/3 x 2/3 x 3/10 = 4/30 = 2/15 = .13



Naive Bayes Classification
• Supervised learning algorithm
• Based on applying Bayes' Theorem.
• First Understand the Bayes' Theorem.

Decision Tree

- Machine learning algorithm used for both classification and regression tasks.
- Tree Structure
- Decision Nodes
- Leaf Nodes
- Splitting
- Entropy and Information Gain
- Pruning



Day	Weather	Temperature	Humidity	Wind	Play Football?
Day 1	Sunny	Hot	High	Weak	No
Day 2	Sunny	Hot	High	Strong	No
Day 3	Cloudy	Hot	High	Weak	Yes
Day 4	Rain	Mild	High	Weak	Yes
Day 5	Rain	Cool	Normal	Weak	Yes
Day 6	Rain	Cool	Normal	Strong	No
Day 7	Cloudy	Cool	Normal	Strong	Yes
Day 8	Sunny	Mild	High	Weak	No
Day 9	Sunny	Cool	Normal	Weak	Yes
Day 10	Rain	Mild	Normal	Weak	Yes
Day 11	Sunny	Mild	Normal	Strong	Yes
Day 12	Cloudy	~~Mild~~	High	Strong	Yes
Day 13	Cloudy	Hot	Normal	Weak	Yes
Day 14	Rain	Mild	High	Strong	No

Calculate IG of Weather

- **Step1: Entropy of entire dataset**

$$S\{+9,-5\} = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$$

- **Step2: Entropy of all attributes:**
- Entropy of Sunny $\{+2,-3\} = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$
- Entropy of Cloudy$\{+4,-0\} = -\frac{4}{4}\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4} = 0$
- Entropy of Rain$\{+3,-2\} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$
- Information Gain= Entropy(whole data) $-\frac{5}{14}\text{Ent}(S) - \frac{4}{14}\text{Ent}(C) - \frac{5}{14}\text{Ent}(R)$

$$= 0.246$$

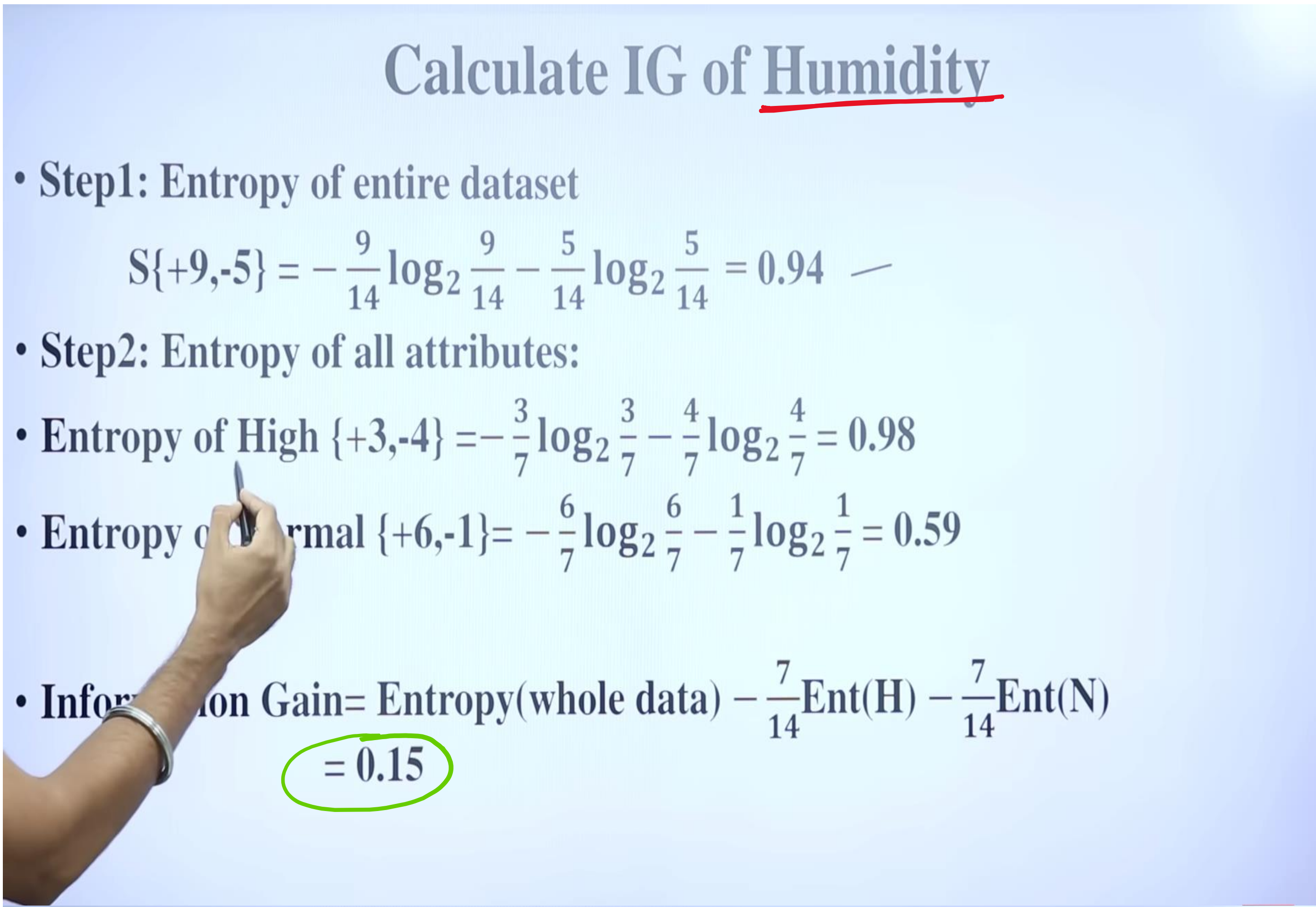
Calculate IG of Temperature

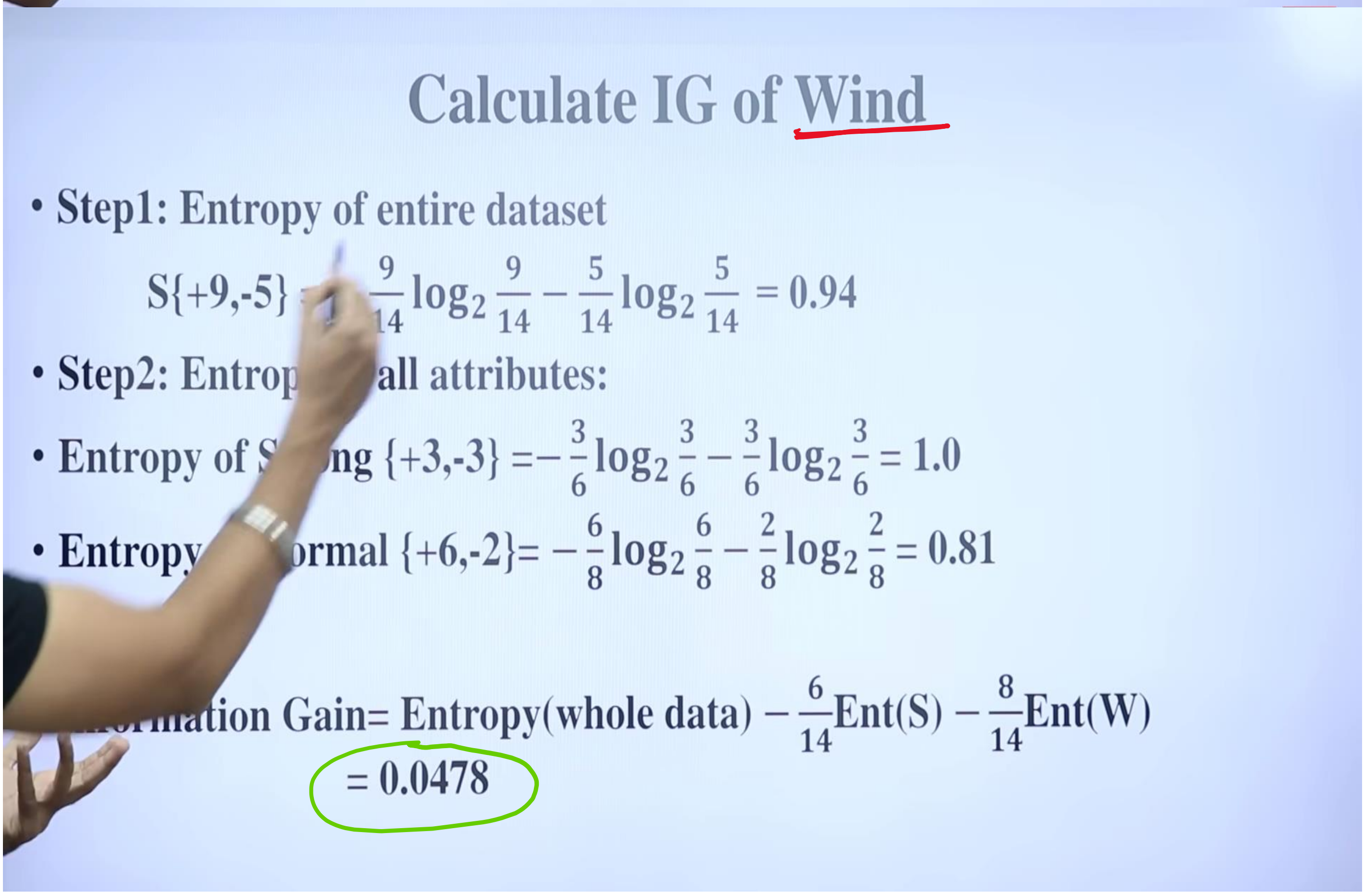
- **Step1: Entropy of entire dataset**

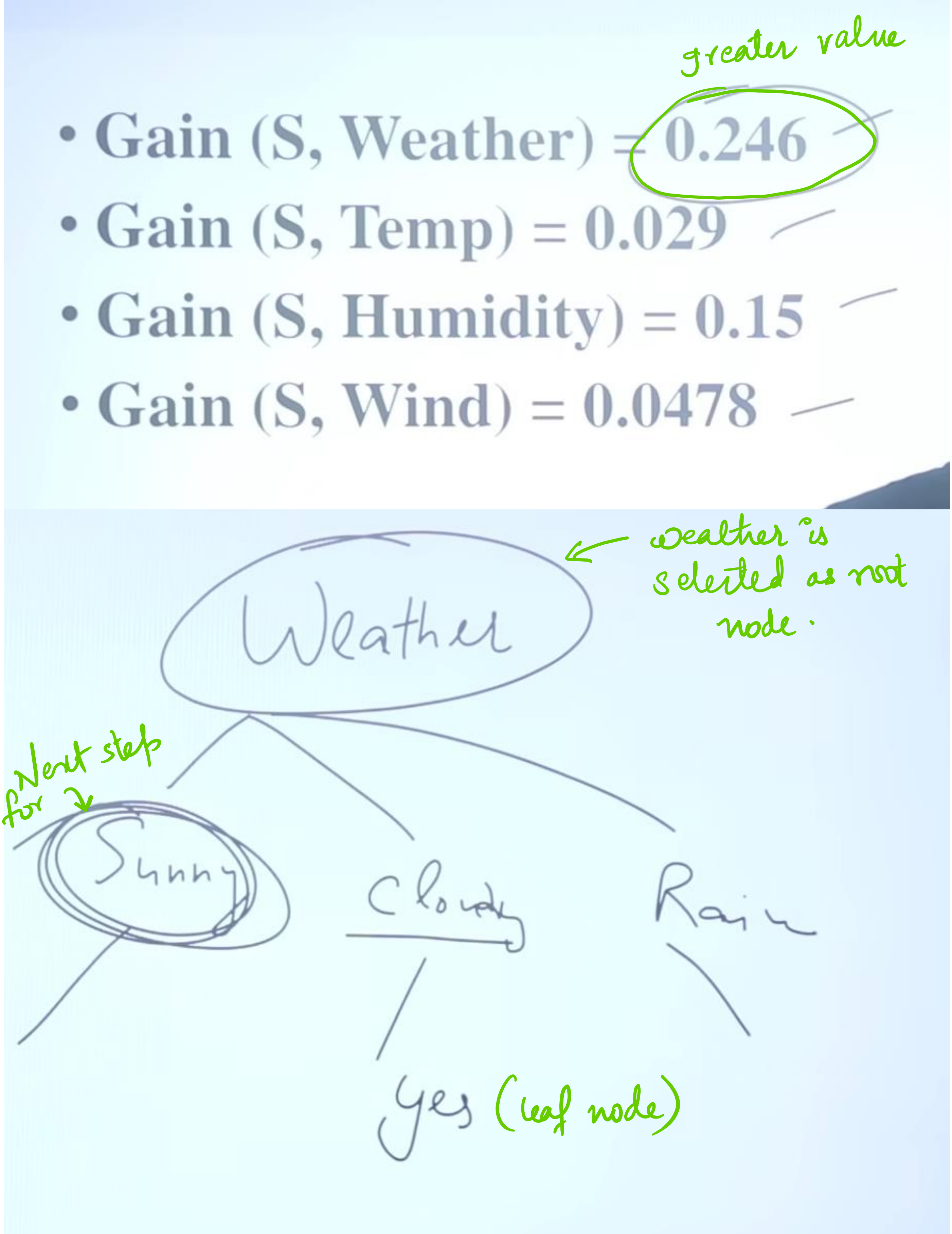
$$S\{+9,-5\} = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$$

- **Step2: Entropy of all attributes:**
- Entropy of Hot $\{+2,-2\} = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1.0$
- Entropy of Mild$\{+4,-2\} = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} = 0.91$
- Entropy of Cold$\{+3,-1\} = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{3}\log_2\frac{1}{3} = 0.81$
- Information Gain= Entropy(whole data) $-\frac{4}{14}\text{Ent}(H) - \frac{6}{14}\text{Ent}(M) - \frac{4}{14}\text{Ent}(C)$

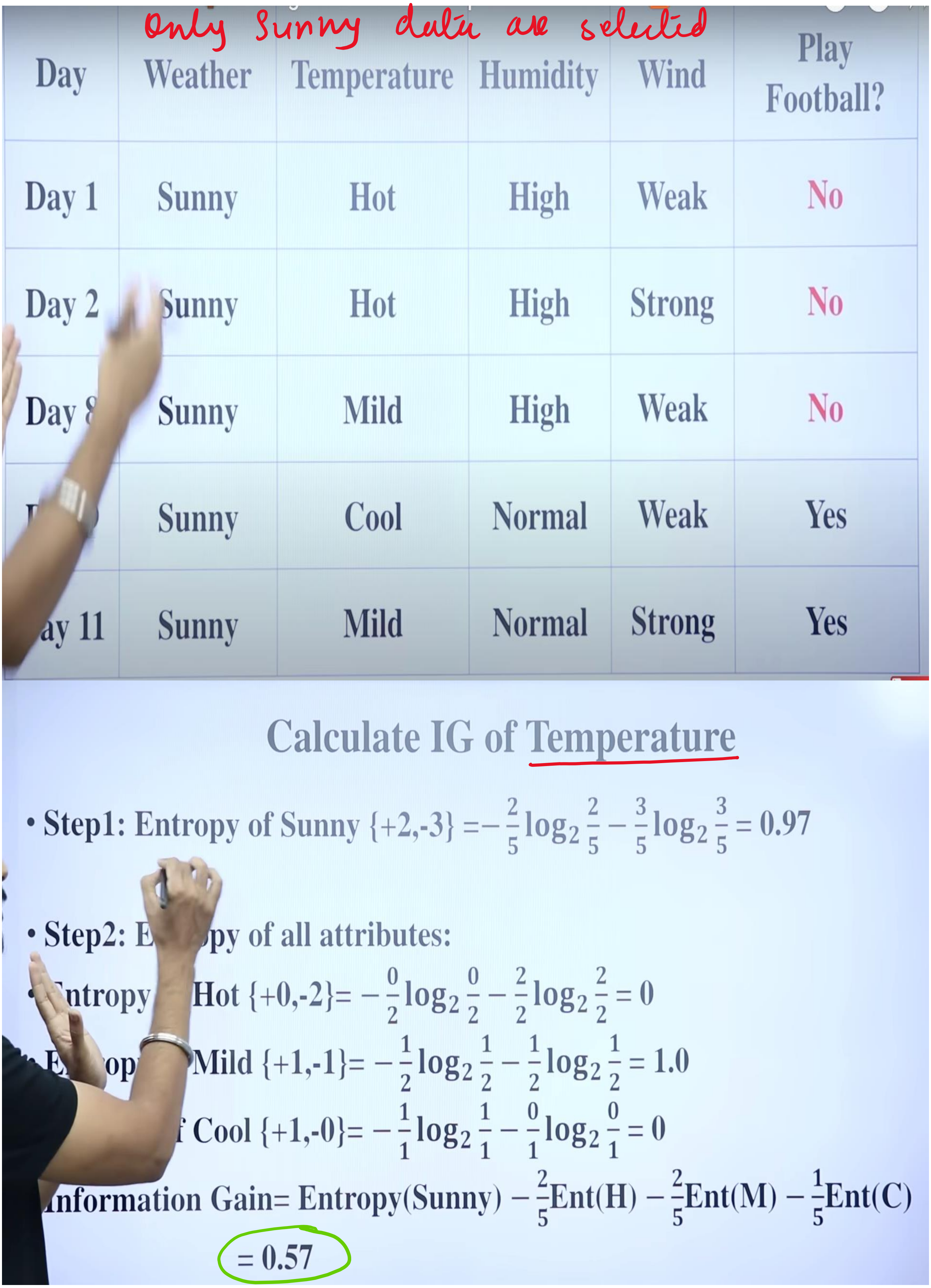
$$= 0.029$$

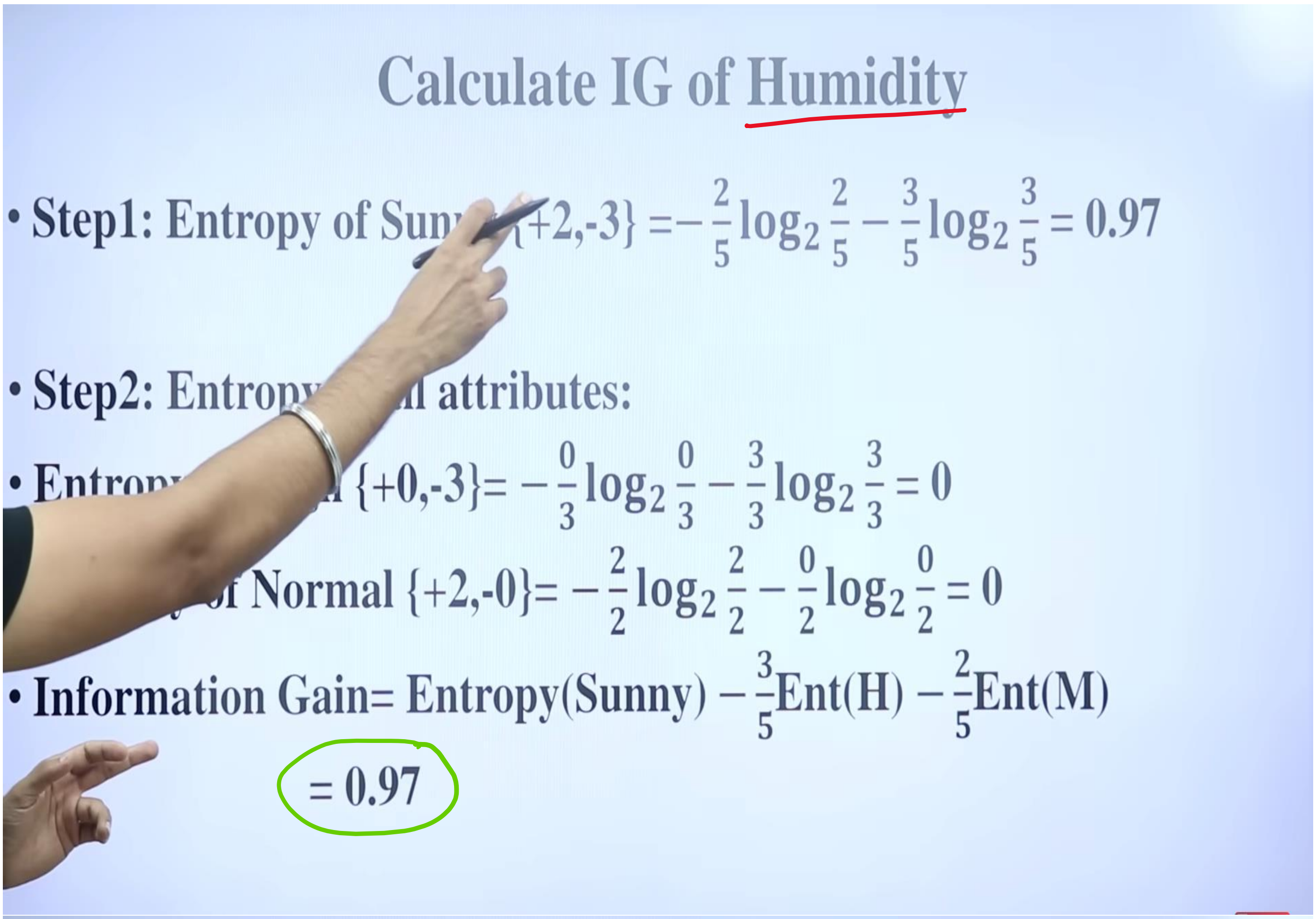

Calculate IG of Humidity
• Step1: Entropy of entire dataset
$S\{+9,-5\} = -\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$
• Step2: Entropy of all attributes:
• Entropy of High $\{+3,-4\} = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.98$
• Entropy o rmal $\{+6,-1\} = -\frac{6}{7}\log_2\frac{6}{7} - \frac{1}{7}\log_2\frac{1}{7} = 0.59$
• Info on Gain= Entropy(whole data) $- \frac{7}{14}Ent(H) - \frac{7}{14}Ent(N)$
$= 0.15$

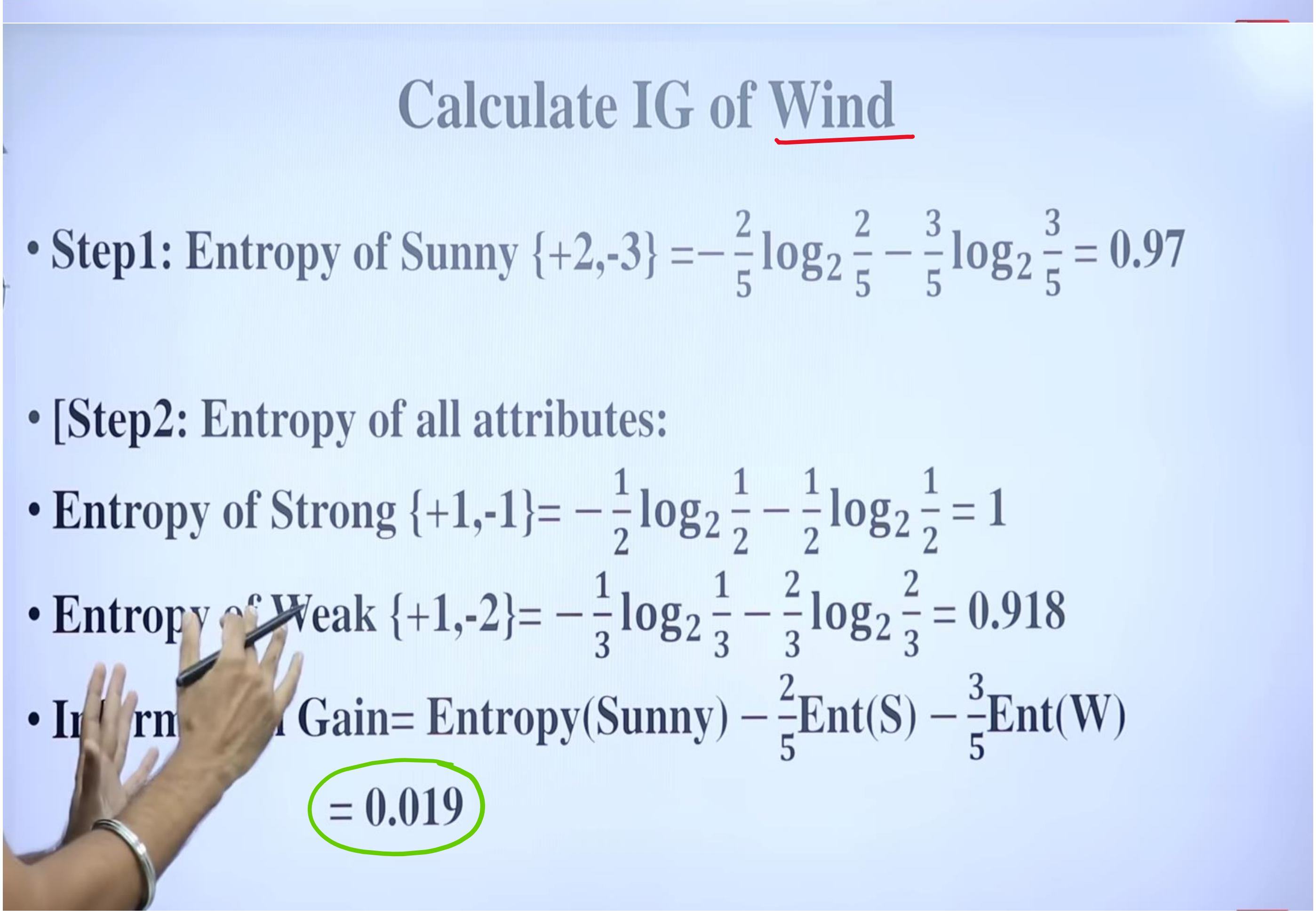

Calculate IG of Wind
• Step1: Entropy of entire dataset
$S\{+9,-5\}$ $\frac{9}{14}\log_2\frac{9}{14} - \frac{5}{14}\log_2\frac{5}{14} = 0.94$
• Step2: Entrop all attributes:
• Entropy of S ng $\{+3,-3\} = -\frac{3}{6}\log_2\frac{3}{6} - \frac{3}{6}\log_2\frac{3}{6} = 1.0$
• Entropy ormal $\{+6,-2\} = -\frac{6}{8}\log_2\frac{6}{8} - \frac{2}{8}\log_2\frac{2}{8} = 0.81$
mation Gain= Entropy(whole data) $- \frac{6}{14}Ent(S) - \frac{8}{14}Ent(W)$
$= 0.0478$

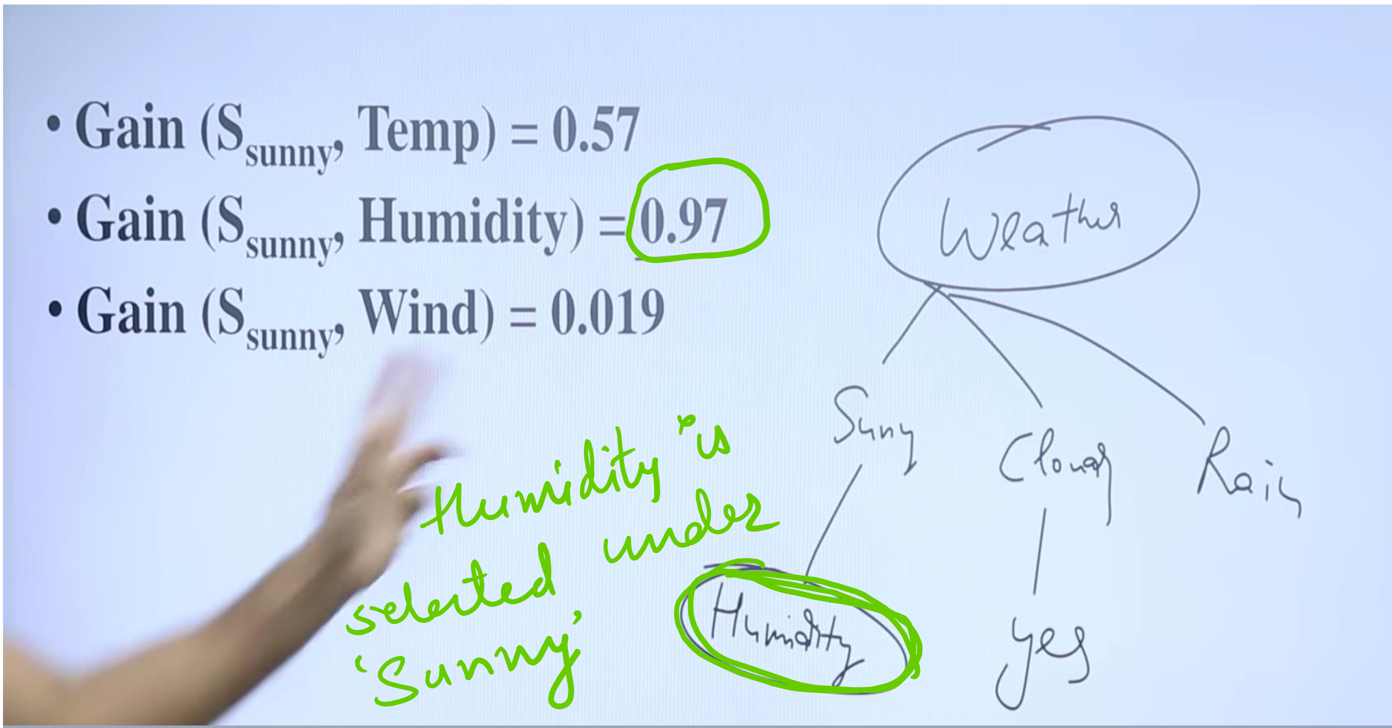


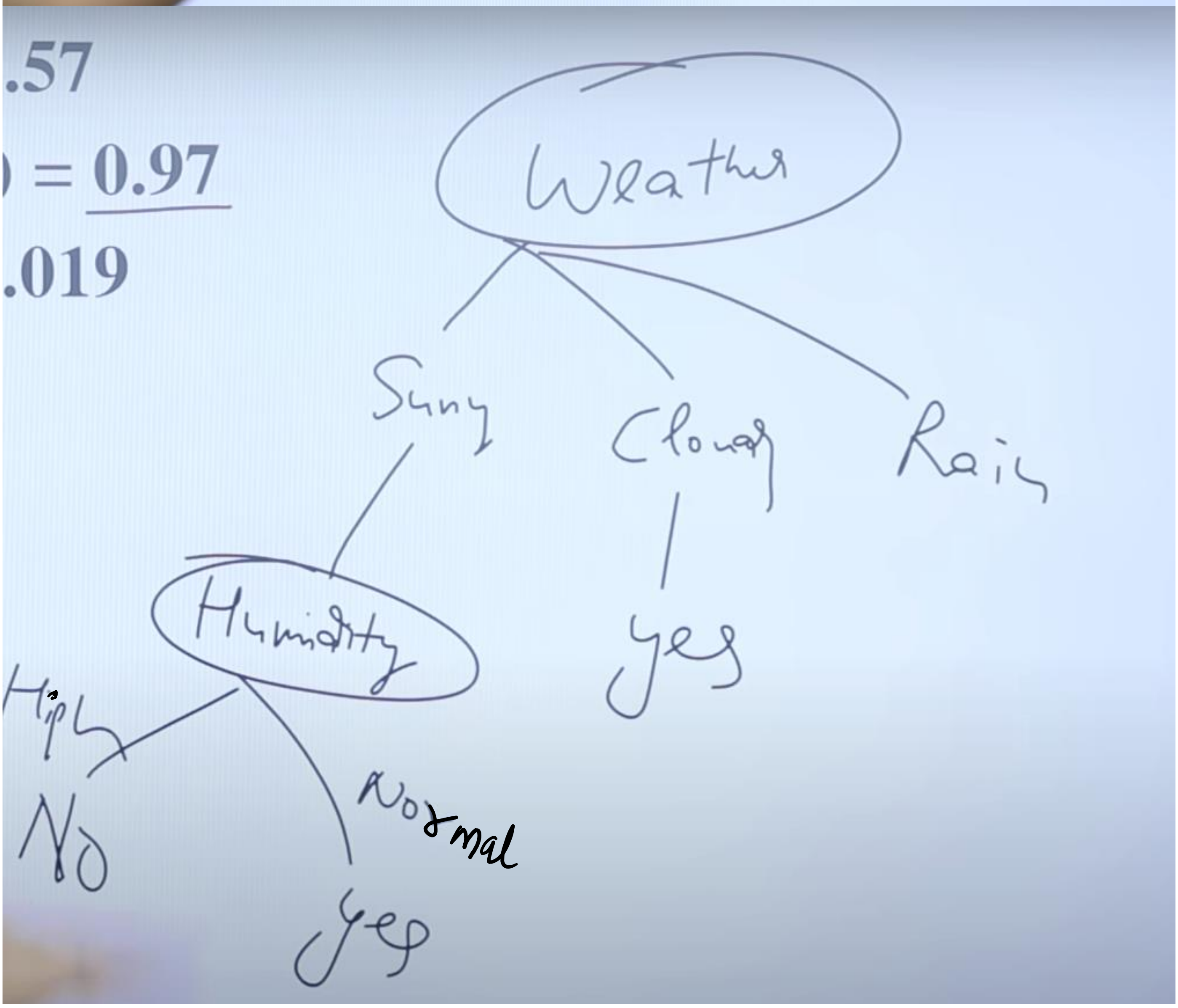
greater value
• Gain (S, Weather) = 0.246
• Gain (S, Temp) = 0.029
• Gain (S, Humidity) = 0.15
• Gain (S, Wind) = 0.0478
weather is selected as root node.
Weather
Next step for
Sunny
Cloudy
Rain
yes (leaf node)

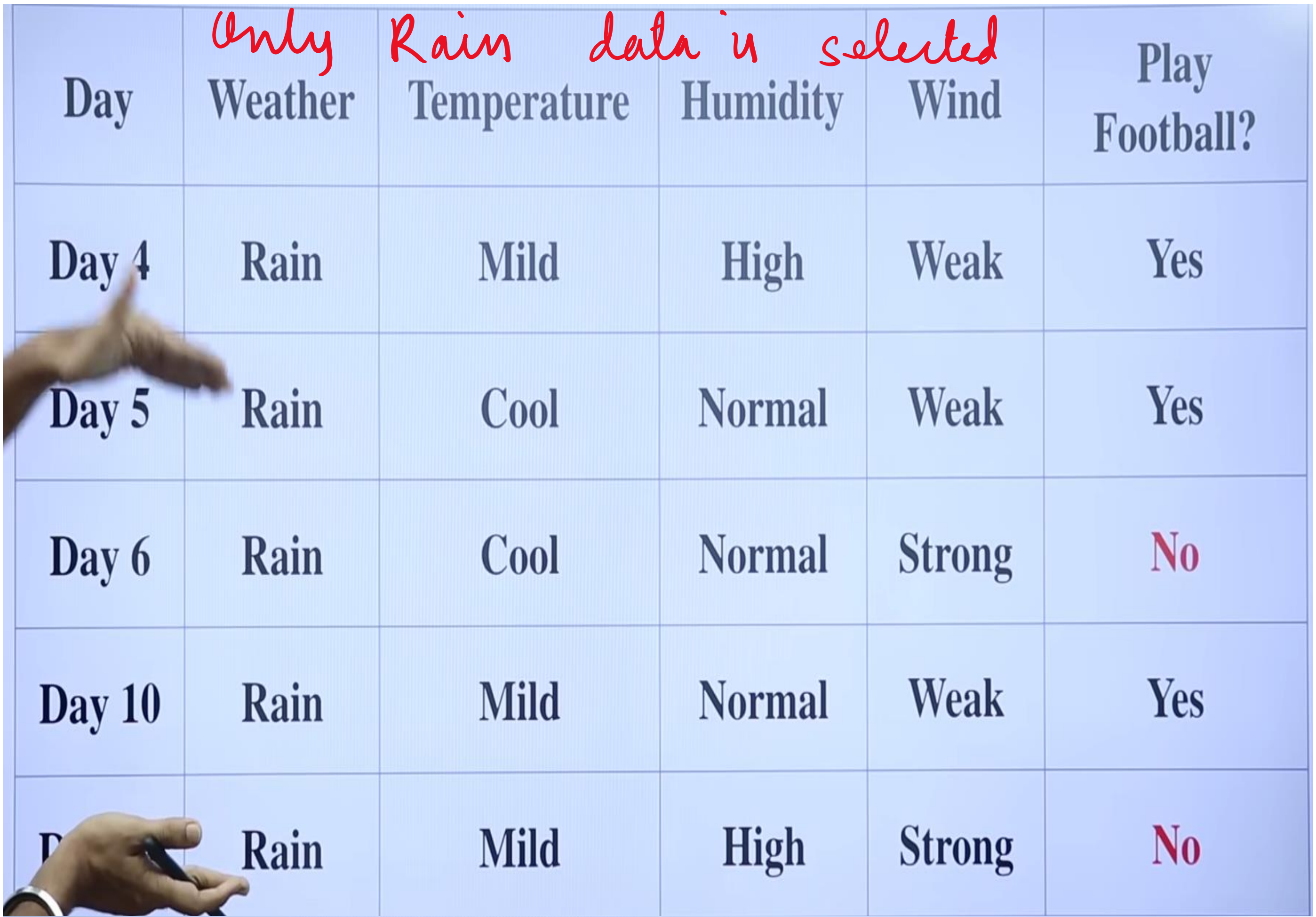

Only Sunny data are selected
Day | Weather | Temperature | Humidity | Wind | Play Football?
Day 1 | Sunny | Hot | High | Weak | No
Day 2 | Sunny | Hot | High | Strong | No
Day | Sunny | Mild | High | Weak | No
Sunny | Cool | Normal | Weak | Yes
ay 11 | Sunny | Mild | Normal | Strong | Yes
Calculate IG of Temperature
• Step1: Entropy of Sunny {+2,-3} $= -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.97$
• Step2: E py of all attributes:
ntropy Hot {+0,-2}= $-\frac{0}{2}\log_2\frac{0}{2} - \frac{2}{2}\log_2\frac{2}{2} = 0$
Mild {+1,-1}= $-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1.0$
Cool {+1,-0}= $-\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1} = 0$
nformation Gain= Entropy(Sunny) $- \frac{2}{5}$Ent(H) $- \frac{2}{5}$Ent(M) $- \frac{1}{5}$Ent(C)
= 0.57


Calculate IG of Humidity
• Step1: Entropy of Sun
+2,-3} $=-\frac{2}{5}\log_2\frac{2}{5}-\frac{3}{5}\log_2\frac{3}{5}=0.97$
• Step2: Entrop
l attributes:
• Entrop
{+0,-3}= $-\frac{0}{3}\log_2\frac{0}{3}-\frac{3}{3}\log_2\frac{3}{3}=0$
Normal {+2,-0}= $-\frac{2}{2}\log_2\frac{2}{2}-\frac{0}{2}\log_2\frac{0}{2}=0$
• Information Gain= Entropy(Sunny) $-\frac{3}{5}$Ent(H) $-\frac{2}{5}$Ent(M)
= 0.97

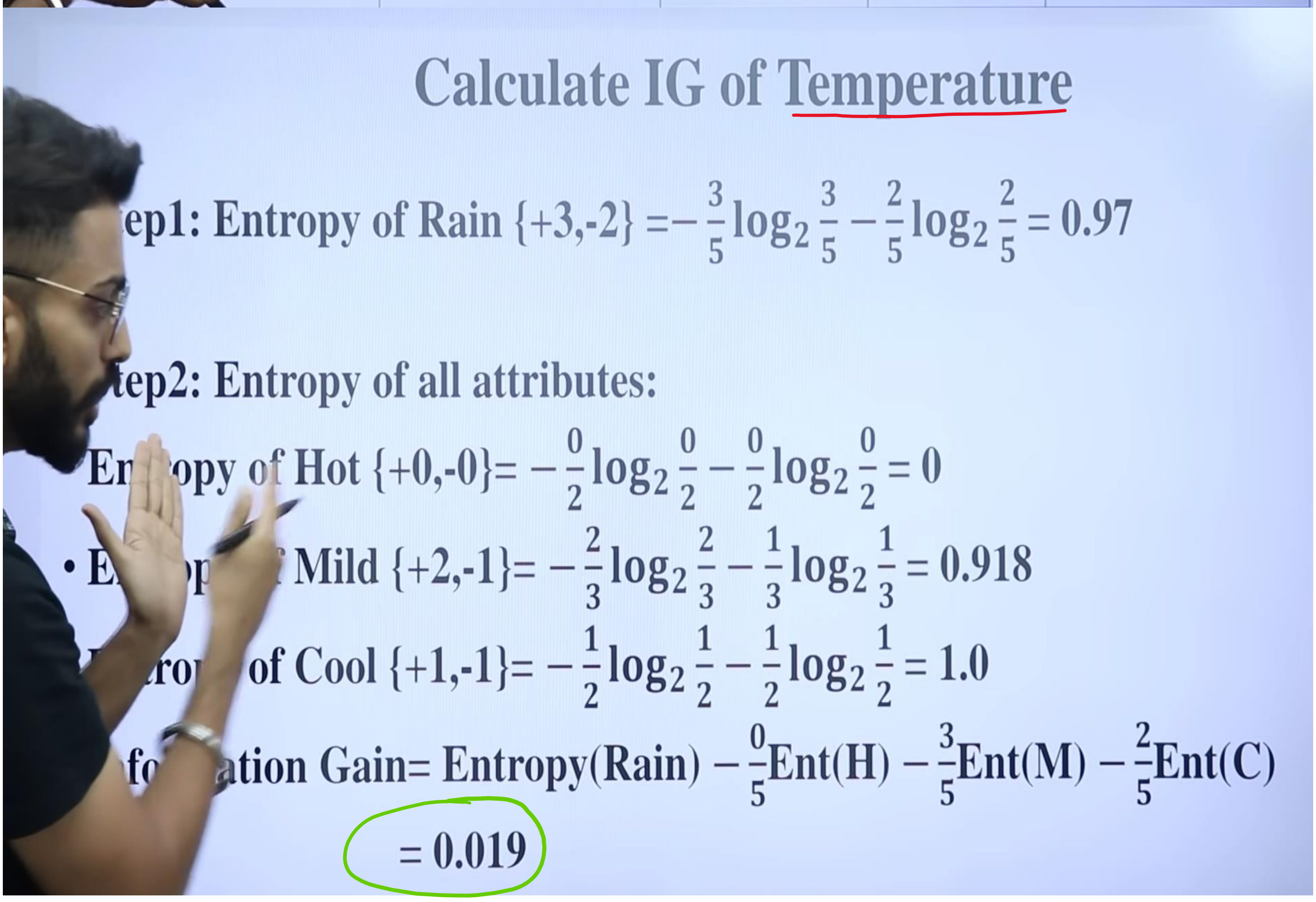

Calculate IG of Wind
• Step1: Entropy of Sunny {+2,-3} $=-\frac{2}{5}\log_2\frac{2}{5}-\frac{3}{5}\log_2\frac{3}{5}=0.97$
• [Step2: Entropy of all attributes:
• Entropy of Strong {+1,-1}= $-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}=1$
• Entrop
eak {+1,-2}= $-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}=0.918$
Gain= Entropy(Sunny) $-\frac{2}{5}$Ent(S) $-\frac{3}{5}$Ent(W)
= 0.019

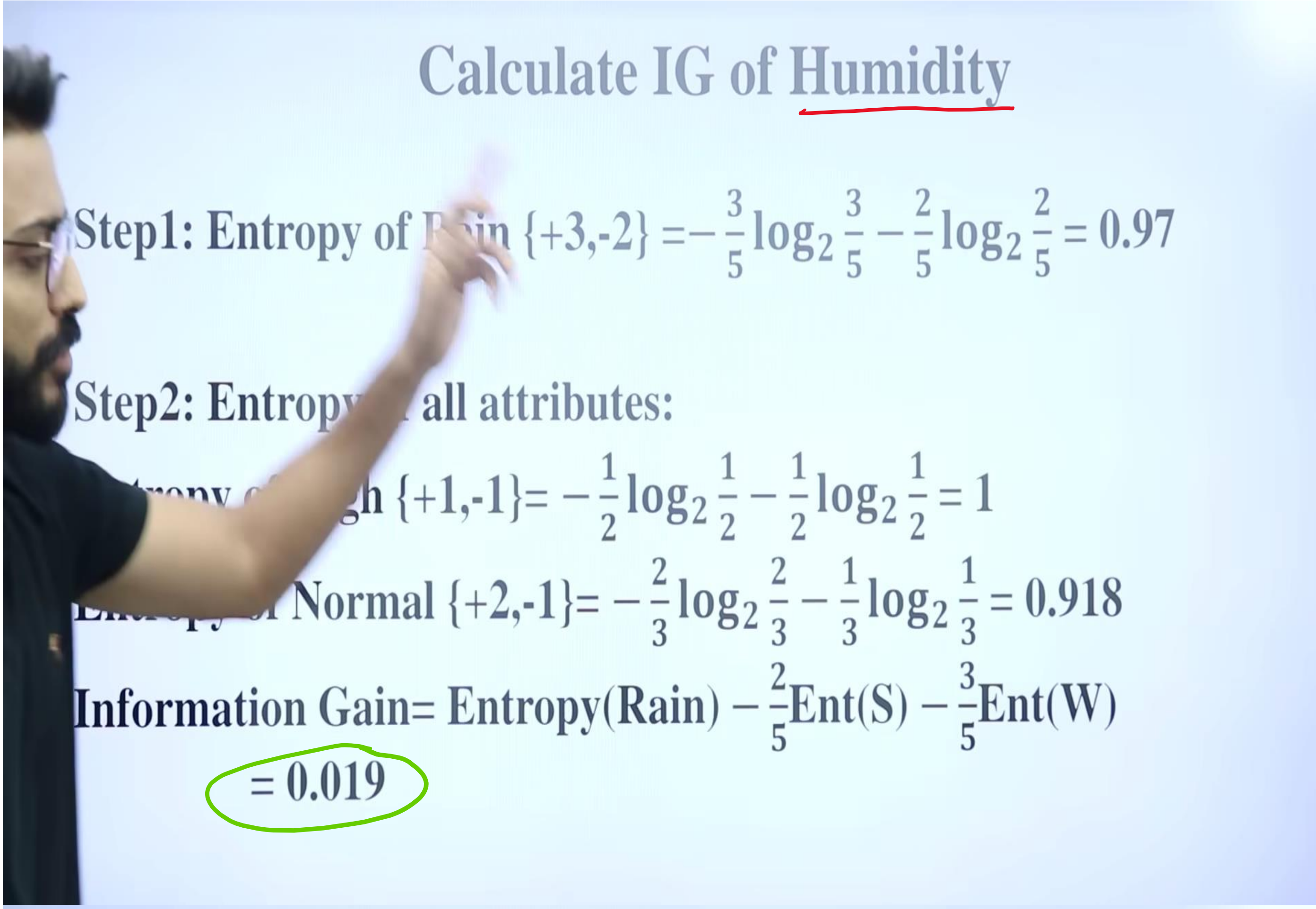

• Gain (S_{sunny}, Temp) = 0.57
• Gain (S_{sunny}, Humidity) = 0.97
• Gain (S_{sunny}, Wind) = 0.019
Weather
Sunny
Cloudy
Rain
Humidity
yes
Humidity is selected under 'Sunny'

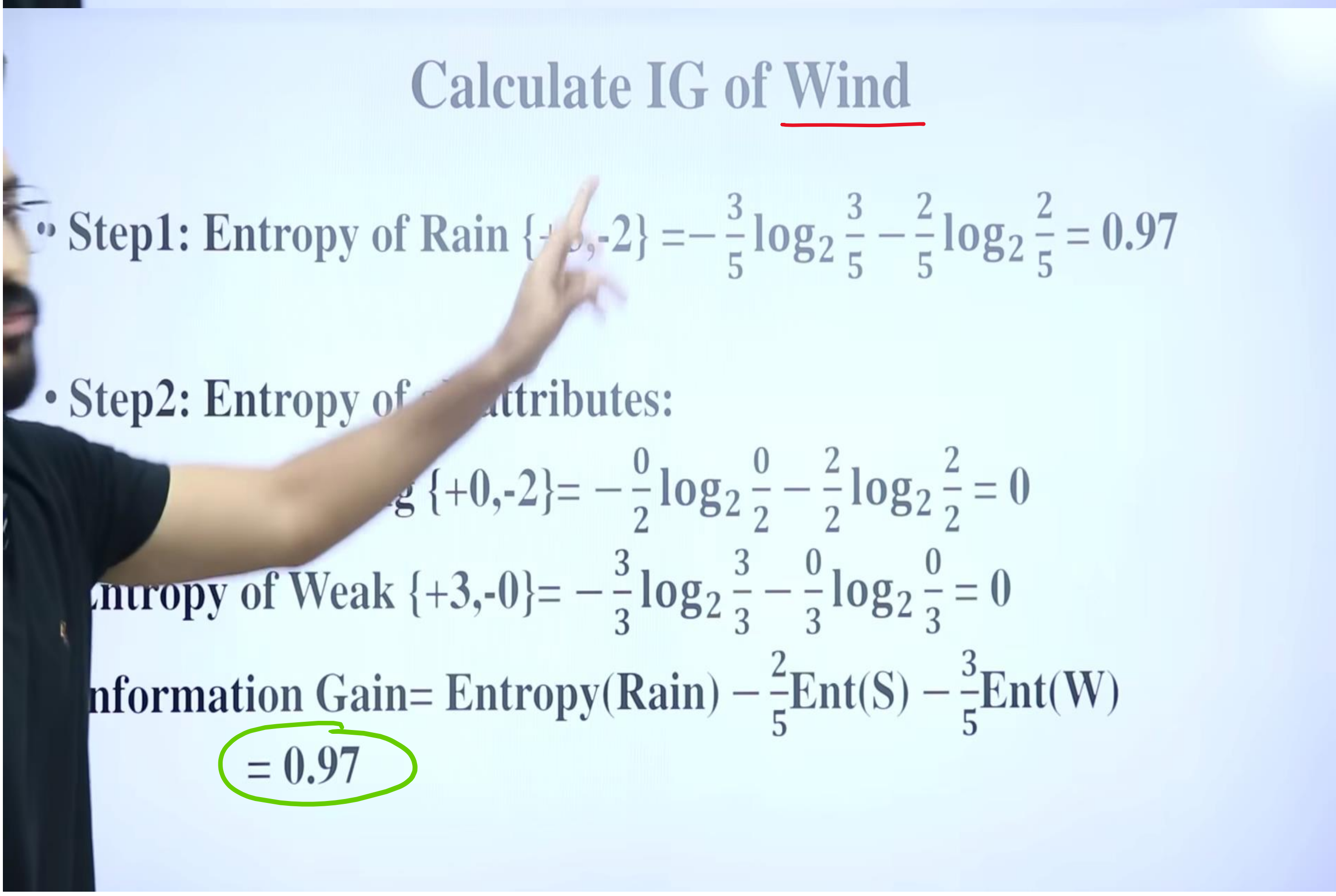

.57
= 0.97
.019
Weather
Sunny
Cloudy
Rain
Humidity
yes
High
Normal
No
yes

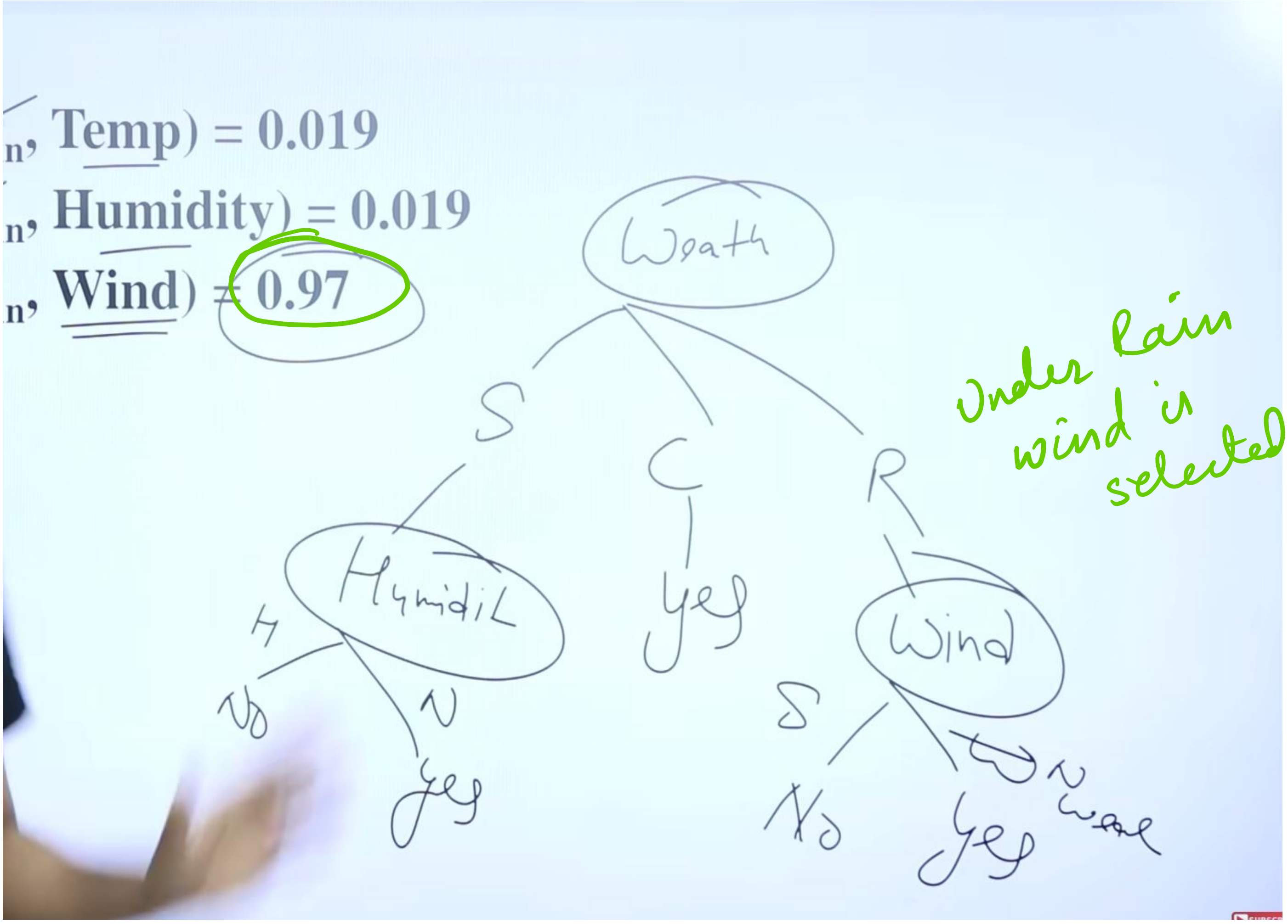


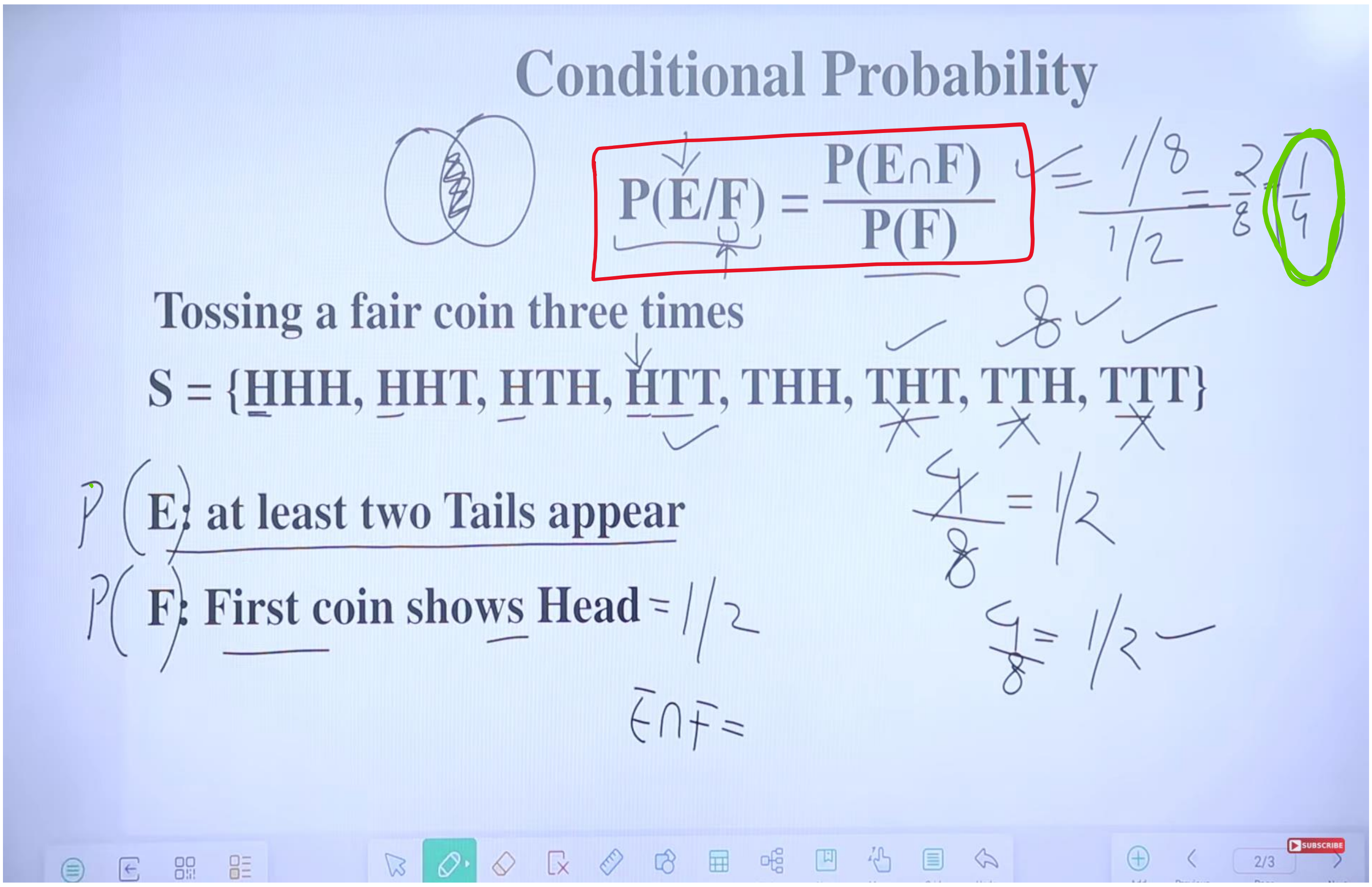
Day	Weather	Temperature	Humidity	Wind	Play Football?
Day 4	Rain	Mild	High	Weak	Yes
Day 5	Rain	Cool	Normal	Weak	Yes
Day 6	Rain	Cool	Normal	Strong	No
Day 10	Rain	Mild	Normal	Weak	Yes
[illegible]	Rain	Mild	High	Strong	No

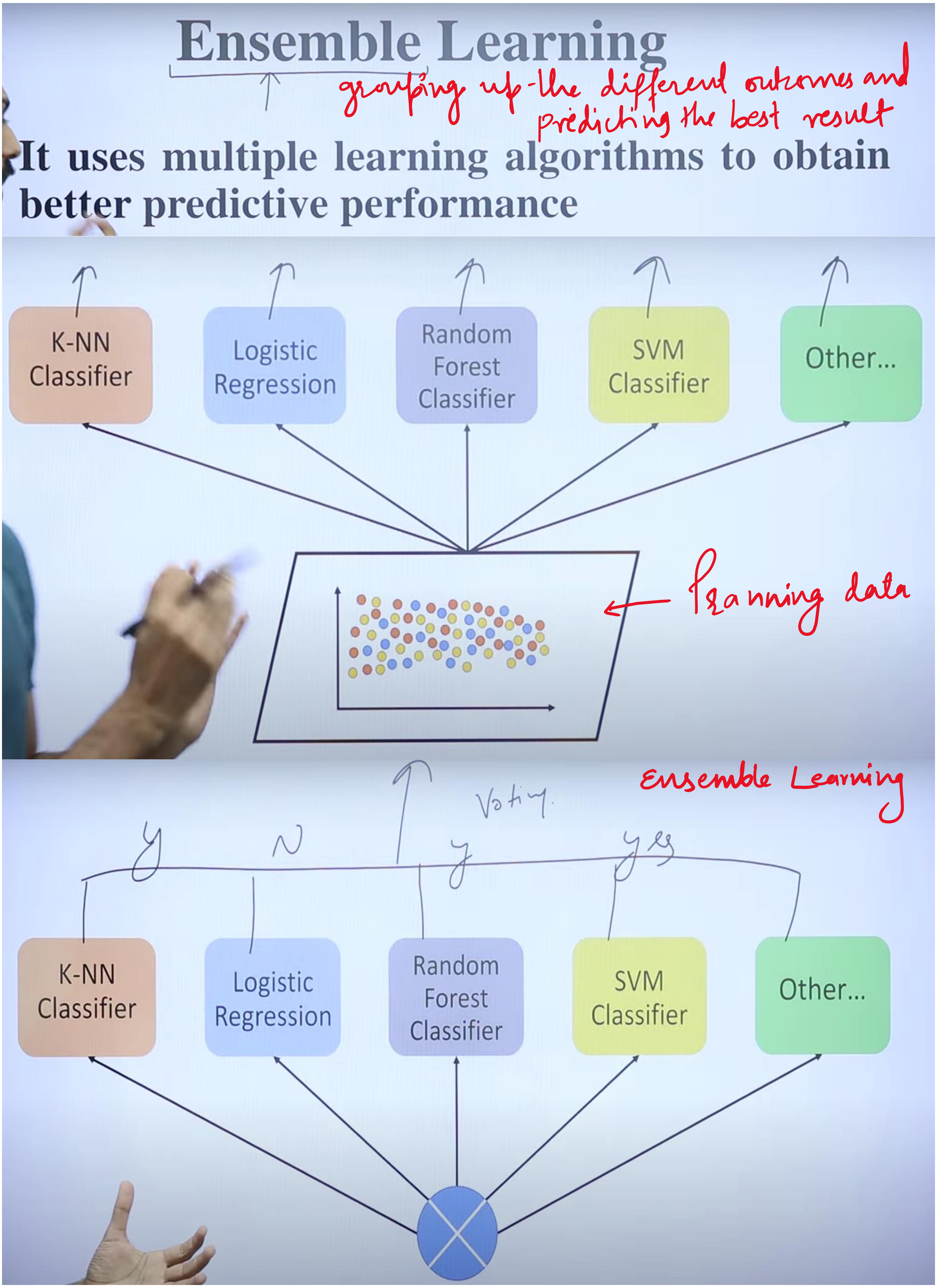


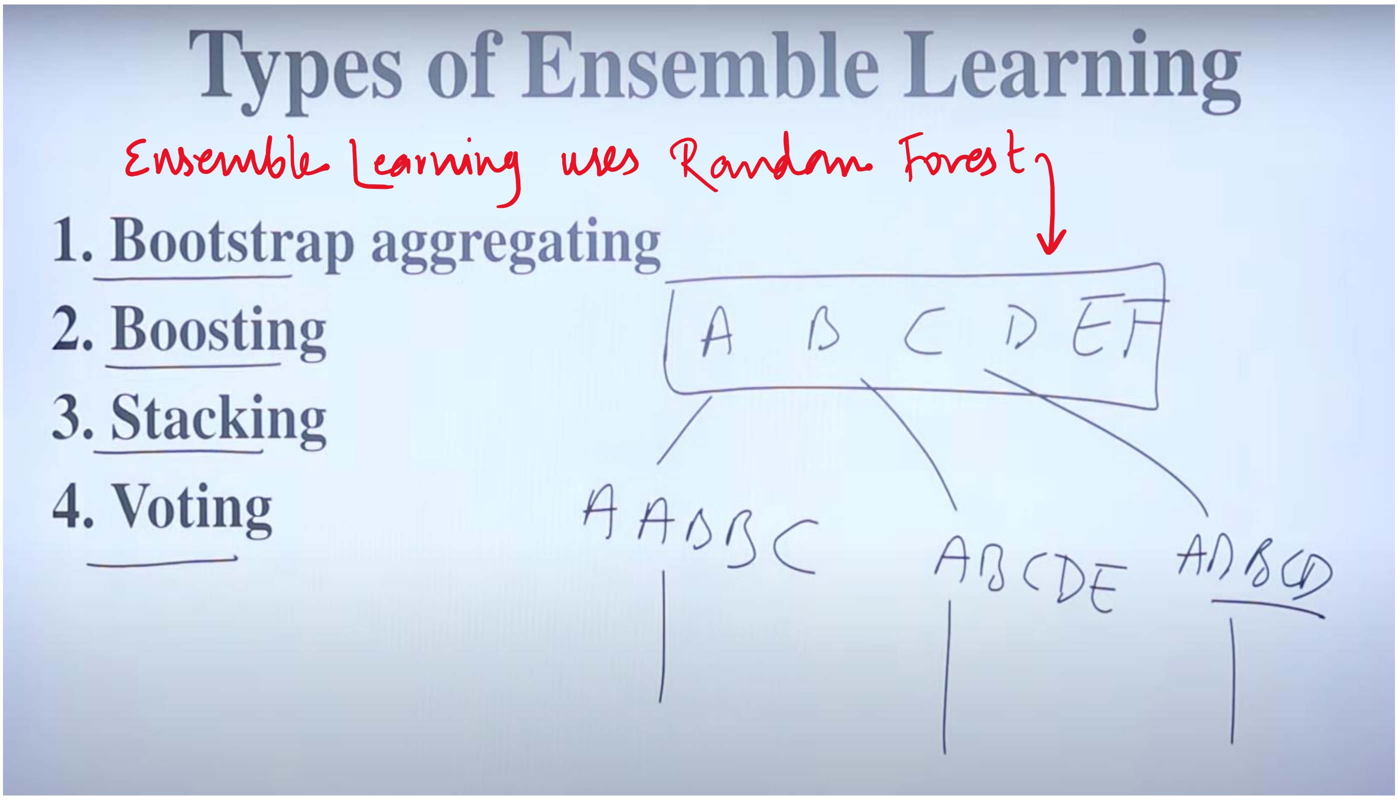

Calculate IG of Humidity
Step1: Entropy of
$\{+3,-2\} = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$
Step2: Entropy
all attributes:
$\{+1,-1\} = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$
Normal $\{+2,-1\} = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918$
Information Gain= Entropy(Rain) $-\frac{2}{5}$Ent(S) $-\frac{3}{5}$Ent(W)
$= 0.019$

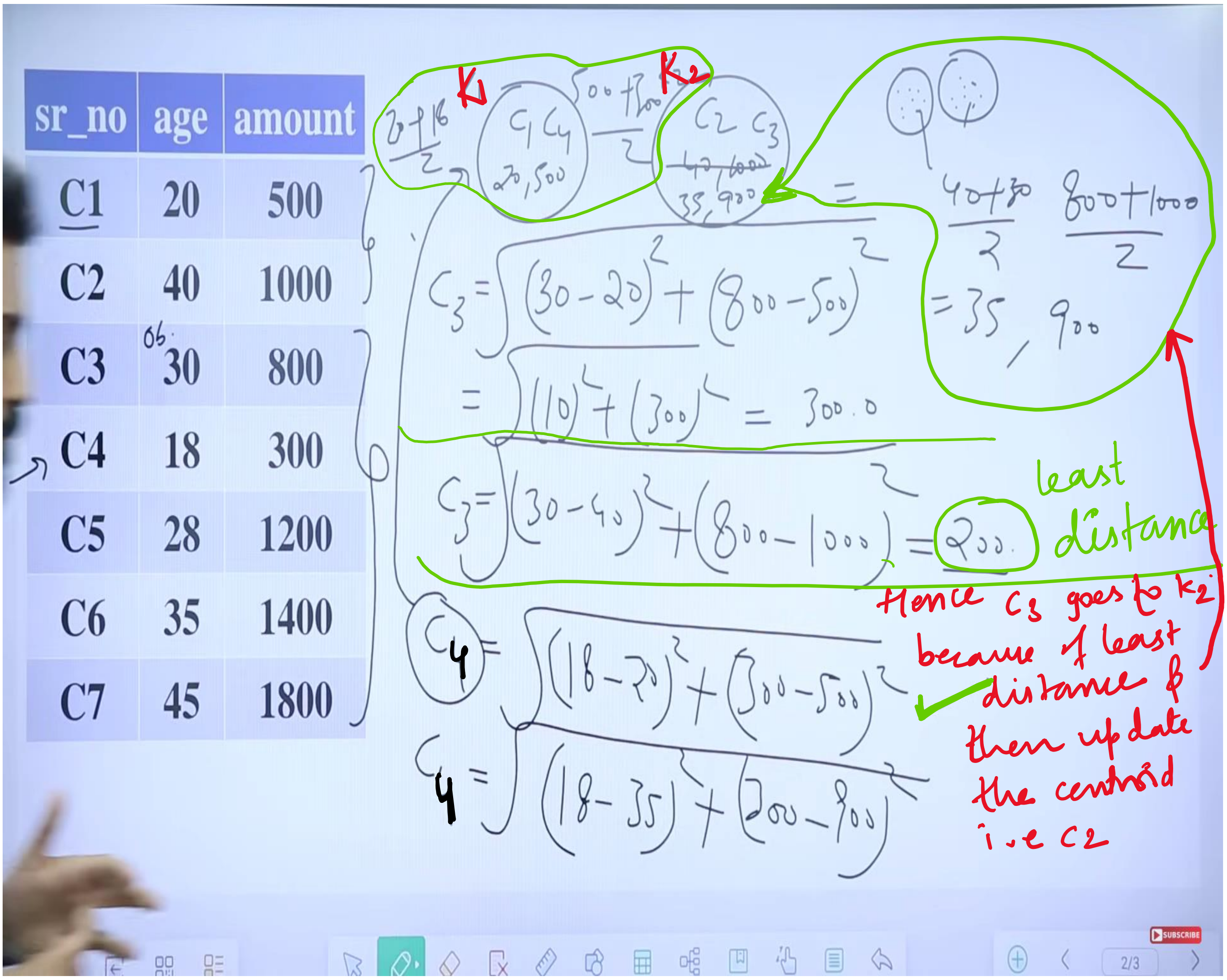

Calculate IG of Wind
Step1: Entropy of Rain
$= -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.97$
Step2: Entropy of
ttributes:
$\{+0,-2\} = -\frac{0}{2}\log_2\frac{0}{2} - \frac{2}{2}\log_2\frac{2}{2} = 0$
ntropy of Weak $\{+3,-0\} = -\frac{3}{3}\log_2\frac{3}{3} - \frac{0}{3}\log_2\frac{0}{3} = 0$
nformation Gain= Entropy(Rain) $-\frac{2}{5}$Ent(S) $-\frac{3}{5}$Ent(W)
$= 0.97$

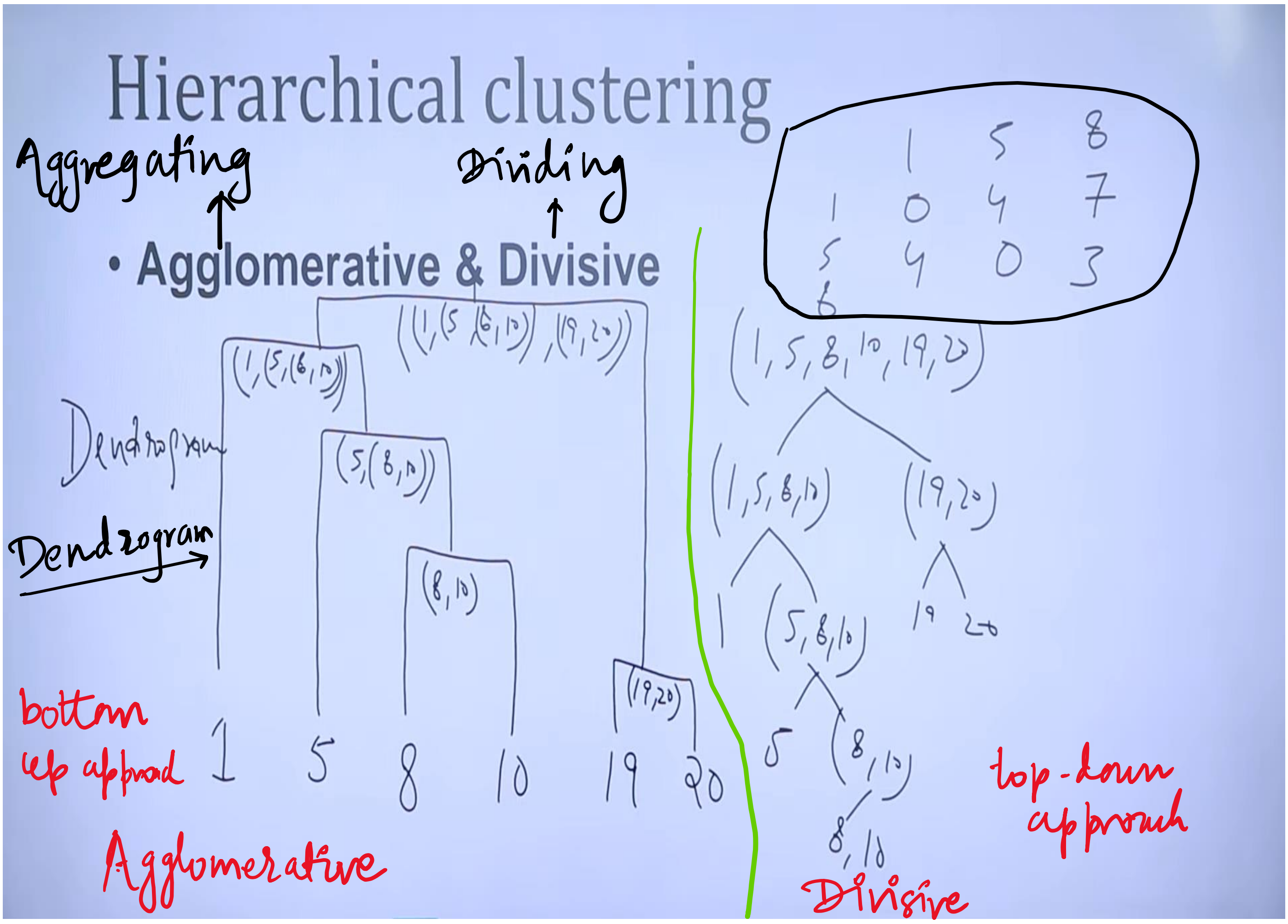

n, Temp) = 0.019
n, Humidity) = 0.019
n, Wind) = 0.97
Weath
S
C
R
Humidit
yes
Wind
H
No
N
yes
S
No
yes
Weak
Under Rain wind is selected

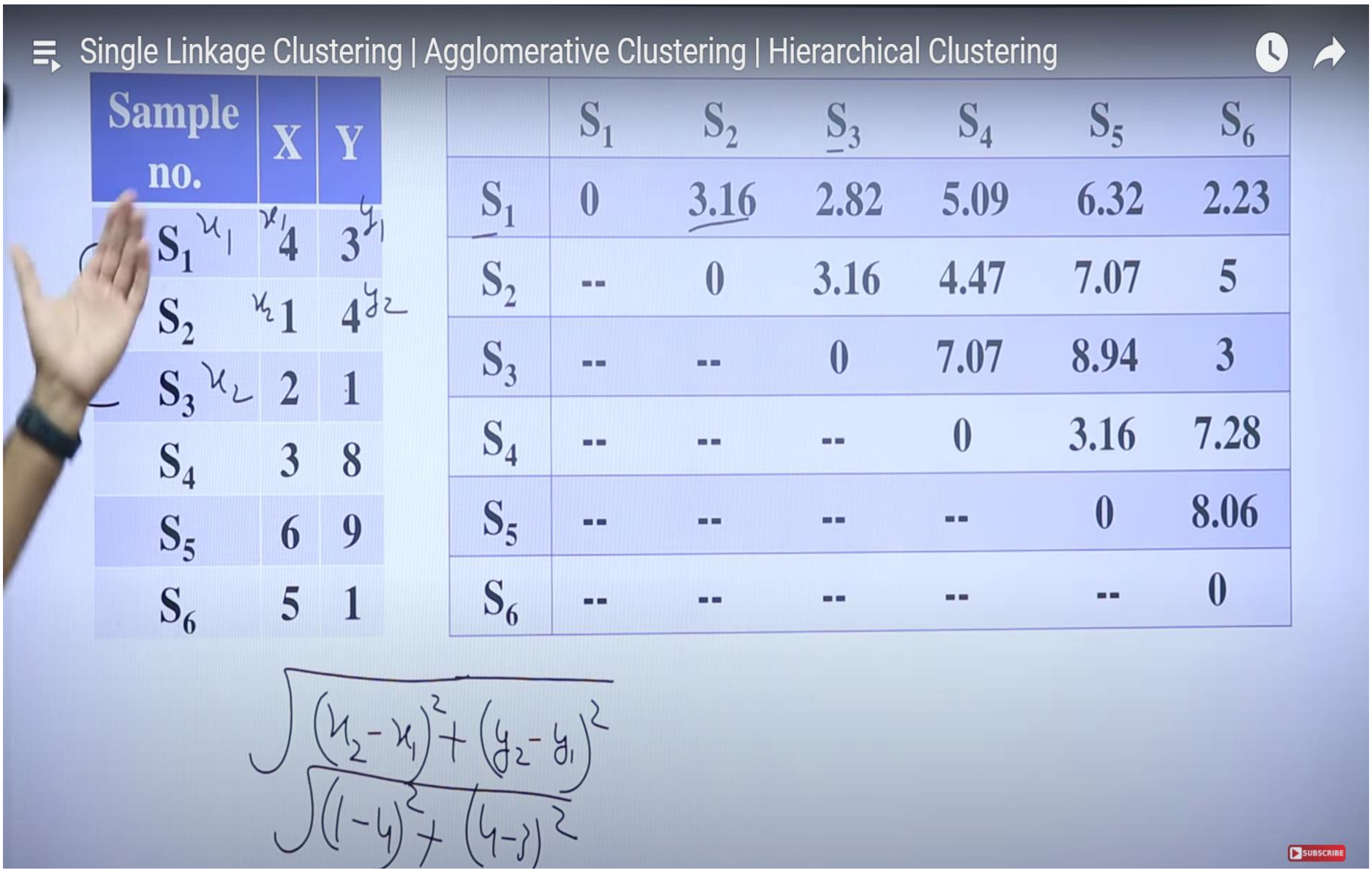

Conditional Probability
P(E/F) = P(E∩F) / P(F)
= 1/8 / 1/2 = 2/8 = 1/4
Tossing a fair coin three times
S = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}
P(E: at least two Tails appear)
P(F: First coin shows Head) = 1/2
4/8 = 1/2
4/8 = 1/2
E∩F =


Ensemble Learning
grouping up the different outcomes and predicting the best result
It uses multiple learning algorithms to obtain better predictive performance
K-NN Classifier
Logistic Regression
Random Forest Classifier
SVM Classifier
Other...
Tranning data
Ensemble Learning
Voting
Y
N
Y
yes
K-NN Classifier
Logistic Regression
Random Forest Classifier
SVM Classifier
Other...


Types of Ensemble Learning
Ensemble Learning uses Random Forest
1. Bootstrap aggregating
2. Boosting
3. Stacking
4. Voting
A B C D E F
AABBC
ABCDE
ADBCD


sr_no age amount
C1 20 500
C2 40 1000
C3 30 800
C4 18 300
C5 28 1200
C6 35 1400
C7 45 1800
K1
K2
C1 C4
C2 C3
= 40+30/2 800+1000/2
= 35, 900
= 300.0
= 200
least distance
Hence C3 goes to k2 because of least distance & then update the centroid i.e C2





Sample no.	X	Y
S_1	4	3
S_2	1	4
S_3	2	1
S_4	3	8
S_5	6	9
S_6	5	1

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	0	3.16	2.82	5.09	6.32	2.23
S_2	--	0	3.16	4.47	7.07	5
S_3	--	--	0	7.07	8.94	3
S_4	--	--	--	0	3.16	7.28
S_5	--	--	--	--	0	8.06
S_6	--	--	--	--	--	0

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

$$\sqrt{(1-4)^2+(4-3)^2}$$

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	0	3.16	2.82	5.09	6.32	2.23
S_2	--	0	3.16	4.47	7.07	5
S_3	--	--	0	7.07	8.94	3
S_4	--	--	--	0	3.16	7.28
S_5	--	--	--	--	0	8.06
[illegible]	--	--	--	--	--	0

Minimum distance

$+ (y_2 - y_1)^2$ (S_1, S_6)

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	0	3.16	2.82	5.09	6.32	2.23
S_2	--	0	3.16	4.47	7.07	5
S_3	--	--	0	7.07	8.94	3
S_4	--	--	--	0	3.16	7.28
S_5	--	--	--	--	0	8.06
S_6	--	--	--	--	--	0

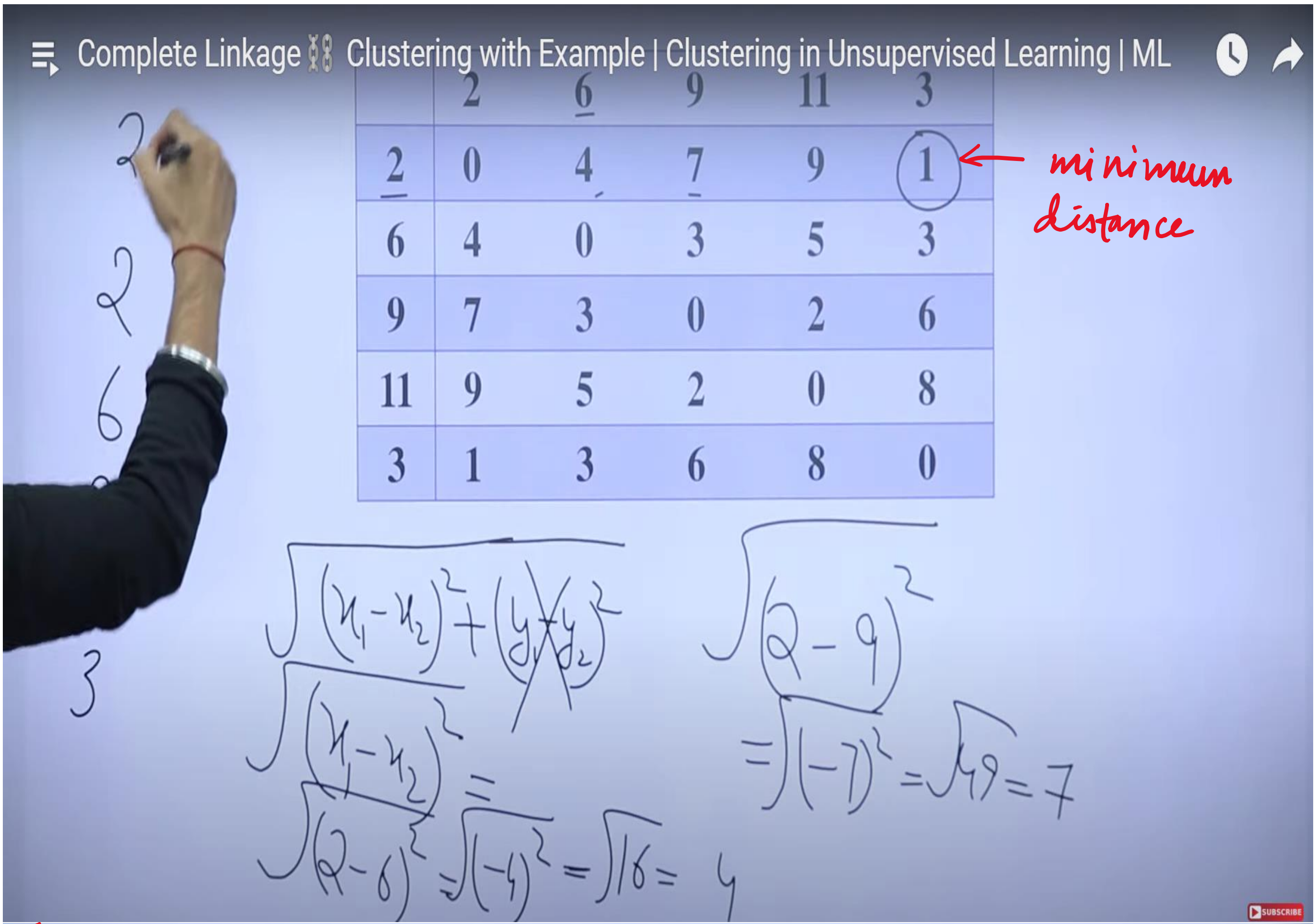
	$(S_1\ S_6)$	S_2	S_3	S_4	S_5
$(S_1\ S_6)$	0	3.16	2.82	5.09	6.32
S_2	--	0	3.16	4.47	7.07
S_3	--	--	0	7.07	8.94
S_4	--	--	--	0	3.16
S_5	--	--	--	--	0

	$(S_1\ S_6)$	S_2	S_3	S_4	S_5
$(S_1\ S_6)$	0	3.16	2.82	5.09	6.32
S_2	--	0	3.16	4.47	7.07
S_3	--	--	0	7.07	8.94
S_4	--	--	--	0	3.16
S_5	--	--	--	--	0

	$(S_1\ S_6)S_3$	S_2	S_4	S_5
$(S_1\ S_6)S_3$	0	[illegible]	5.09	6.32
S_2	[illegible]	[illegible]	4.47	7.07
[illegible]	--	--	0	3.16
S_5	--	--	--	0

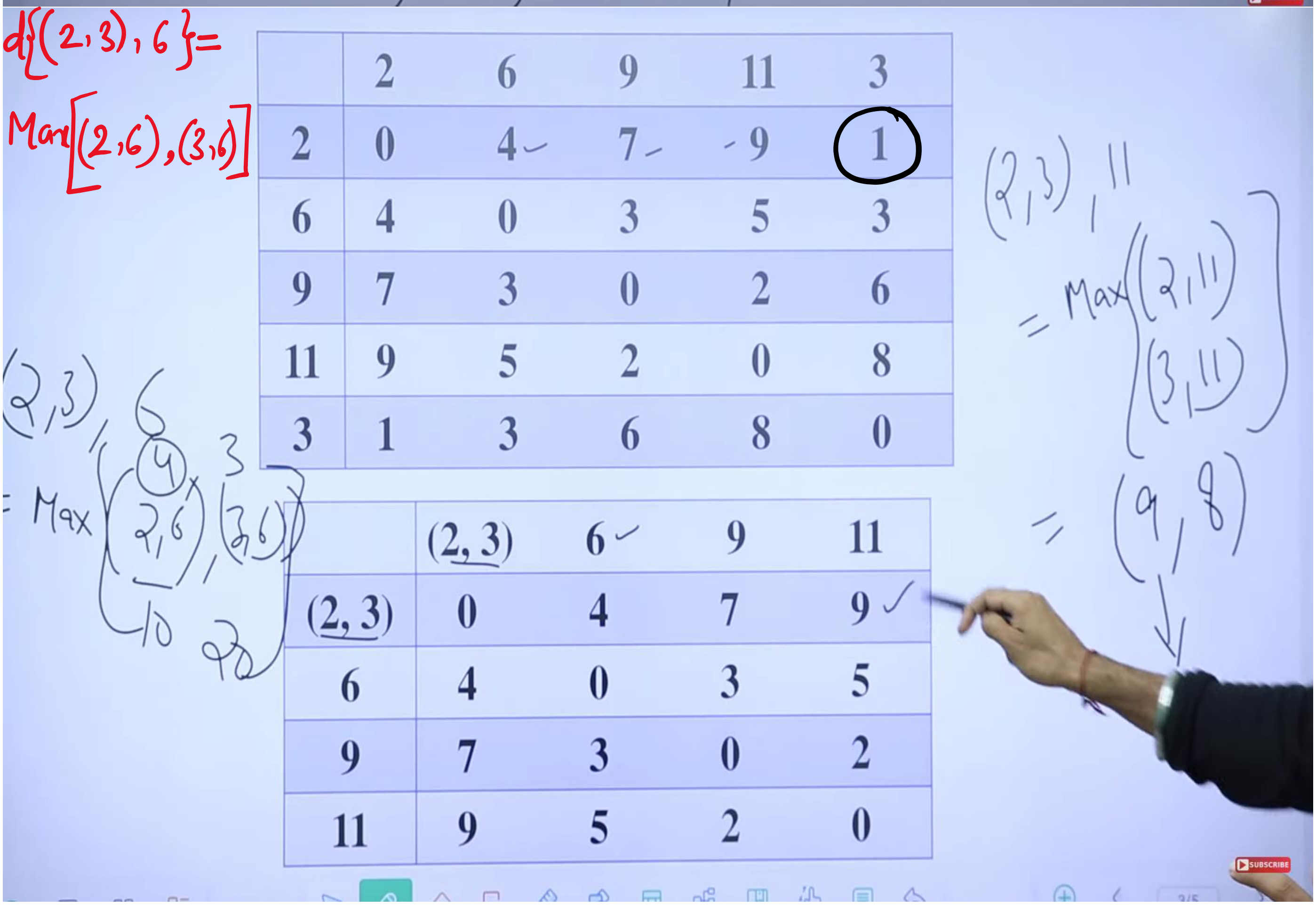
	$(S_1\ S_6)S_3$	S_2	S_4	S_5
$(S_1\ S_6)S_3$	0	3.16	5.09	6.32
S_2	--	0	4.47	7.07
S_4	--	--	0	3.16
S_5	--	--	--	0

	$((S_1\ S_6)S_3)S_2$	S_4S_5
$((S_1\ S_6)S_3)S_2$	0	5.09
S_4S_5	--	0


Complete Linkage Clustering with Example | Clustering in Unsupervised Learning | ML

	2	6	9	11	3
2	0	4	7	9	1
6	4	0	3	5	3
9	7	3	0	2	6
11	9	5	2	0	8
3	1	3	6	8	0

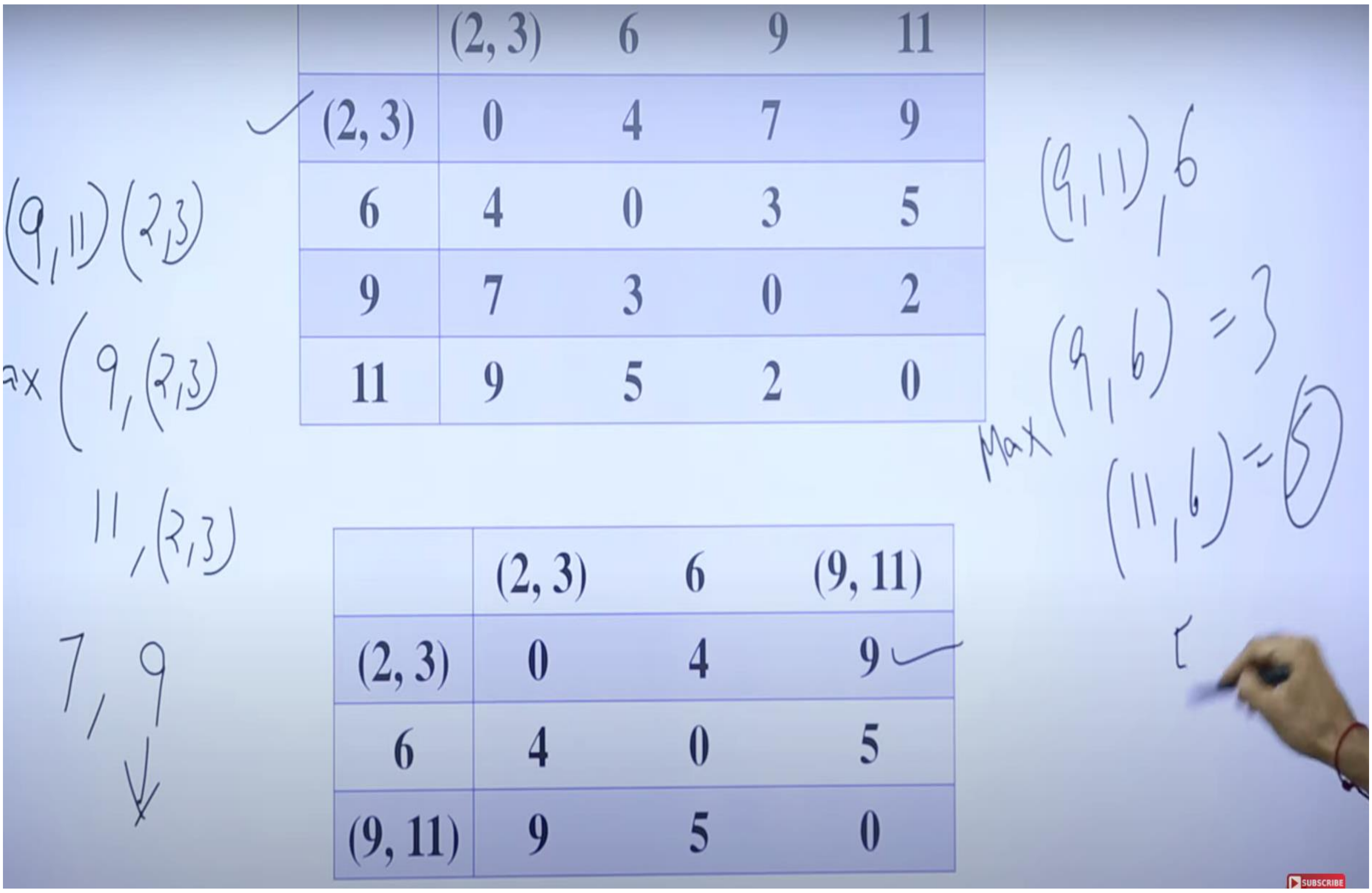
minimum distance
$\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$
$\sqrt{(x_1-x_2)^2} = \sqrt{(2-6)^2} = \sqrt{(-4)^2} = \sqrt{16} = 4$
$\sqrt{(2-9)^2} = \sqrt{(-7)^2} = \sqrt{49} = 7$


$d\{(2,3),6\} = Max[(2,6),(3,6)]$

	2	6	9	11	3
2	0	4	7	9	1
6	4	0	3	5	3
9	7	3	0	2	6
11	9	5	2	0	8
3	1	3	6	8	0

$(2,3), 6 = Max((2,6),(3,6))$
$(2,3), 11 = Max((2,11),(3,11)) = (9, 8)$

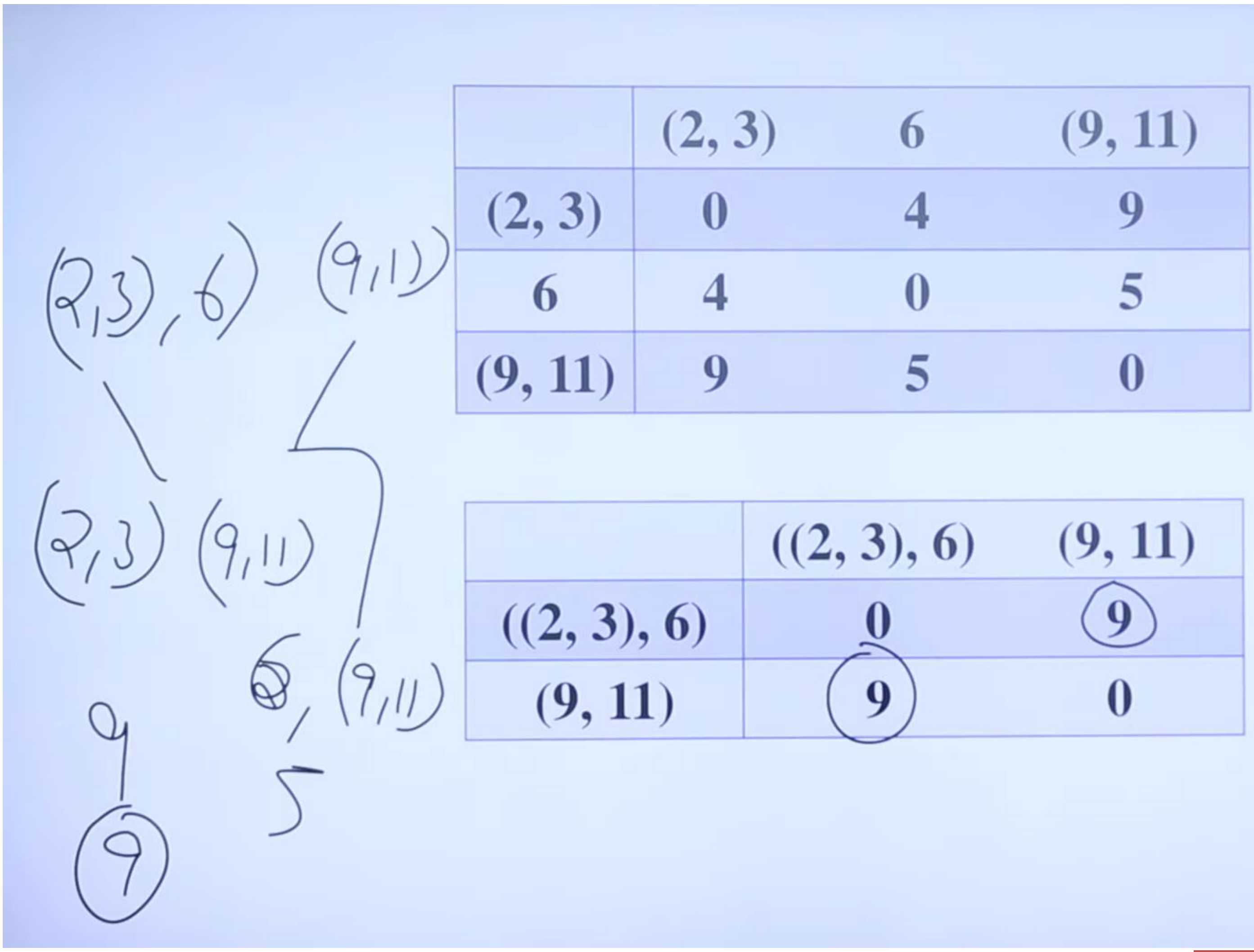
	(2, 3)	6	9	11
(2, 3)	0	4	7	9
6	4	0	3	5
9	7	3	0	2
11	9	5	2	0



	(2, 3)	6	9	11
(2, 3)	0	4	7	9
6	4	0	3	5
9	7	3	0	2
11	9	5	2	0

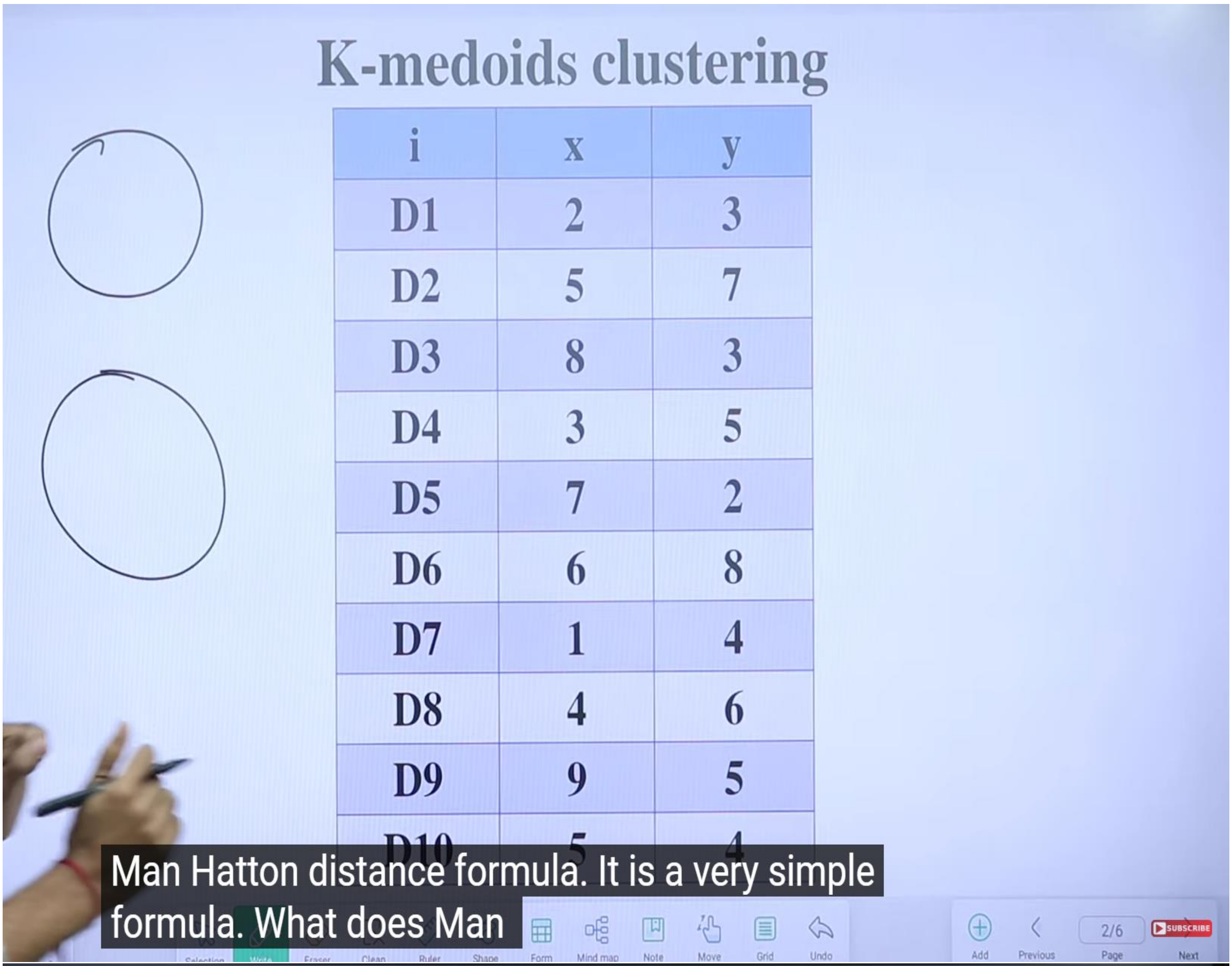
	(2, 3)	6	(9, 11)
(2, 3)	0	4	9
6	4	0	5
(9, 11)	9	5	0

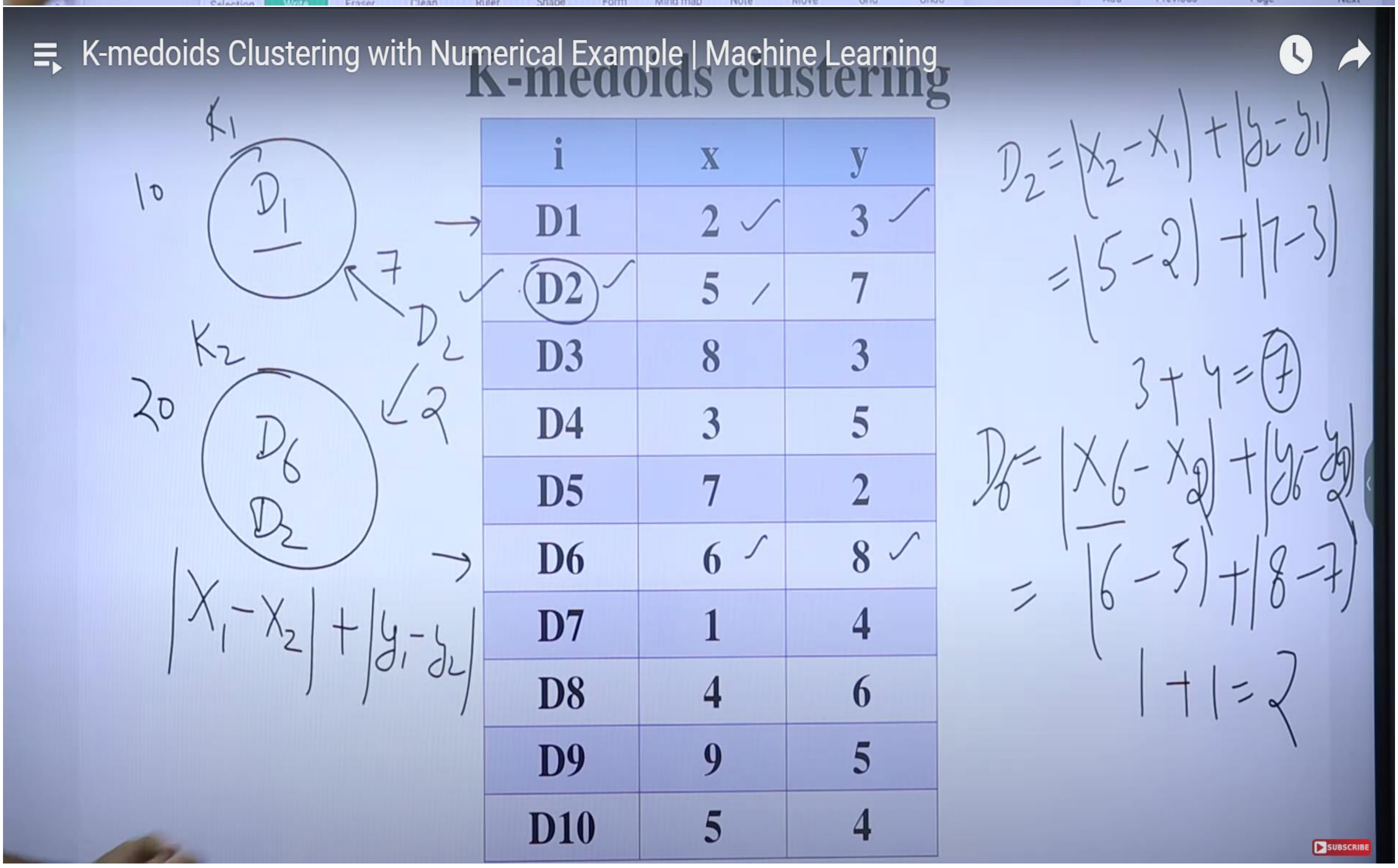
s

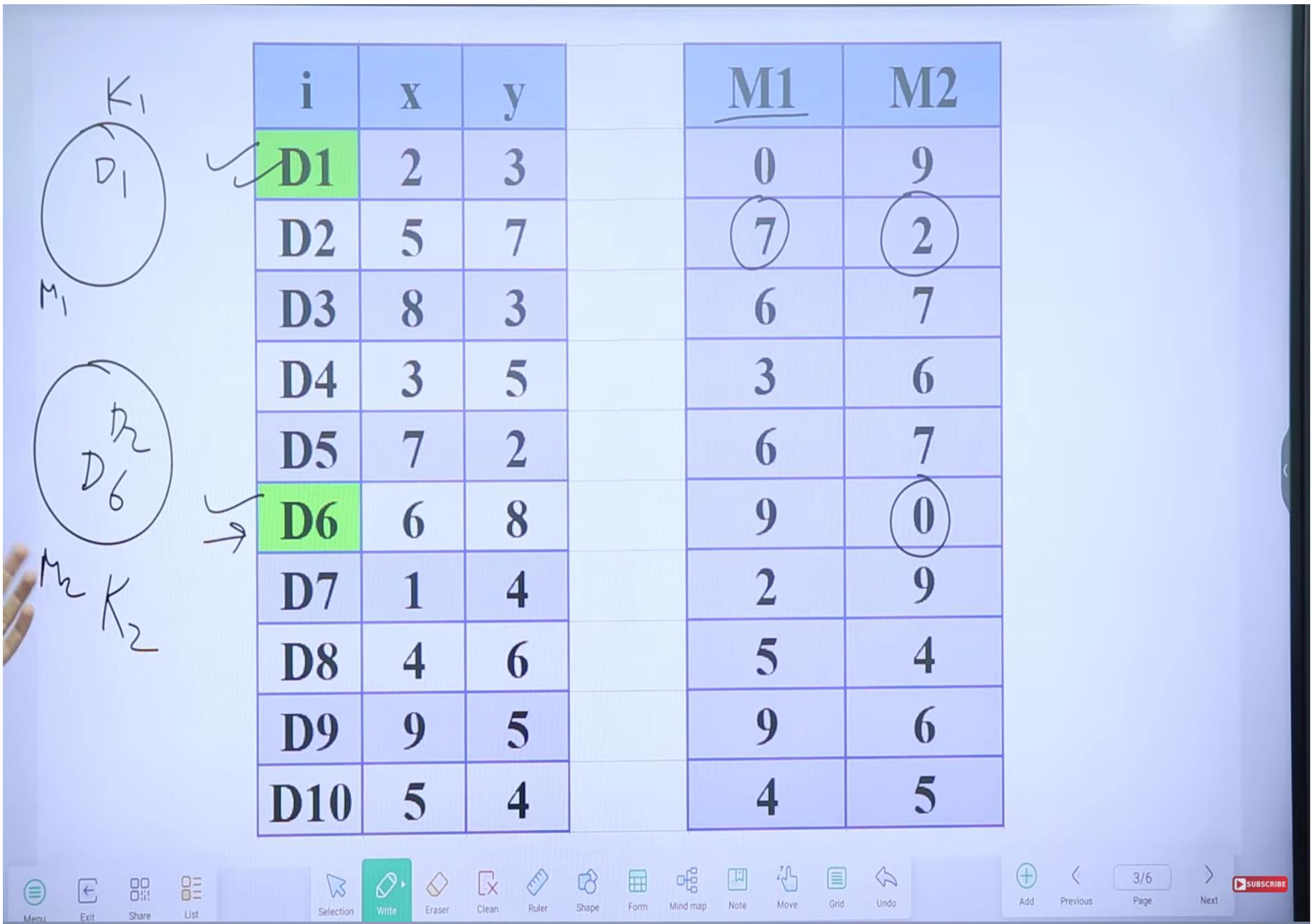


	(2, 3)	6	(9, 11)
(2, 3)	0	4	9
6	4	0	5
(9, 11)	9	5	0

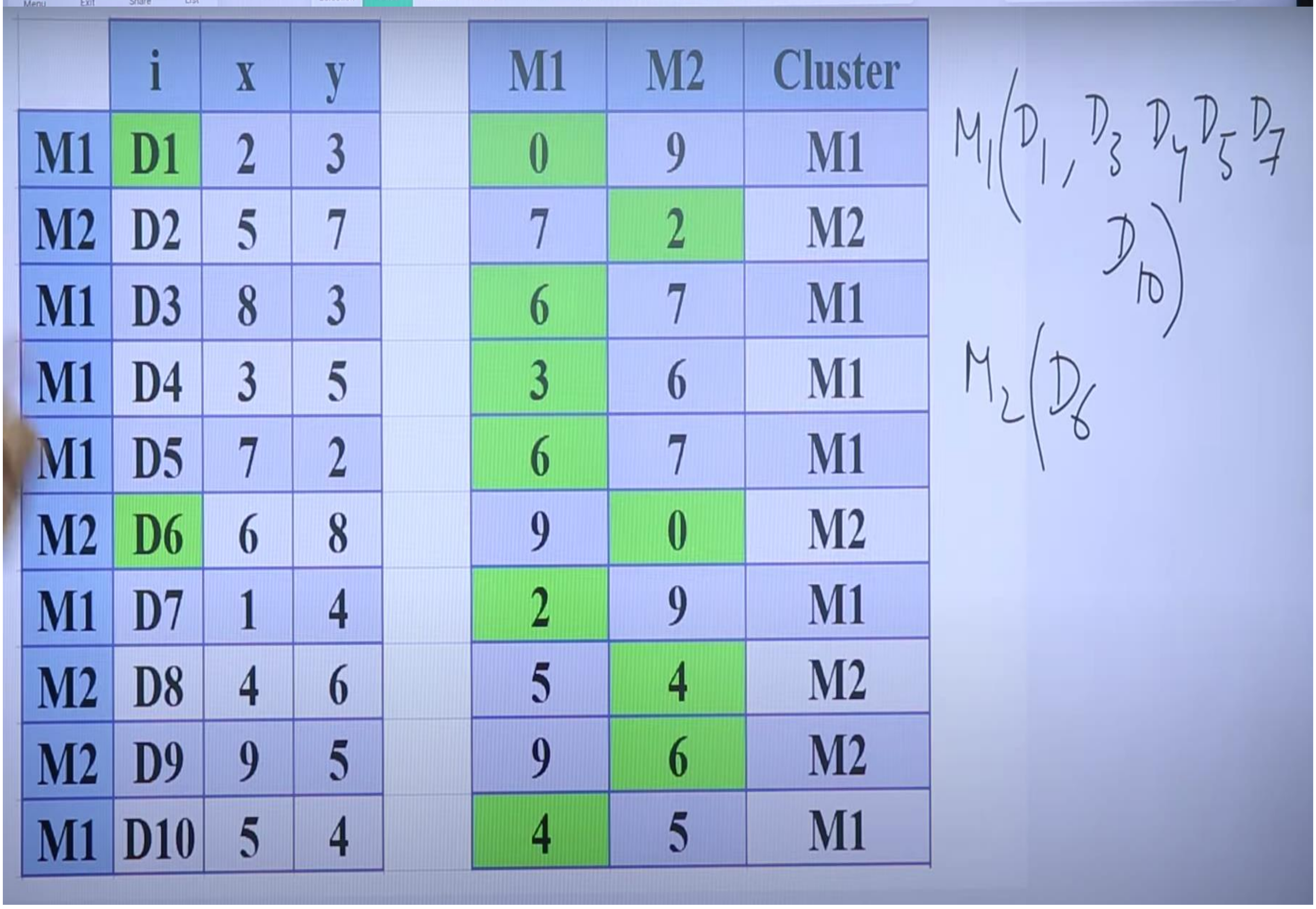
	((2, 3), 6)	(9, 11)
((2, 3), 6)	0	9
(9, 11)	9	0


K-medoids clustering
i x y
D1 2 3
D2 5 7
D3 8 3
D4 3 5
D5 7 2
D6 6 8
D7 1 4
D8 4 6
D9 9 5
D10 5 4
Man Hatton distance formula. It is a very simple
formula. What does Man


K-medoids Clustering with Numerical Example | Machine Learning
K-medoids clustering
i x y
D1 2 3
D2 5 7
D3 8 3
D4 3 5
D5 7 2
D6 6 8
D7 1 4
D8 4 6
D9 9 5
D10 5 4
K1
10
D1
K2
20
D6
D2
7
2
|X1-X2| + |y1-y2|
D2 = |X2-X1| + |y2-y1|
= |5-2| + |7-3|
3 + 4 = 7
D6 = |X6-X2| + |y6-y2|
= |6-5| + |8-7|
1 + 1 = 2



i	x	y	M1	M2
D1	2	3	0	9
D2	5	7	7	2
D3	8	3	6	7
D4	3	5	3	6
D5	7	2	6	7
D6	6	8	9	0
D7	1	4	2	9
D8	4	6	5	4
D9	9	5	9	6
D10	5	4	4	5



	i	x	y	M1	M2	Cluster
M1	D1	2	3	0	9	M1
M2	D2	5	7	7	2	M2
M1	D3	8	3	6	7	M1
M1	D4	3	5	3	6	M1
M1	D5	7	2	6	7	M1
M2	D6	6	8	9	0	M2
M1	D7	1	4	2	9	M1
M2	D8	4	6	5	4	M2
M2	D9	9	5	9	6	M2
M1	D10	5	4	4	5	M1

	i	x	y	M1	M2	Cluster	Total Cost
M1	D1	2	3	0	9	M1	0
M2	D2	5	7	7	2	M2	2
M1	D3	8	3	6	7	M1	6
M1	D4	3	5	3	6	M1	3
M1	D5	7	2	6	7	M1	6
M2	D6	6	8	9	0	M2	0
M1	D7	1	4	2	9	M1	2
M2	D8	4	6	5	4	M2	4
M2	D9	9	5	9	6	M2	6
M1	D10	5	4	4	5	M1	4
							33

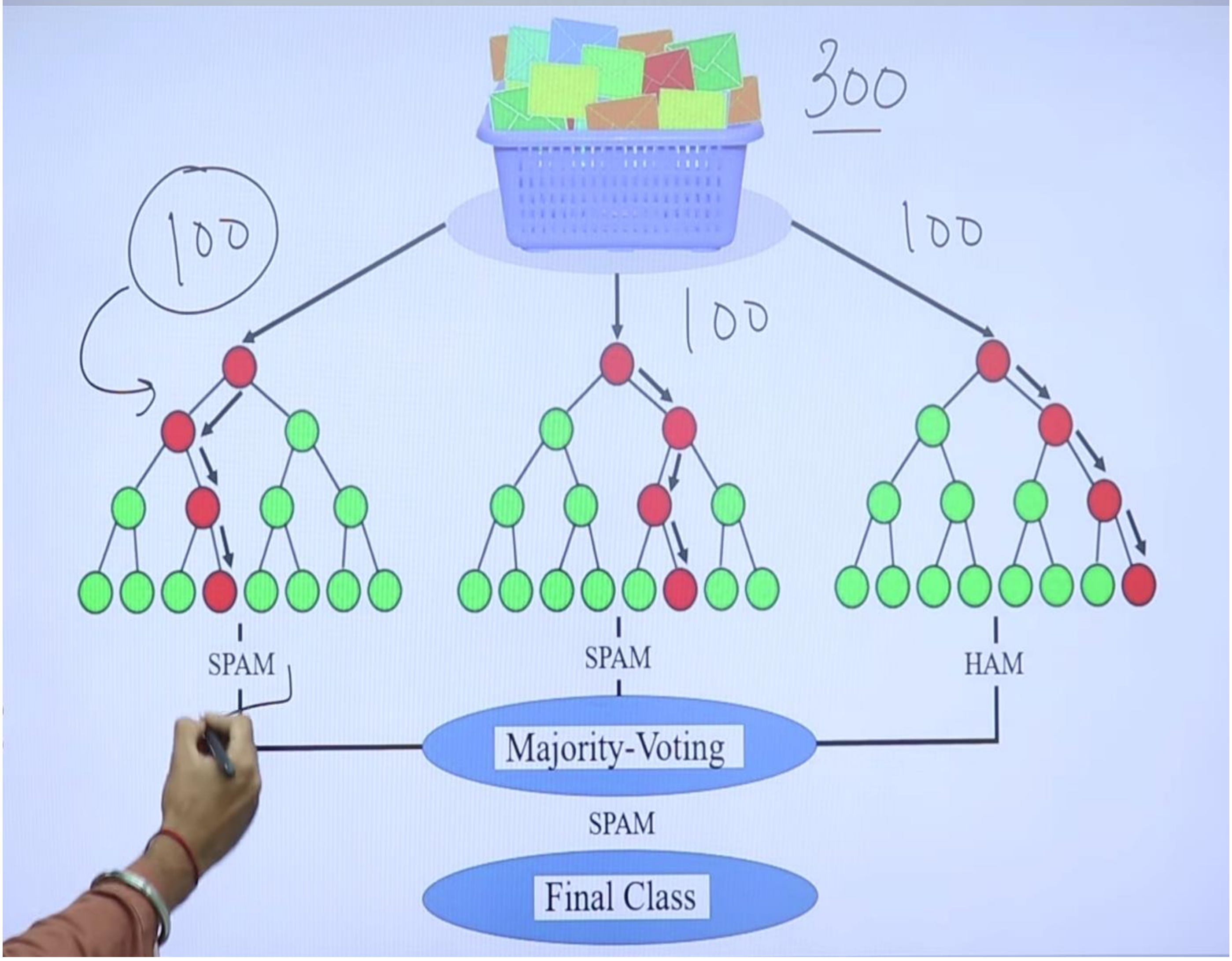
	i	x	y	M1	K	Cluster	Total Cost
M1	D1	2	3	0	3	M1	0
K	D2	5	7	7	4	K	4
M1	D3	8	3	6	7	M1	6
K	D4	3	5	3	0	K	0
M1	D5	7	2	6	7	M1	6
K	D6	6	8	9	6	K	6
M1	D7	1	4	2	3	M1	2
K	D8	4	6	5	2	K	2
K	D9	9	5	9	6	K	6
K	D10	5	4	4	3	K	3
							35

2+2=4

|3-5|+
|5-7|

Random Forest

- Random forests or random decision forests is an ensemble learning method for classification, regression.
- Random decision forests correct for decision trees' habit of overfitting to their training set.
- Step1: Create Bootstrap Dataset from Original data by Randomly choosing data(repetition is allowed).
- Step2: Create Randomized Decision Tree from Bootstrap Dataset.
- Step3: Finally output of the random forest is the class selected by most trees.



Ques. Suppose you have the following dataset with two features (X and Y) and corresponding labels:

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

$$\sqrt{(2-6)^2+(3-7)^2}$$

$$=\sqrt{16+16}$$

$$\sqrt{32}$$

Data Point	X	Y	Label
1	x_2 2.0	y_2 3.0	A
2	3.0 x_2	4.0 y_2	A
3	5.0	6.0	B
4	7.0	8.0	B
5	10.0	10.0	A

$$\sqrt{(3-6)^2+(4-7)^2}=\sqrt{}$$

Now, consider a new data point with *x*1=6.0 and *y*1=7.0. Using kNN with *k*=3, predict the label for this new data point.

Ques. Consider a dataset with a single feature (X) and corresponding target values (Y):

Data Point	X (Input)	Y (Target)
1	2.0	5.0
2	4.0	8.0
3	6.0	12.0
4	8.0	15.0
5	10.0	20.0

$$\sqrt{(2-7)^2}$$

$$\sqrt{(4-7)^2}$$

$$\sqrt{(6-7)^2}$$

$$\sqrt{(8-7)^2}$$

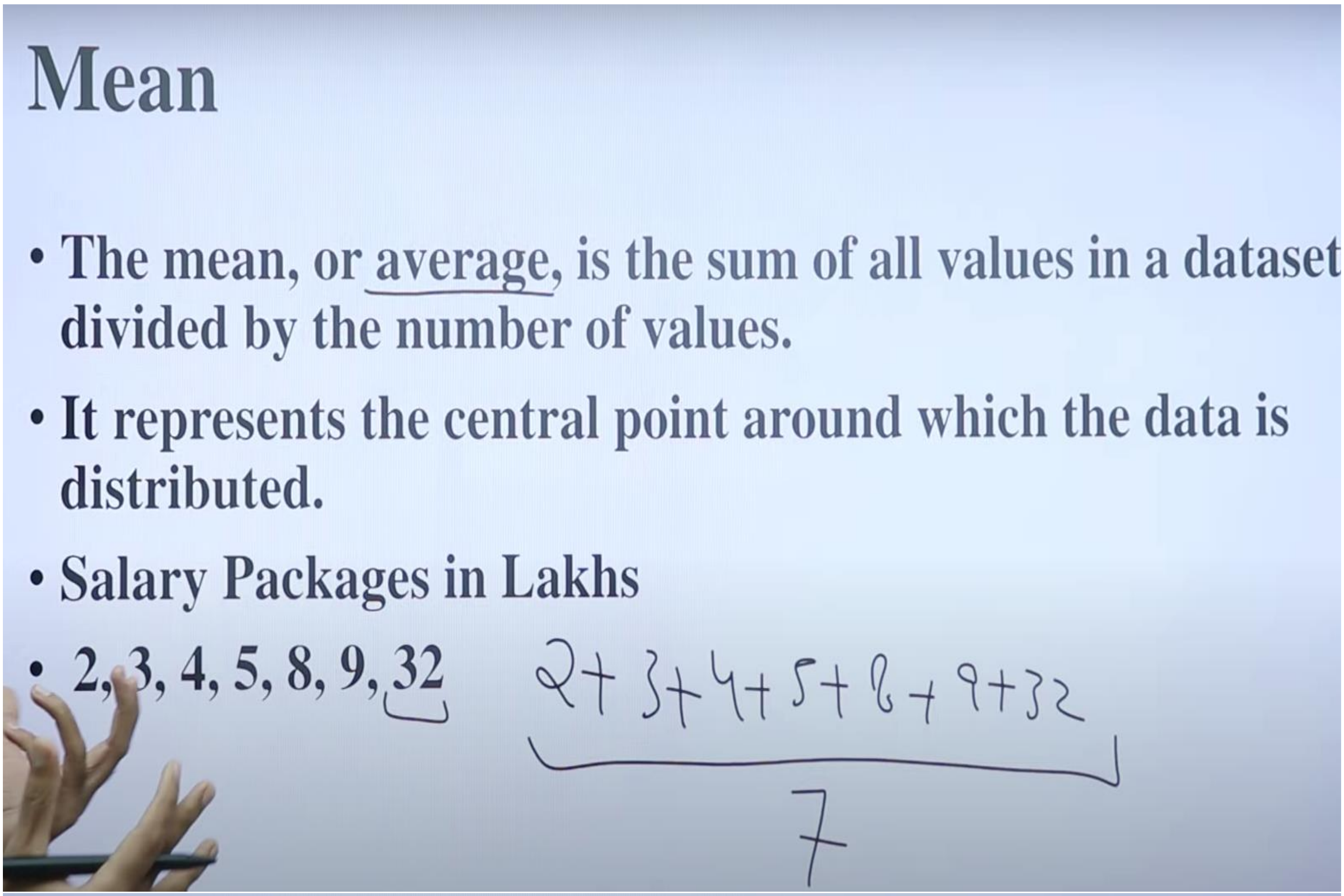
$$\sqrt{(10-7)^2}$$

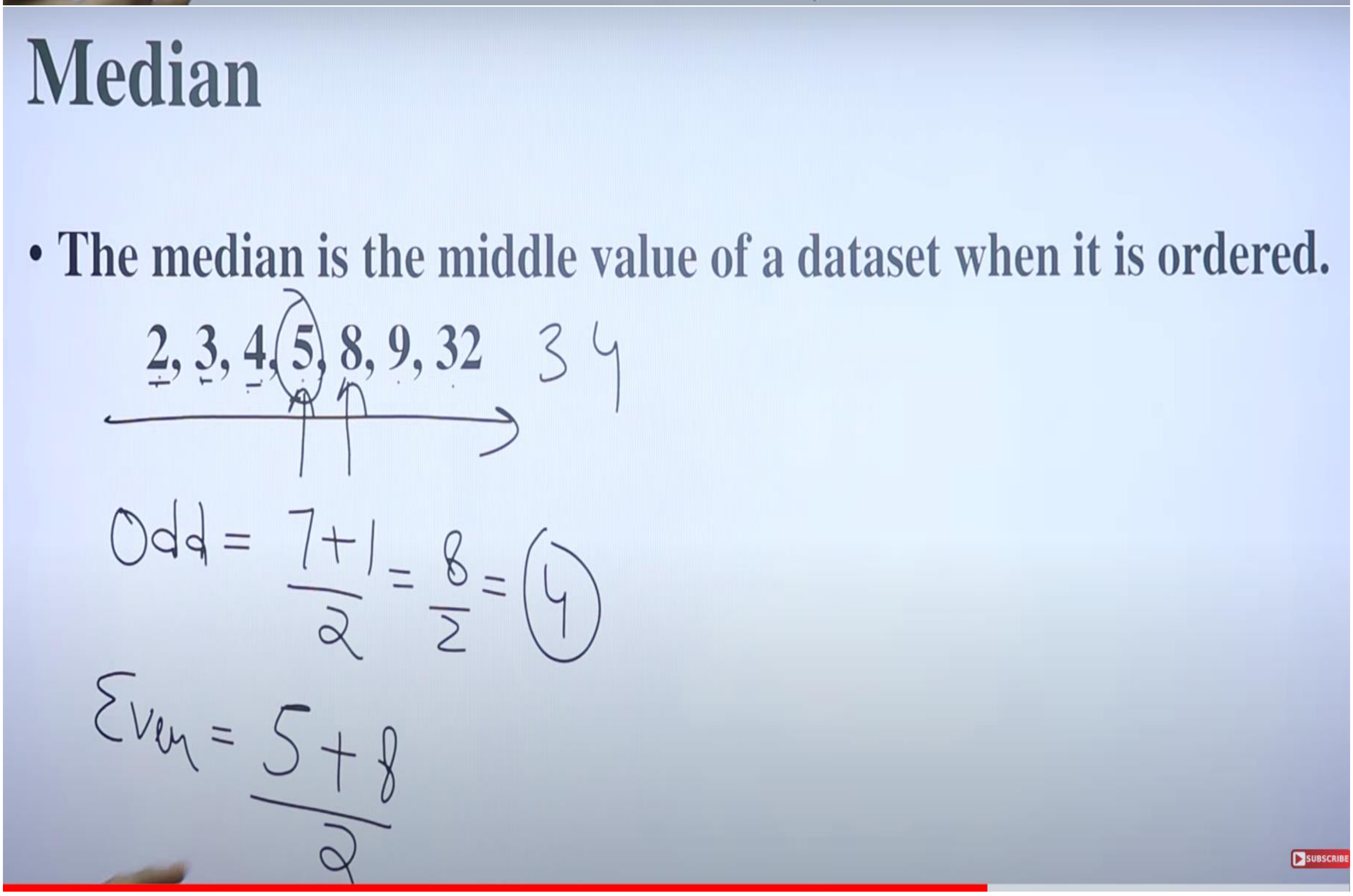
$$\frac{12+15}{2}=\frac{27}{2}$$

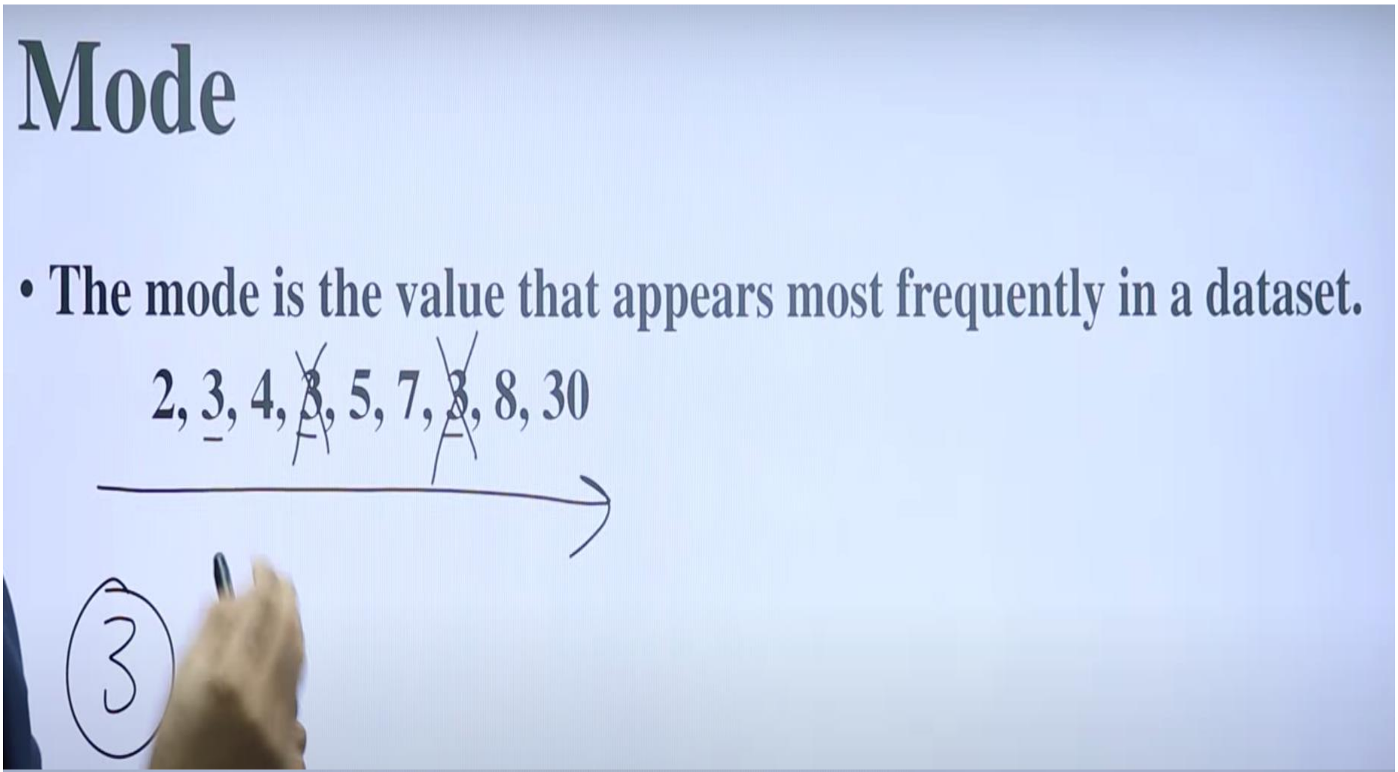
13.5

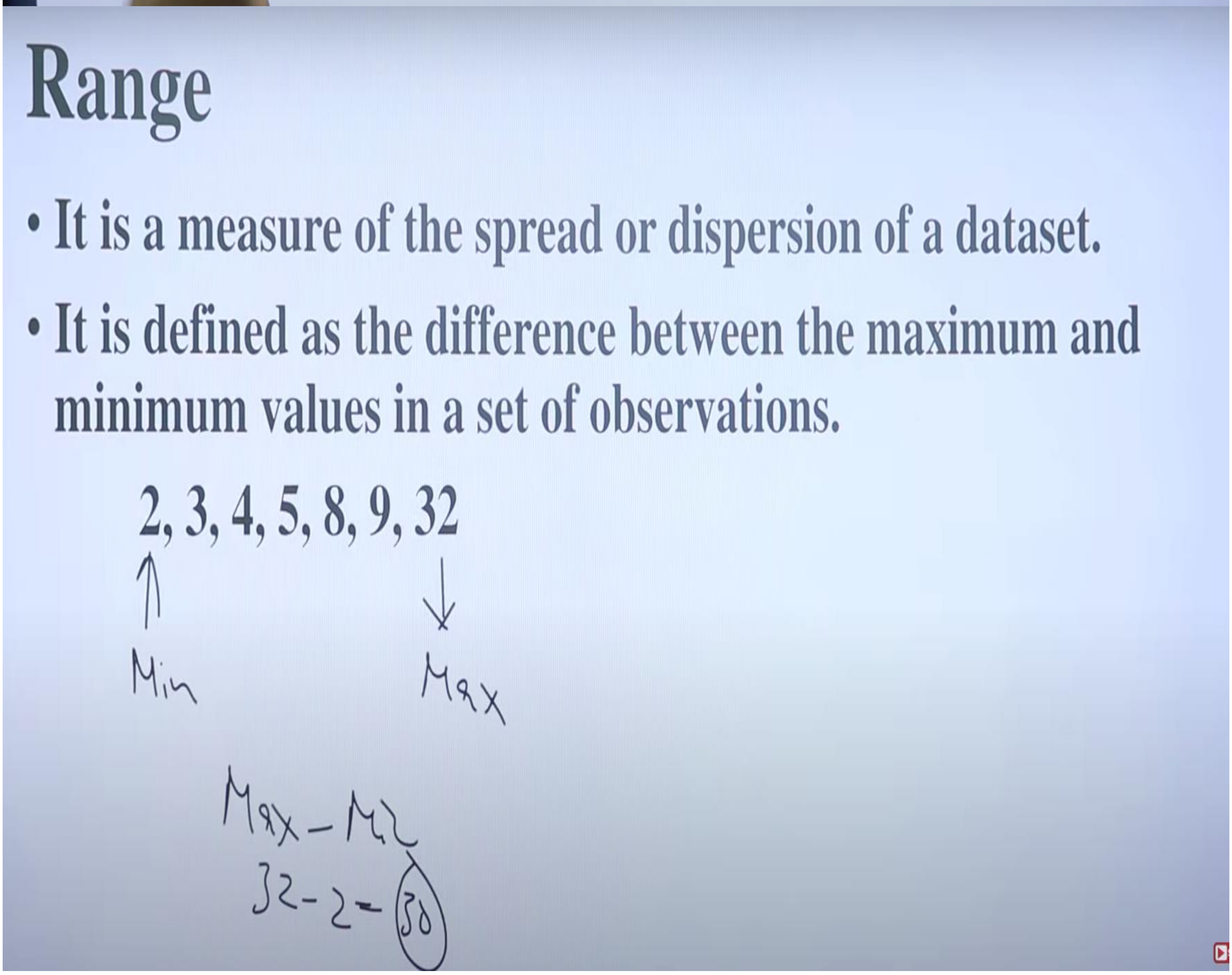
Now, consider a new data point with X1=7.0. Using kNN with k=2, predict the target value (Y) for this new data point.

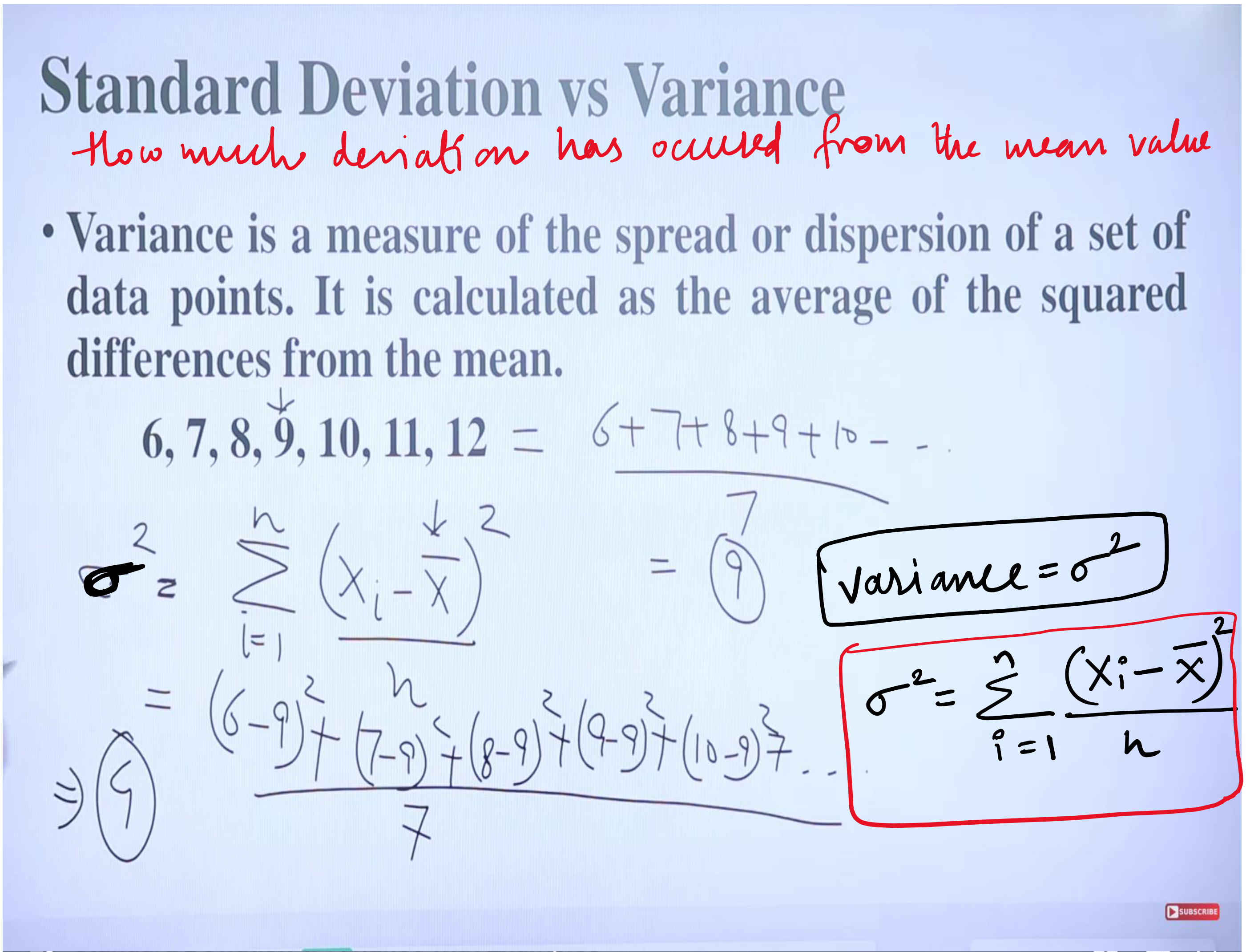
13.5

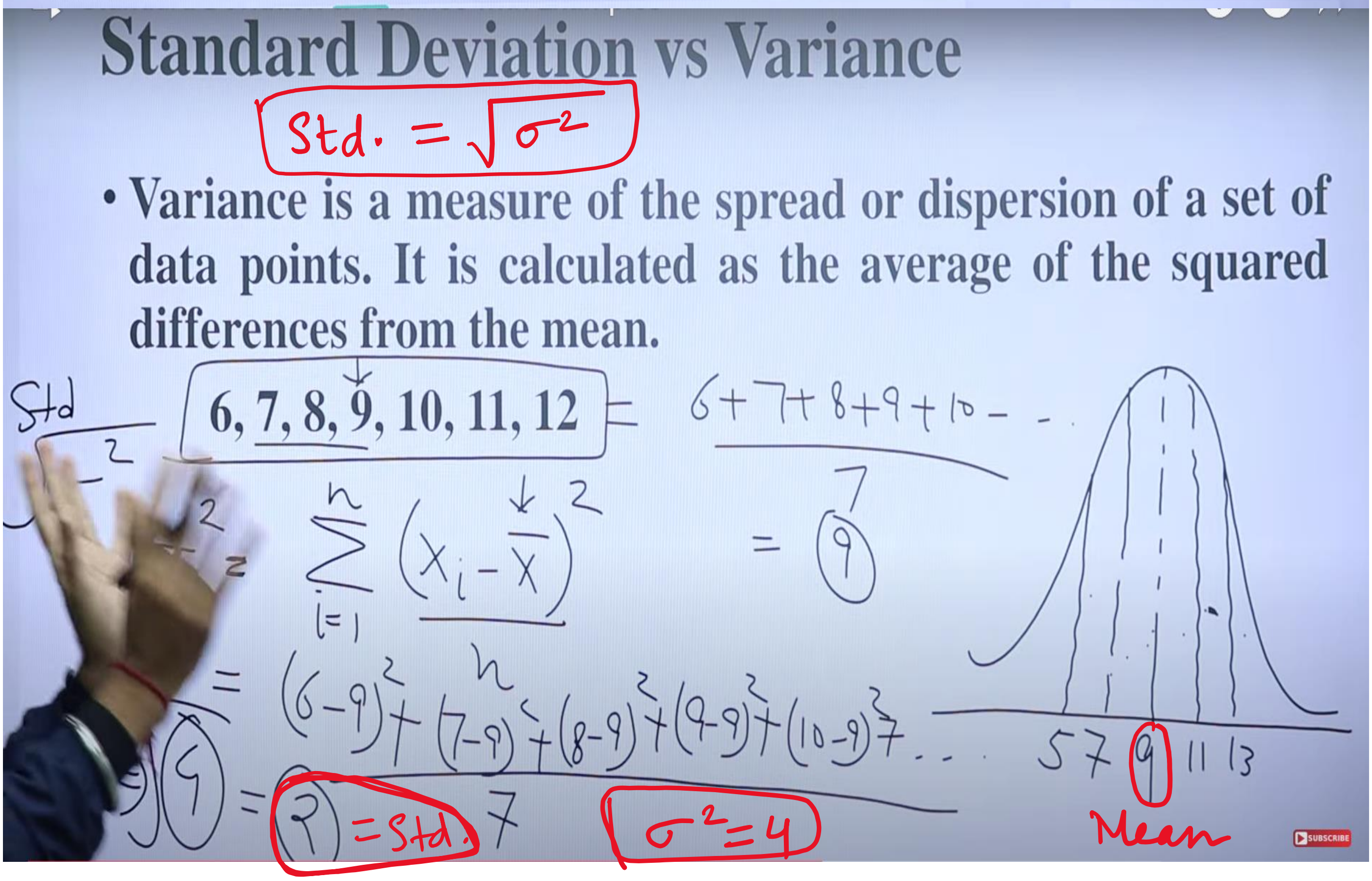

Mean
• The mean, or average, is the sum of all values in a dataset divided by the number of values.
• It represents the central point around which the data is distributed.
• Salary Packages in Lakhs
• 2, 3, 4, 5, 8, 9, 32
2+3+4+5+8+9+32
7

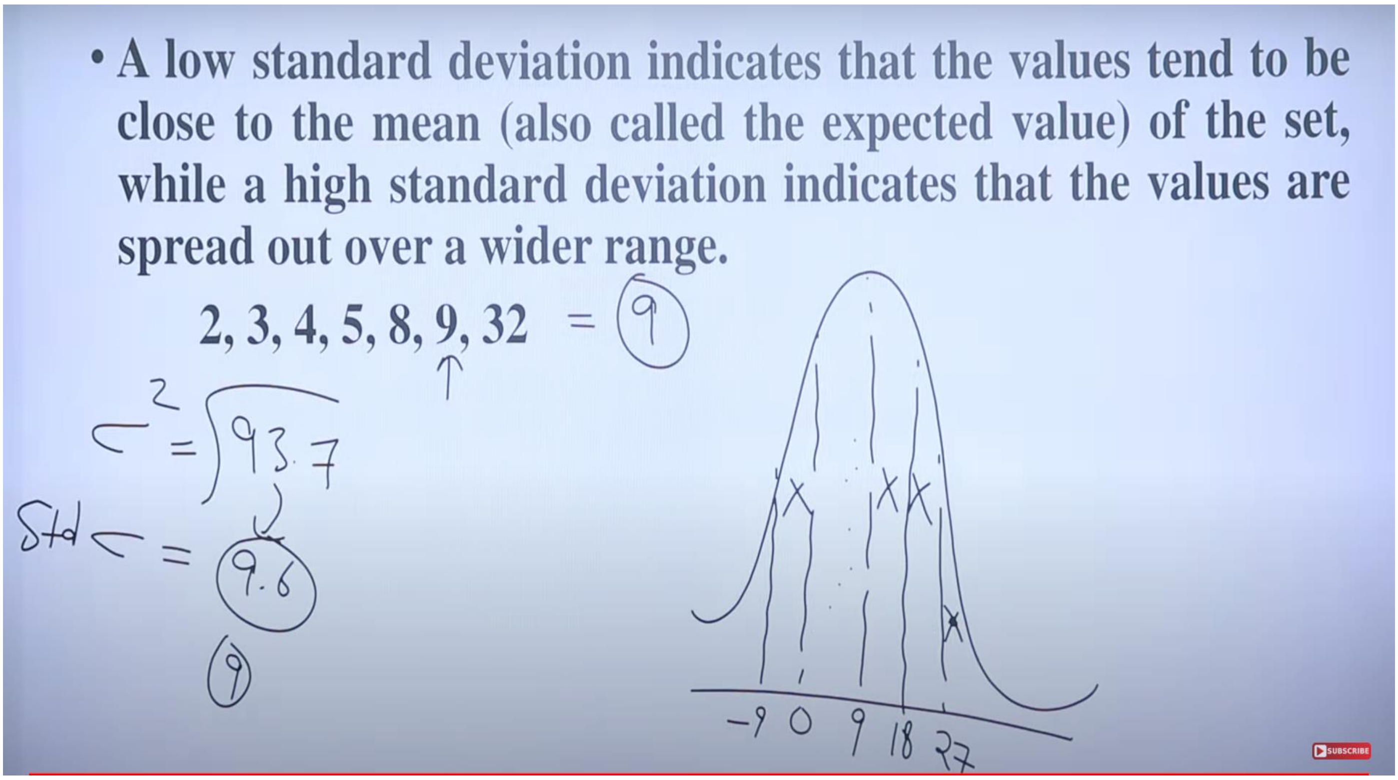

Median
• The median is the middle value of a dataset when it is ordered.
2, 3, 4, 5, 8, 9, 32
34
Odd = 7+1 / 2 = 8 / 2 = 4
Even = 5+8 / 2

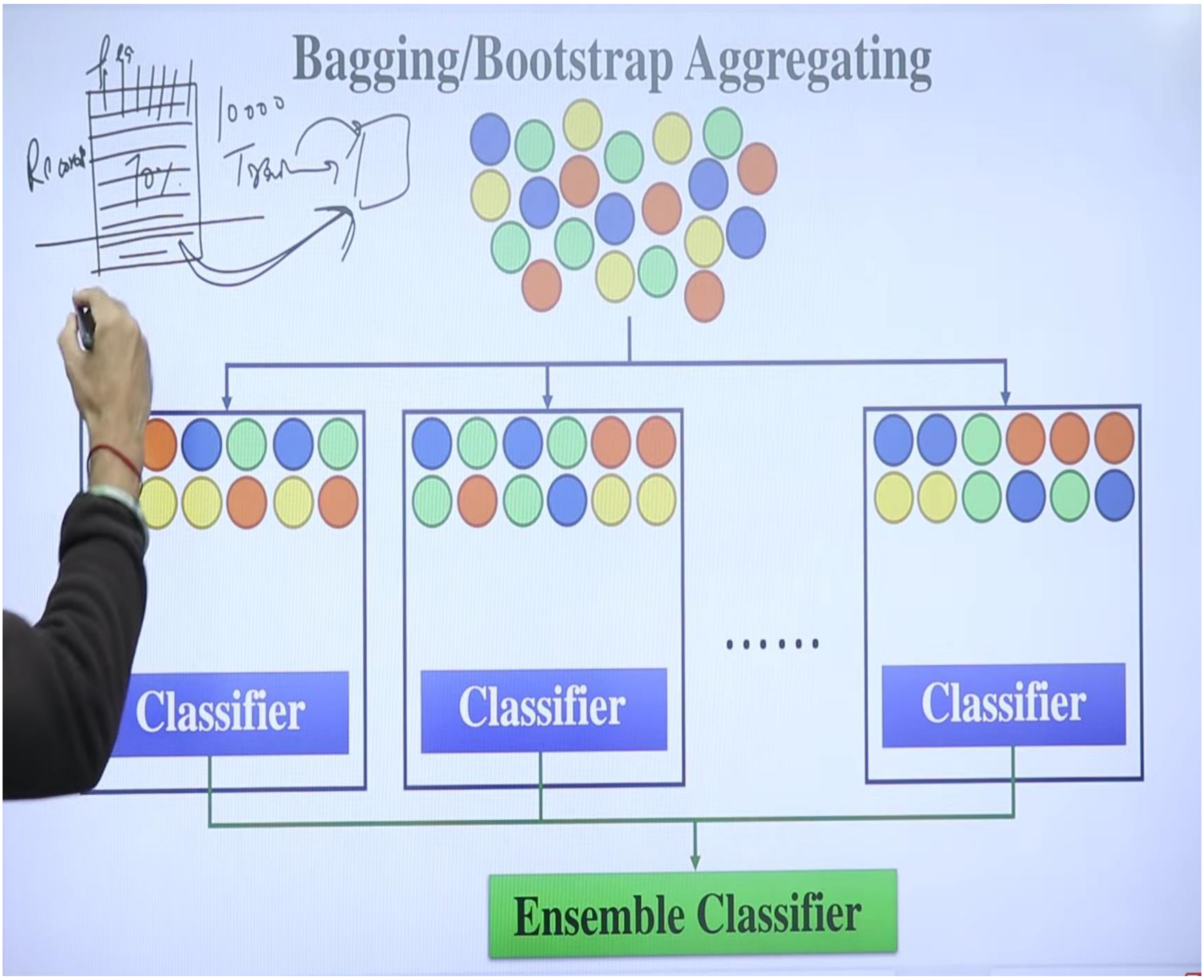

Mode
• The mode is the value that appears most frequently in a dataset.
2, 3, 4, 3, 5, 7, 3, 8, 30
3

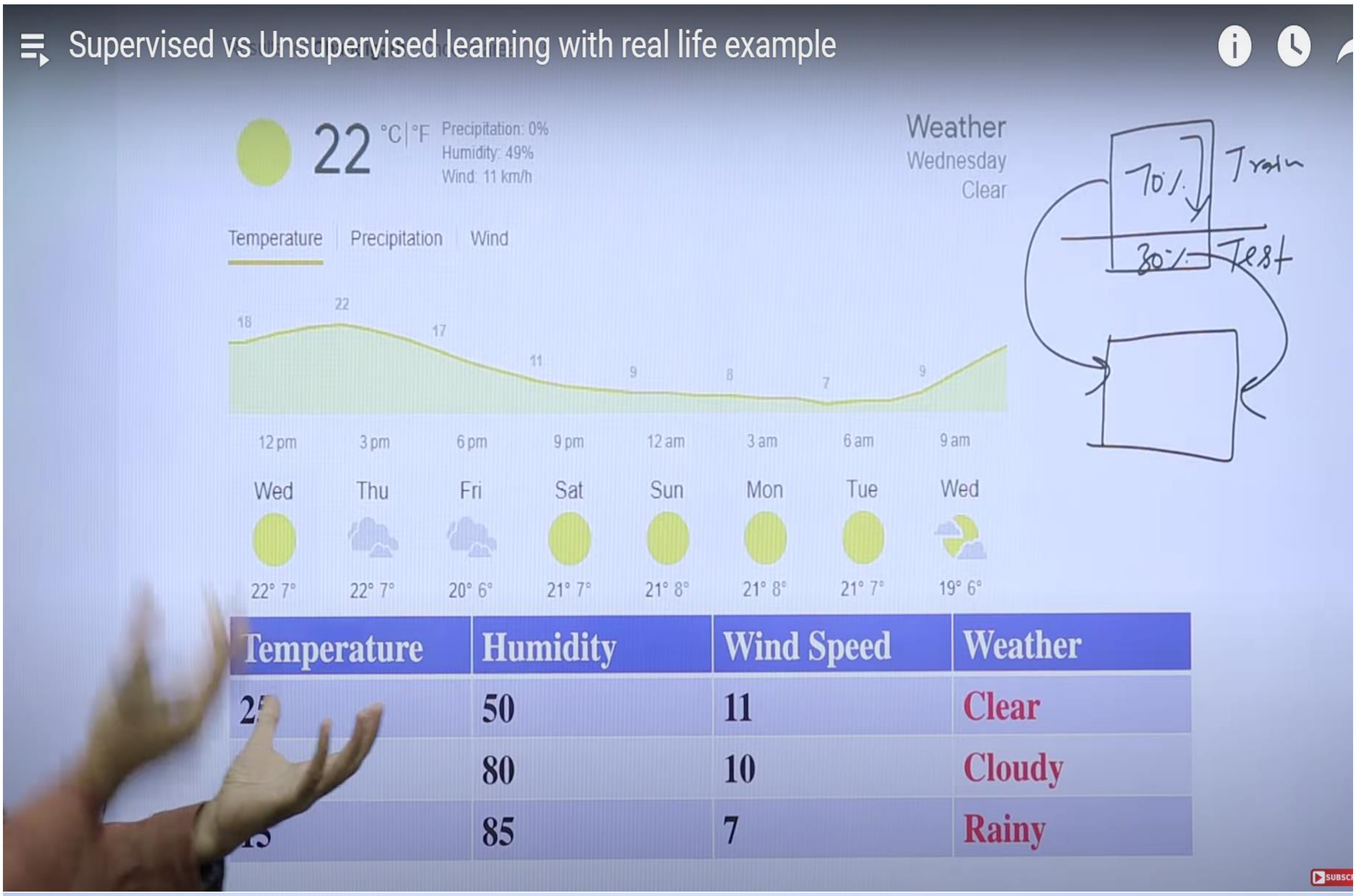

Range
• It is a measure of the spread or dispersion of a dataset.
• It is defined as the difference between the maximum and minimum values in a set of observations.
2, 3, 4, 5, 8, 9, 32
Min
Max
Max – Min
32 – 2 = 30

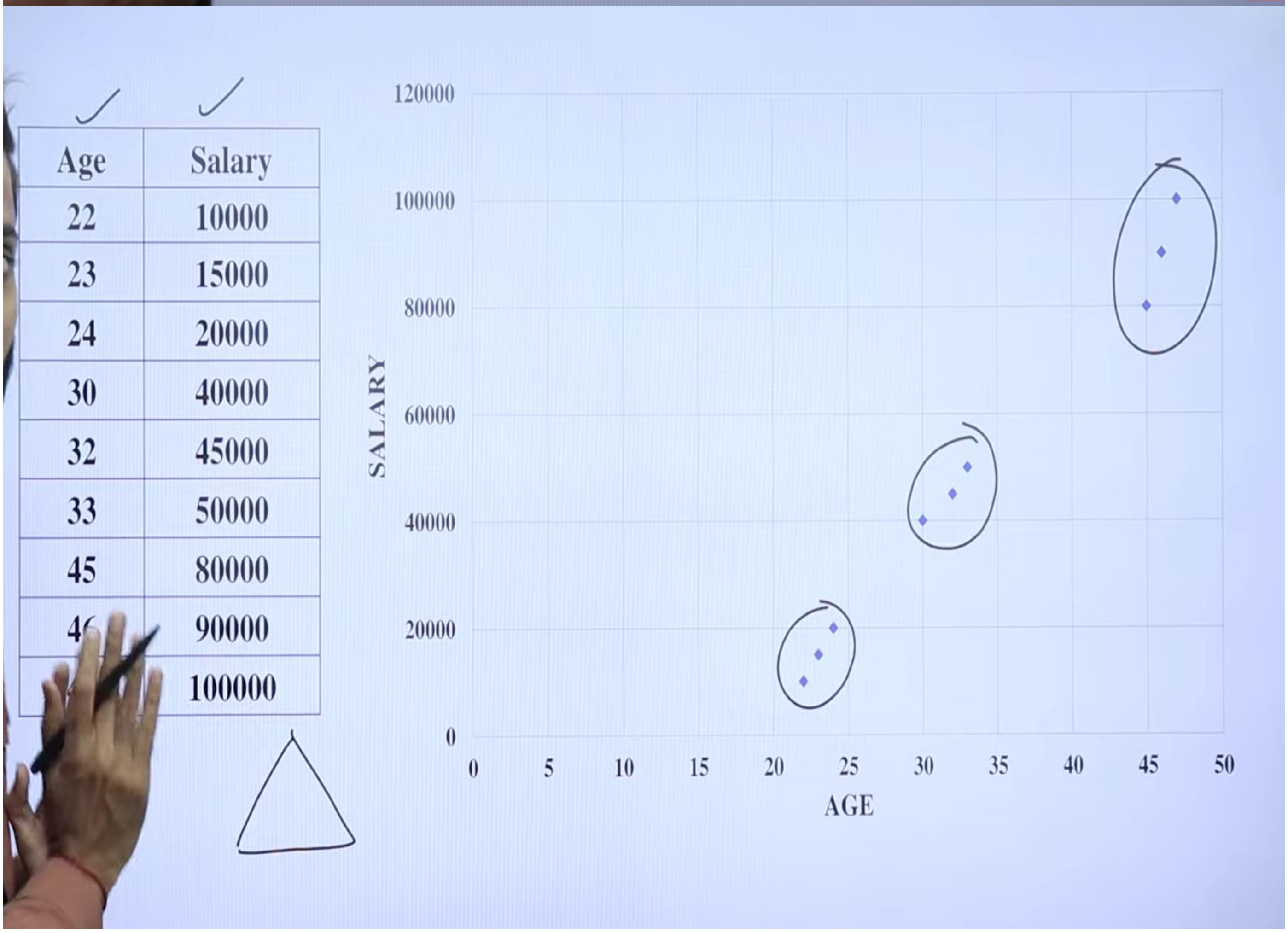

Standard Deviation vs Variance
How much deviation has occured from the mean value
• Variance is a measure of the spread or dispersion of a set of data points. It is calculated as the average of the squared differences from the mean.
6, 7, 8, 9, 10, 11, 12 = (6+7+8+9+10 - -) / 7
= 9
variance = σ²
σ² = Σ_{i=1}^{n} (xᵢ − x̄)² / n
= ((6−9)² + (7−9)² + (8−9)² + (9−9)² + (10−9)² ...) / 7
⇒ 9
SUBSCRIBE

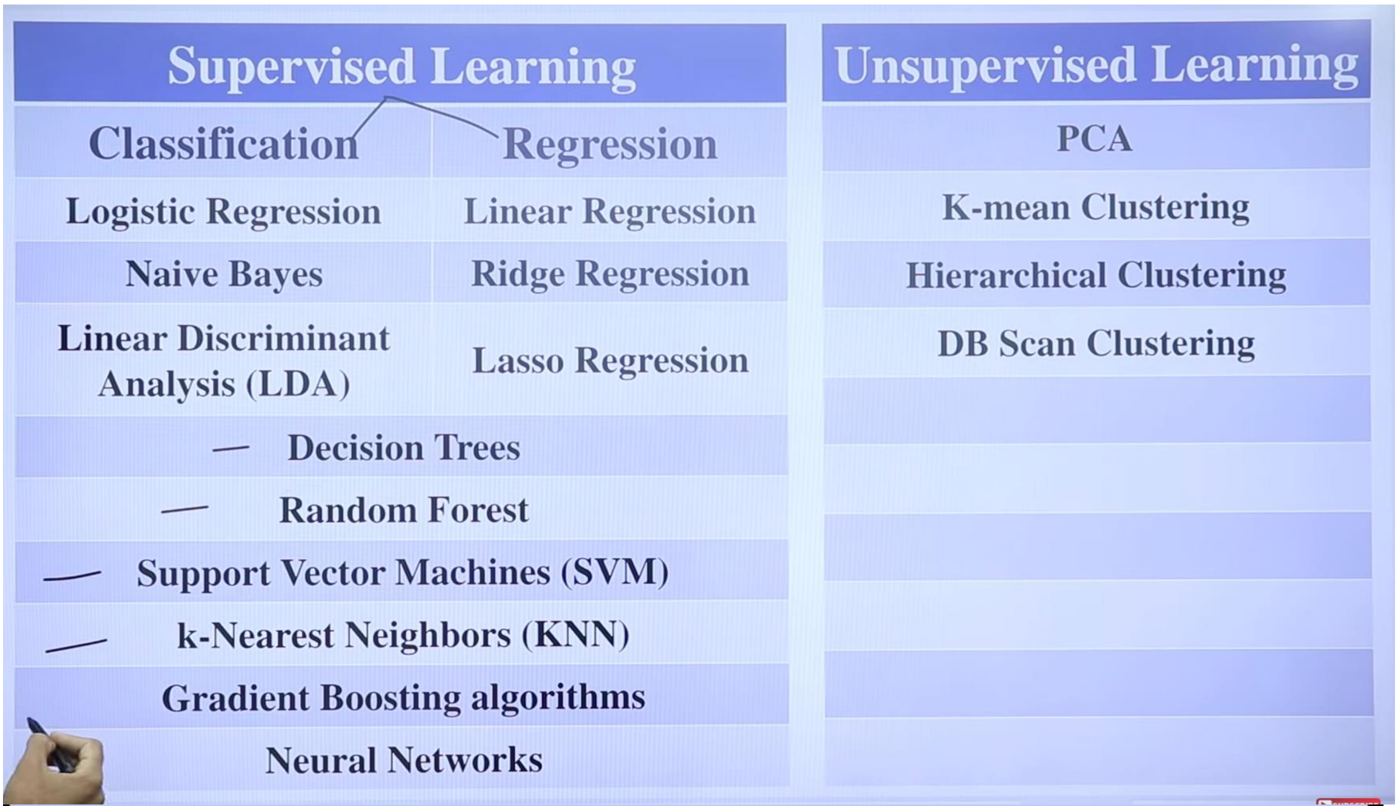

Standard Deviation vs Variance
Std. = √σ²
• Variance is a measure of the spread or dispersion of a set of data points. It is calculated as the average of the squared differences from the mean.
6, 7, 8, 9, 10, 11, 12 = (6+7+8+9+10 - -) / 7
= 9
Std
= ((6−9)² + (7−9)² + (8−9)² + (9−9)² + (10−9)² ...) / 7
? = Std.
σ² = 4
5 7 9 11 13
Mean
SUBSCRIBE

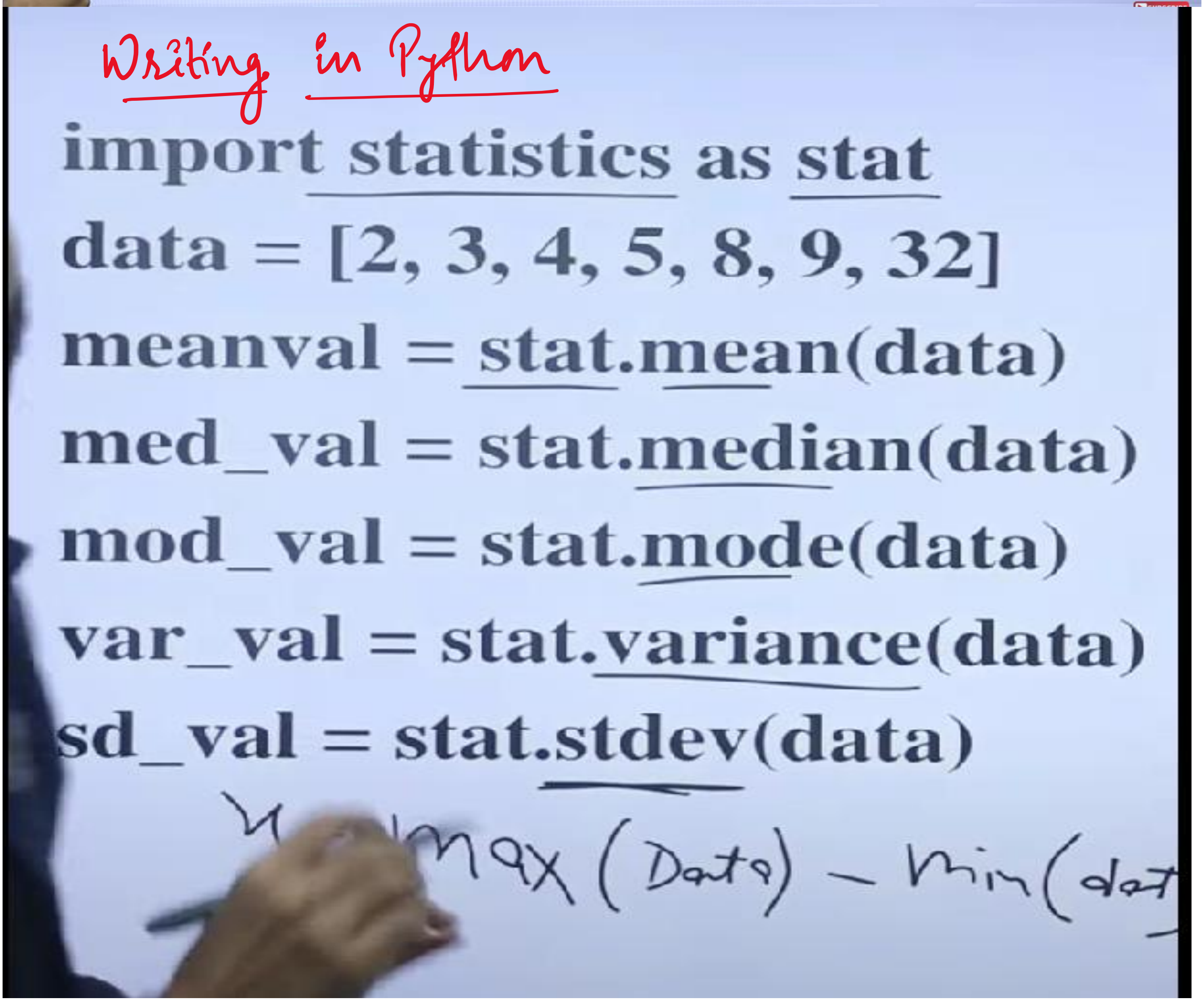

• A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the values are spread out over a wider range.
2, 3, 4, 5, 8, 9, 32 = 9
σ² = 93.7
Std σ = 9.6
9
-9 0 9 18 27

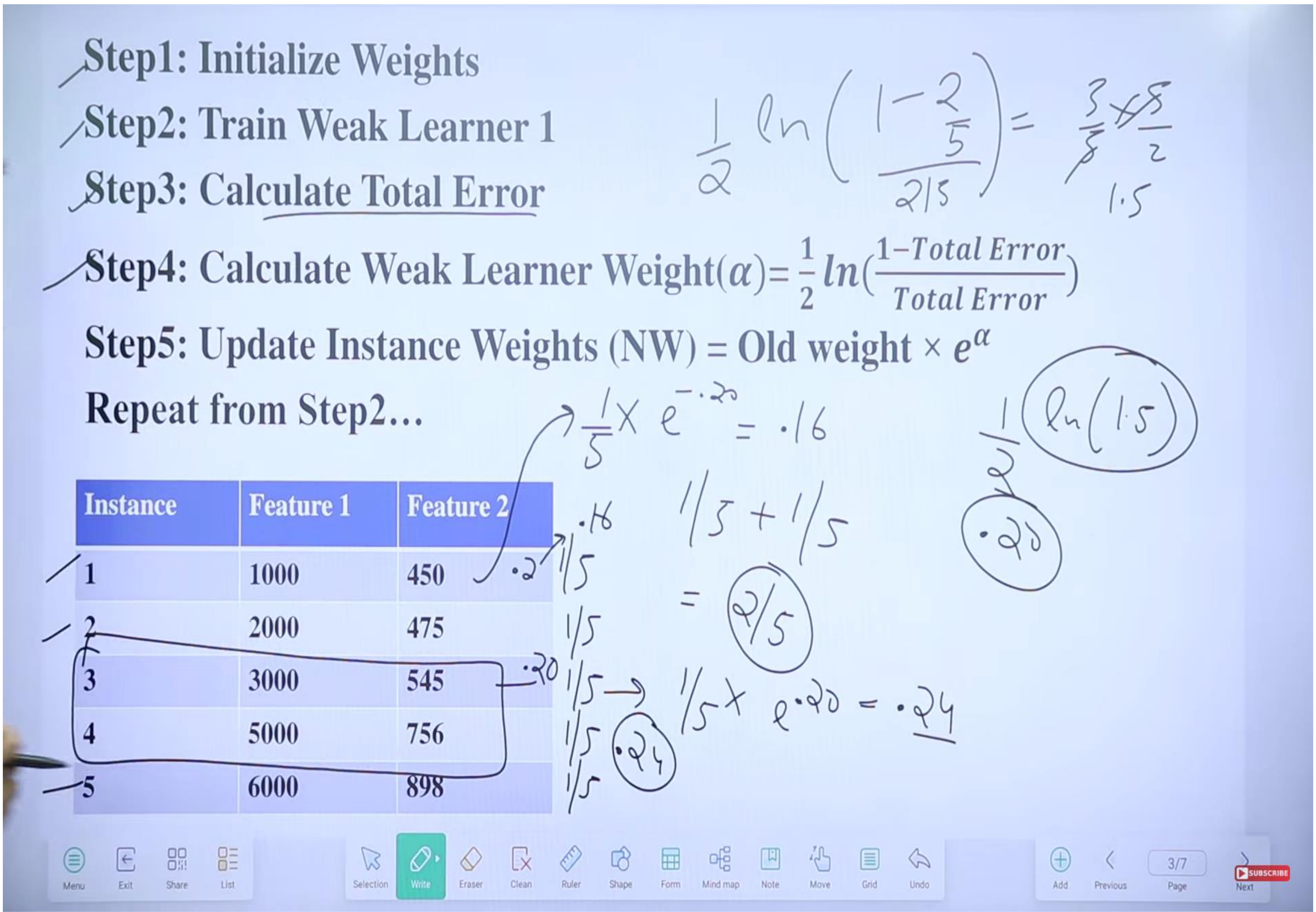

Bagging/Bootstrap Aggregating
10000
70%
Classifier
Classifier
Classifier
Ensemble Classifier

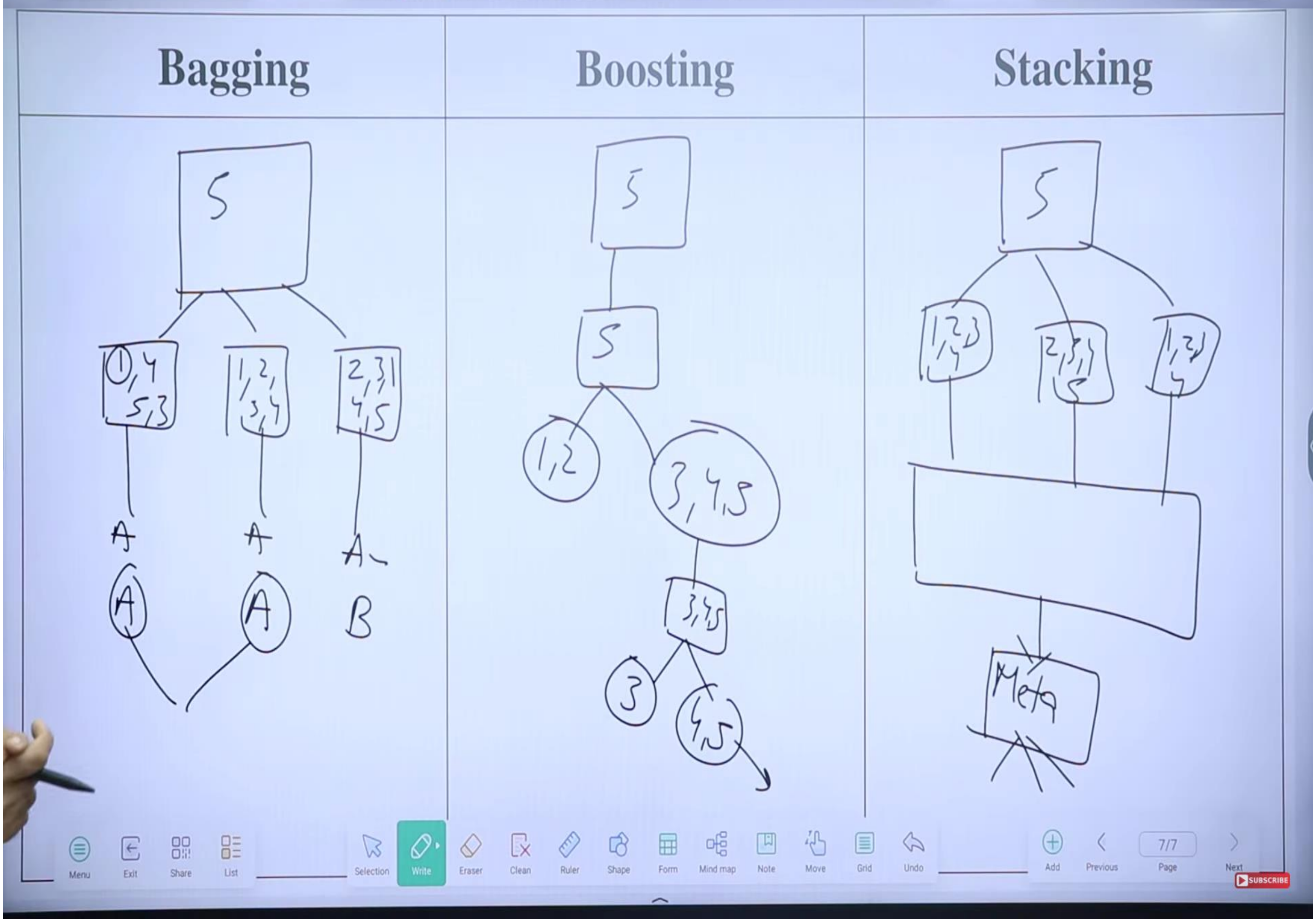

Supervised vs Unsupervised learning with real life example
22 °C | °F
Precipitation: 0%
Humidity: 49%
Wind: 11 km/h
Weather
Wednesday
Clear
Temperature Precipitation Wind
12 pm 3 pm 6 pm 9 pm 12 am 3 am 6 am 9 am
Wed Thu Fri Sat Sun Mon Tue Wed
22° 7° 22° 7° 20° 6° 21° 7° 21° 8° 21° 8° 21° 7° 19° 6°
70% Train
30% Test
Temperature Humidity Wind Speed Weather
50 11 Clear
80 10 Cloudy
85 7 Rainy

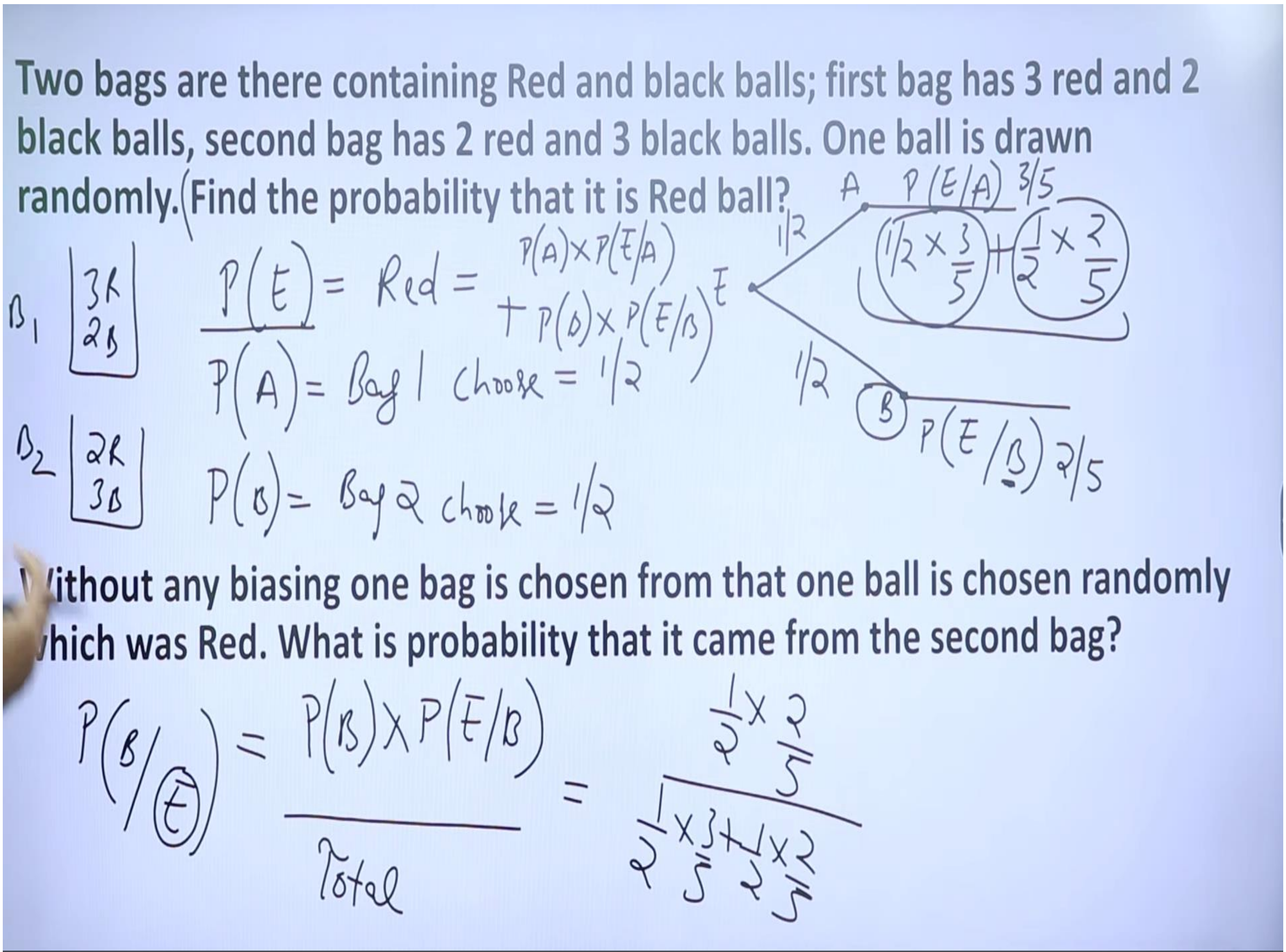

Age Salary
22 10000
23 15000
24 20000
30 40000
32 45000
33 50000
45 80000
90000
100000
SALARY
120000 100000 80000 60000 40000 20000 0
0 5 10 15 20 25 30 35 40 45 50
AGE

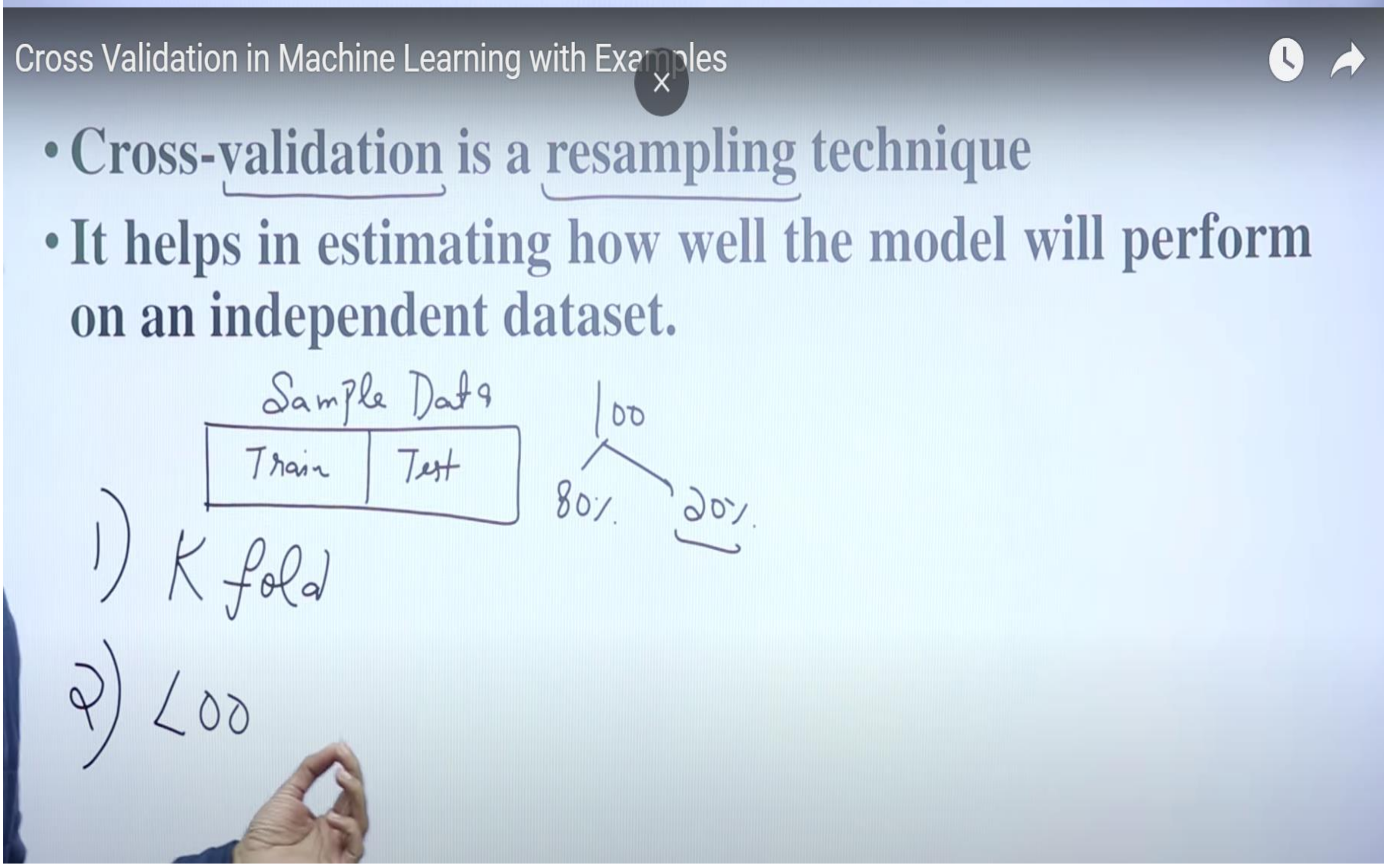

Supervised Learning
Classification
Regression
Logistic Regression
Linear Regression
Naive Bayes
Ridge Regression
Linear Discriminant Analysis (LDA)
Lasso Regression
Decision Trees
Random Forest
Support Vector Machines (SVM)
k-Nearest Neighbors (KNN)
Gradient Boosting algorithms
Neural Networks
Unsupervised Learning
PCA
K-mean Clustering
Hierarchical Clustering
DB Scan Clustering

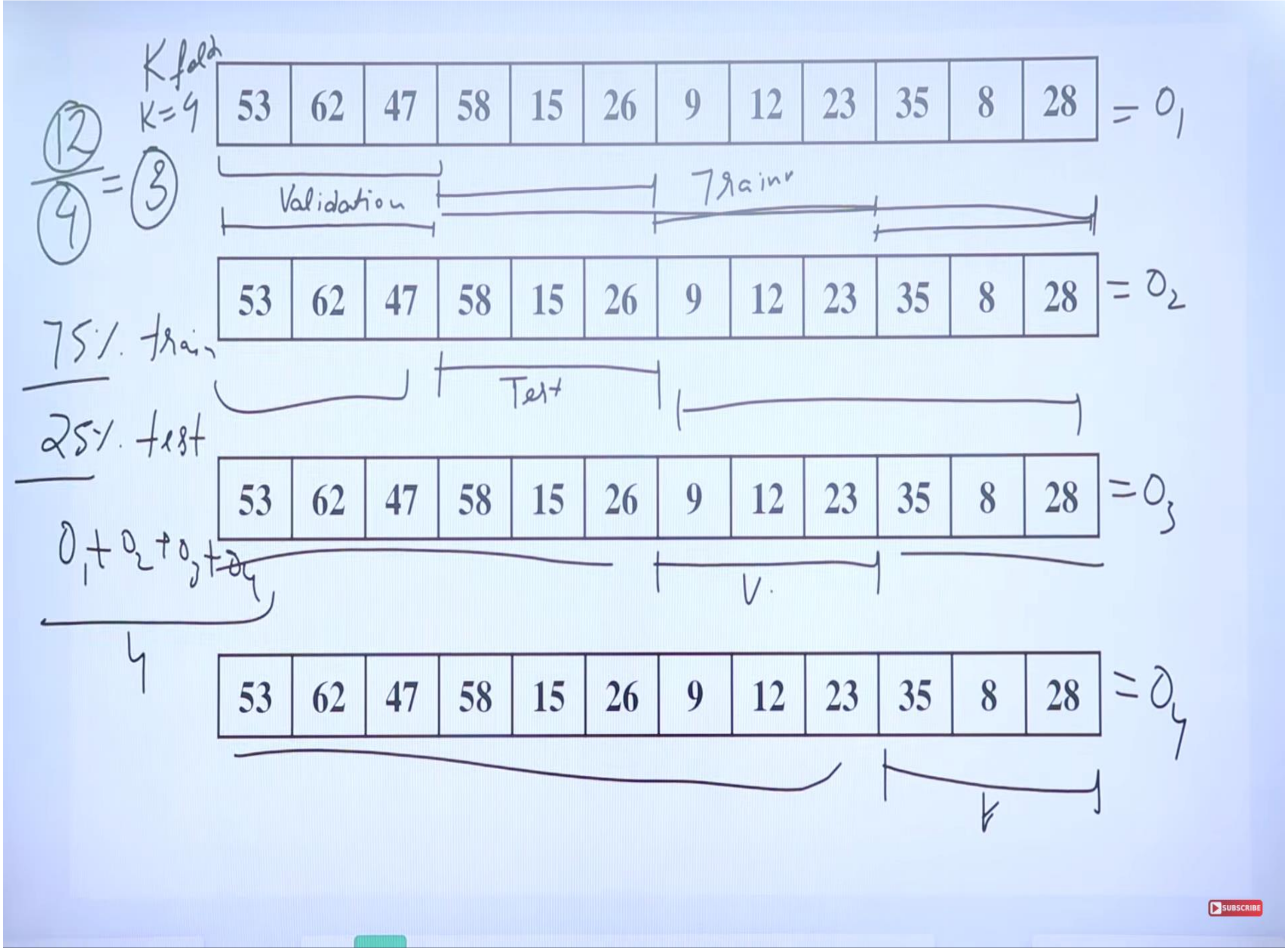

Writing in Python
import statistics as stat
data = [2, 3, 4, 5, 8, 9, 32]
meanval = stat.mean(data)
med_val = stat.median(data)
mod_val = stat.mode(data)
var_val = stat.variance(data)
sd_val = stat.stdev(data)


Step1: Initialize Weights
Step2: Train Weak Learner 1
Step3: Calculate Total Error
Step4: Calculate Weak Learner Weight(α)= 1/2 ln(1−Total Error / Total Error)
Step5: Update Instance Weights (NW) = Old weight × e^α
Repeat from Step2…
Instance Feature 1 Feature 2
1 1000 450
2 2000 475
3 3000 545
4 5000 756
5 6000 898

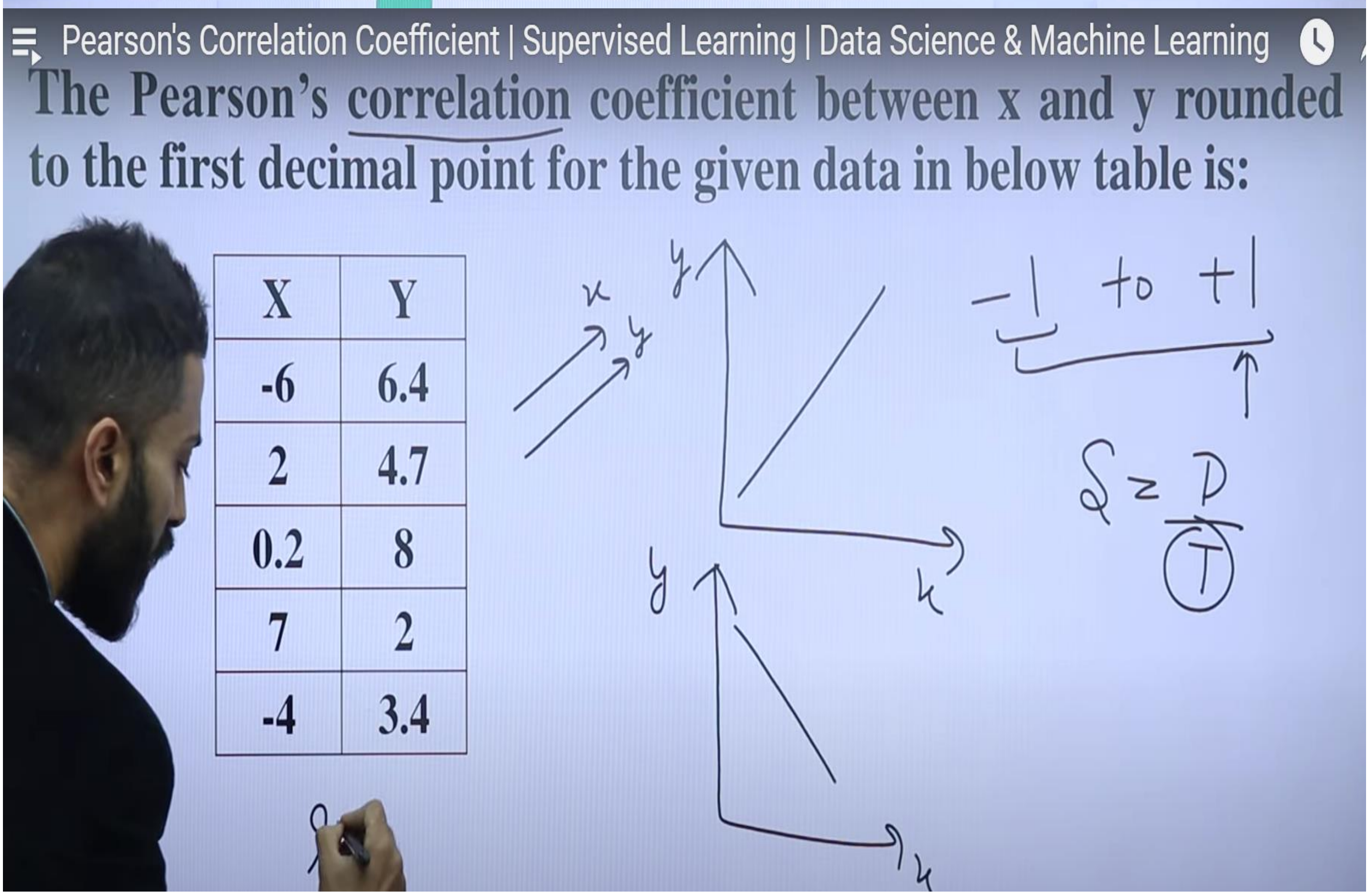

Bagging
Boosting
Stacking
Meta


Two bags are there containing Red and black balls; first bag has 3 red and 2 black balls, second bag has 2 red and 3 black balls. One ball is drawn randomly. Find the probability that it is Red ball?
P(E) = Red = P(A)×P(E/A) + P(B)×P(E/B)
P(A) = Bag 1 choose = 1/2
P(B) = Bag 2 choose = 1/2
B1 3R 2B
B2 2R 3B
A P(E/A) 3/5
B P(E/B) 2/5
1/2 × 3/5 + 1/2 × 2/5
Without any biasing one bag is chosen from that one ball is chosen randomly which was Red. What is probability that it came from the second bag?
P(B/E) = P(B)×P(E/B) / Total = (1/2 × 2/5) / (1/2 × 3/5 + 1/2 × 2/5)


Cross Validation in Machine Learning with Examples
• Cross-validation is a resampling technique
• It helps in estimating how well the model will perform on an independent dataset.
Sample Data
Train Test
100
80% 20%
1) K fold
2) LOO



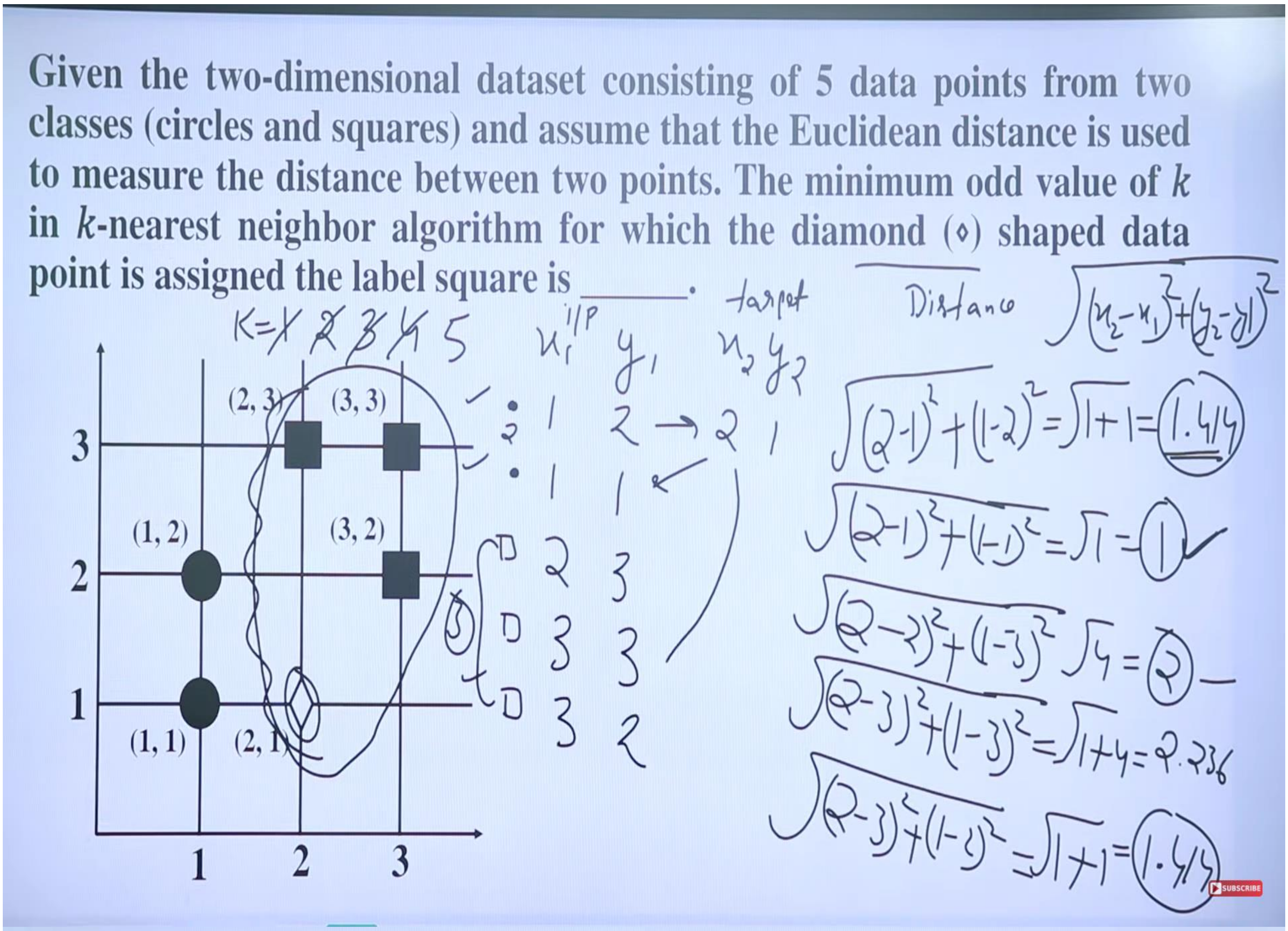
K fold
K=4
12/4 = 3
75% train
25% test
53 62 47 58 15 26 9 12 23 35 8 28 = O1
Validation
Train
53 62 47 58 15 26 9 12 23 35 8 28 = O2
Test
53 62 47 58 15 26 9 12 23 35 8 28 = O3
O1+O2+O3+O4 / 4
V.
53 62 47 58 15 26 9 12 23 35 8 28 = O4



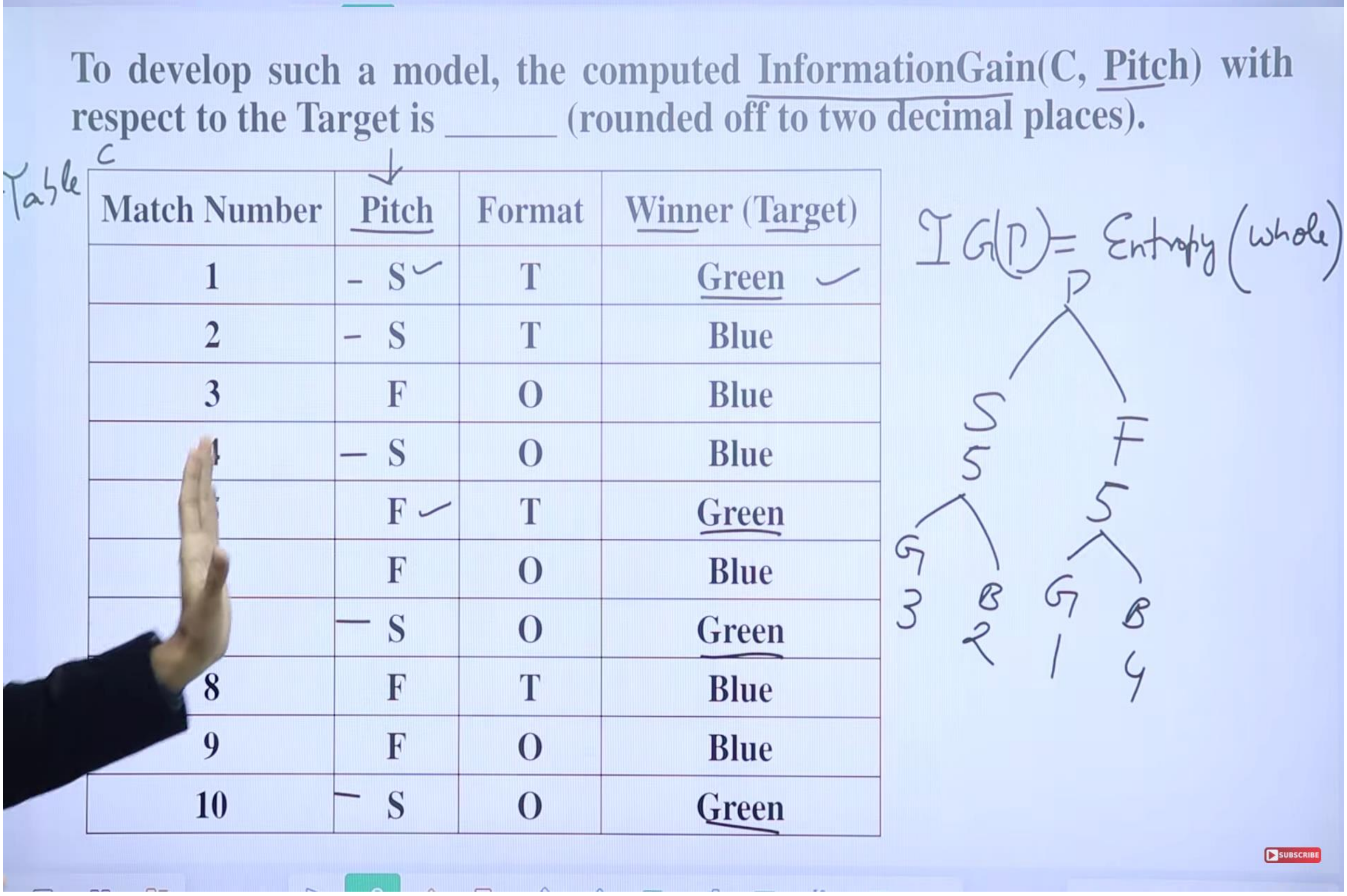
Pearson's Correlation Coefficient | Supervised Learning | Data Science & Machine Learning
The Pearson's correlation coefficient between x and y rounded to the first decimal point for the given data in below table is:
X Y
-6 6.4
2 4.7
0.2 8
7 2
-4 3.4
-1 to +1

$$r = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\dfrac{\sum_{i=1}^{n}(x_i-\mu_x)\times(y_i-\mu_y)}{n}}{\sum_{i=1}^{n}\sqrt{\dfrac{(x_i-\mu_x)^2}{n}} \times \sqrt{\dfrac{(y_i-\mu_y)^2}{n}}}$$

X	Y	μ_x	μ_y	$x-\mu_x$	$y-\mu_y$	$(x-\mu_x)*(y-\mu_y)$	$(x-\mu_x)^2$	$(y-\mu_y)^2$
-6	6.4	-0.16	4.9	-5.84	1.5	-8.76	34.1056	2.25
2	4.7			2.16	-0.2	-0.432	4.6656	0.04
0.2	8			0.36	3.1	1.116	0.1296	9.61
7	2			7.16	-2.9	-20.764	51.2656	8.41
-4	3.4			-3.84	-1.5	5.76	14.7456	2.25
						Total= -23.08	Total= 104.91	Total= 22.56
						Avg= -4.6	Avg= 20.98	Avg= 4.5

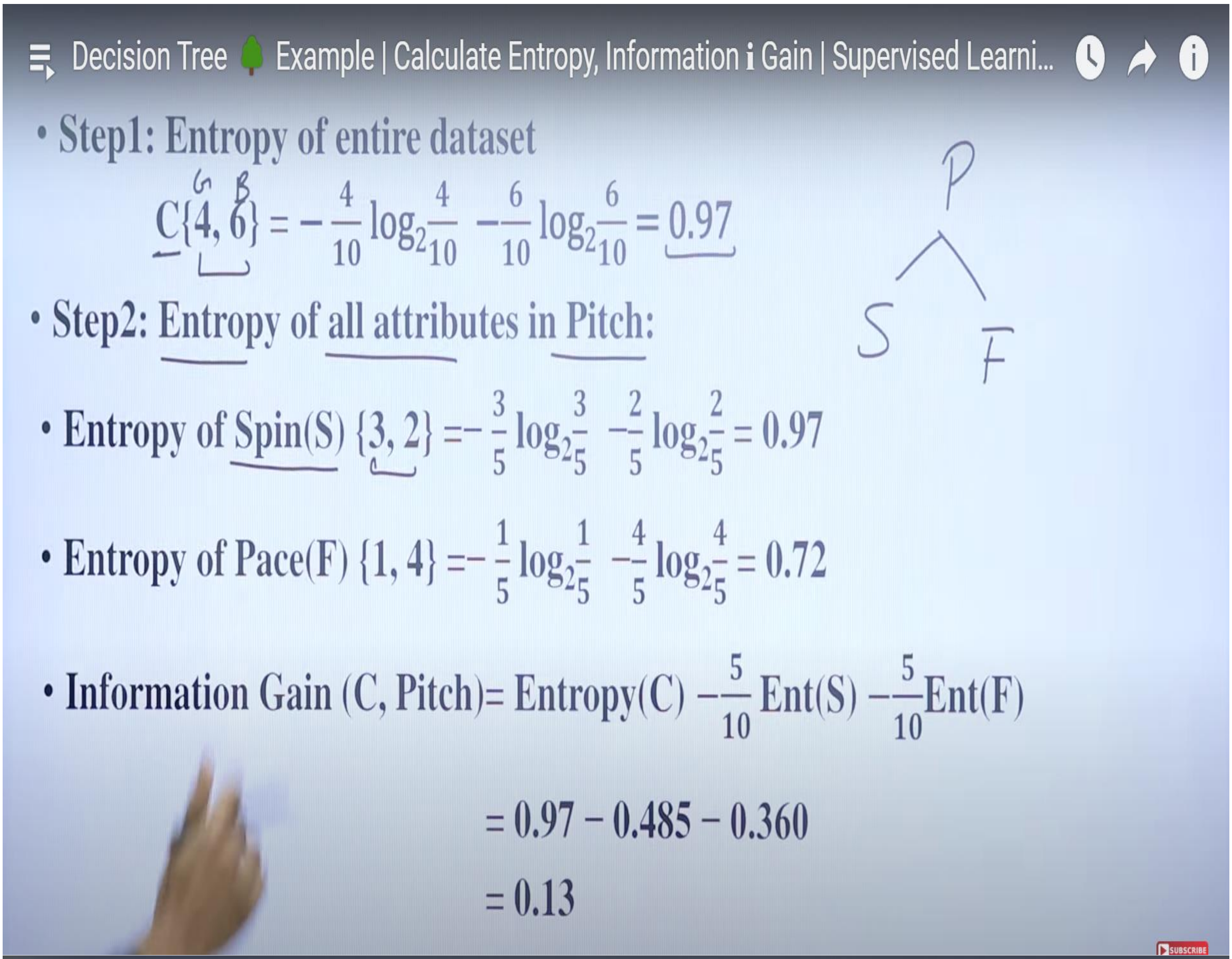


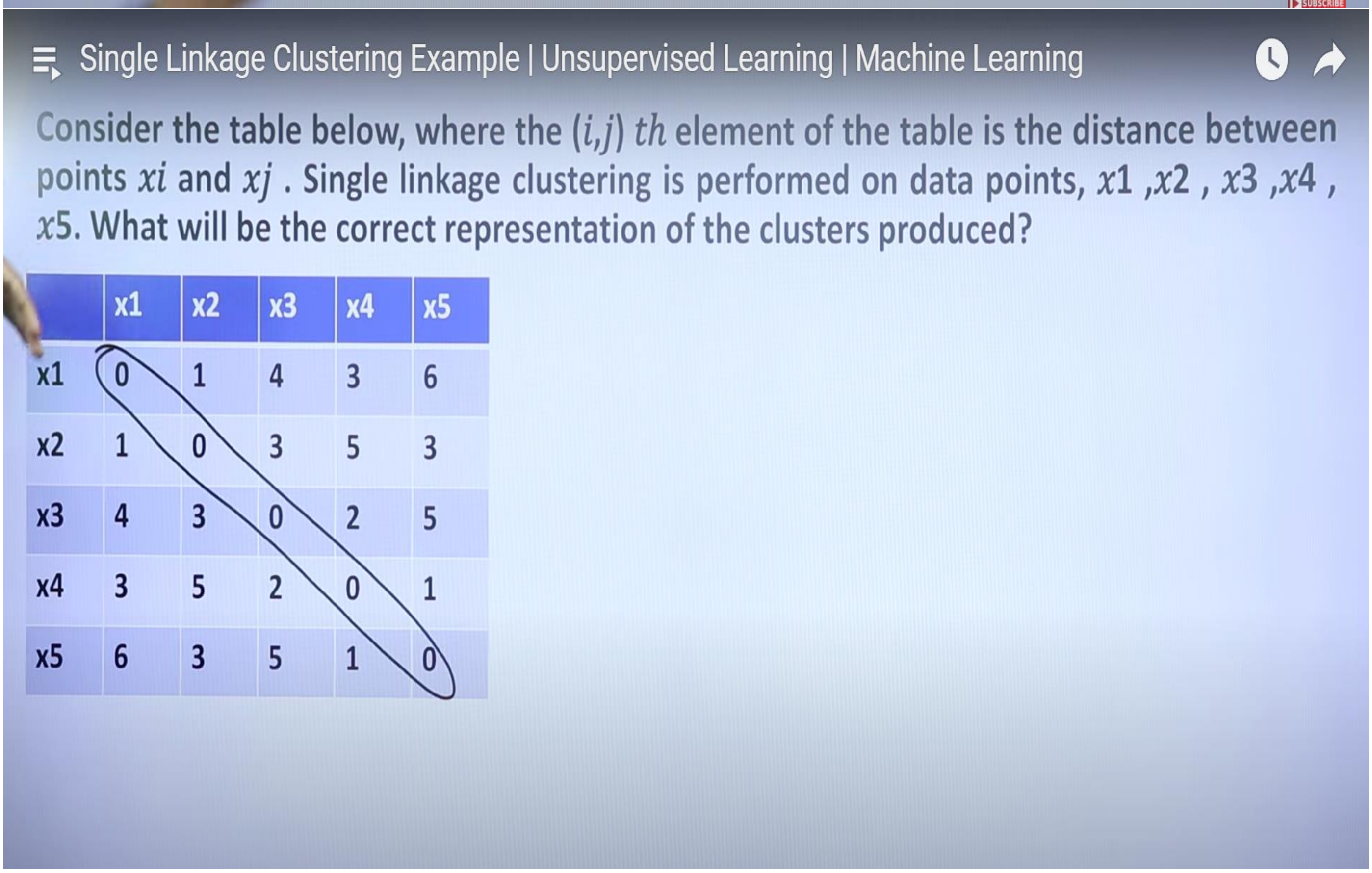
Given the two-dimensional dataset consisting of 5 data points from two classes (circles and squares) and assume that the Euclidean distance is used to measure the distance between two points. The minimum odd value of k in k-nearest neighbor algorithm for which the diamond (◊) shaped data point is assigned the label square is ______.



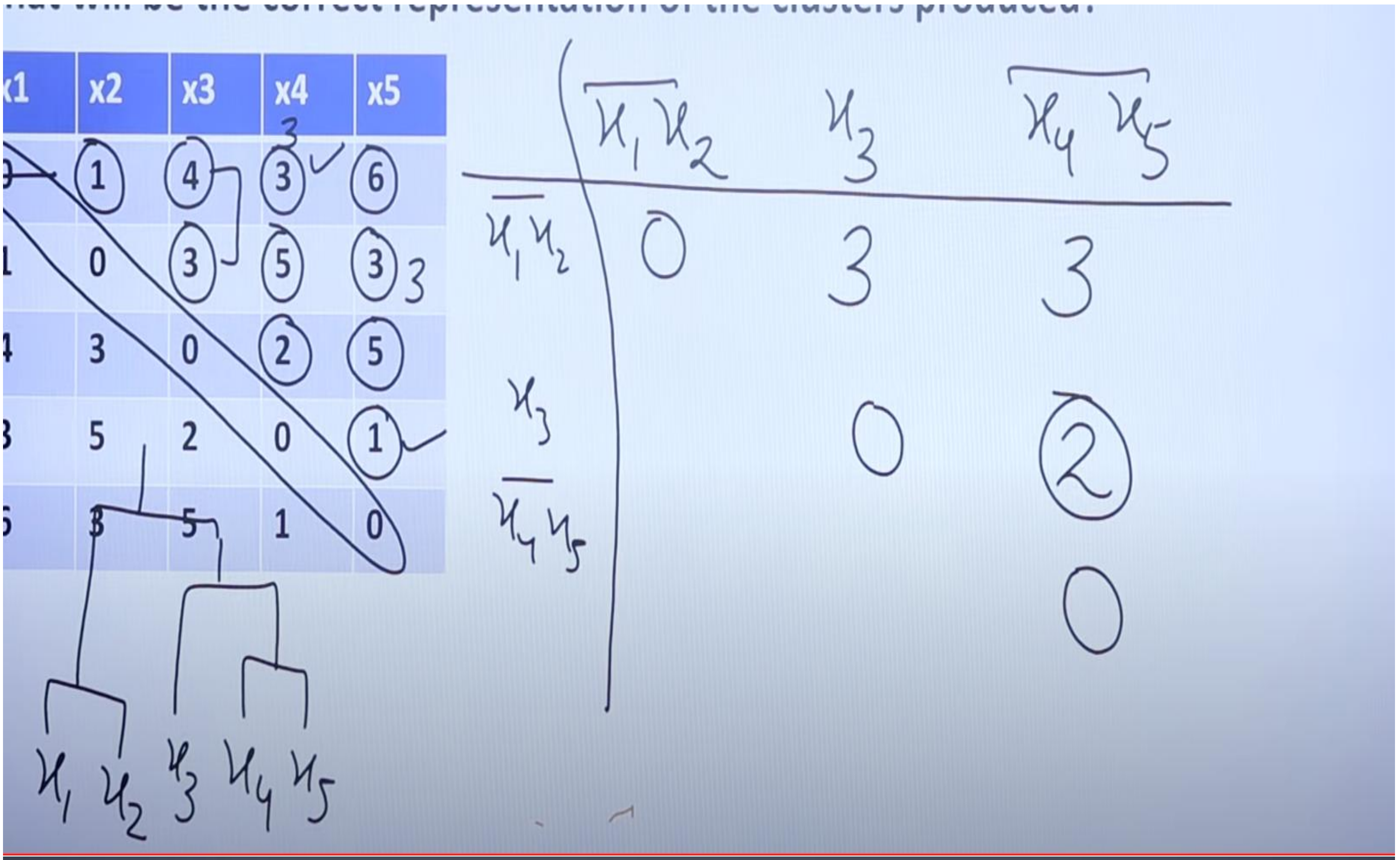
To develop such a model, the computed InformationGain(C, Pitch) with respect to the Target is ______ (rounded off to two decimal places).

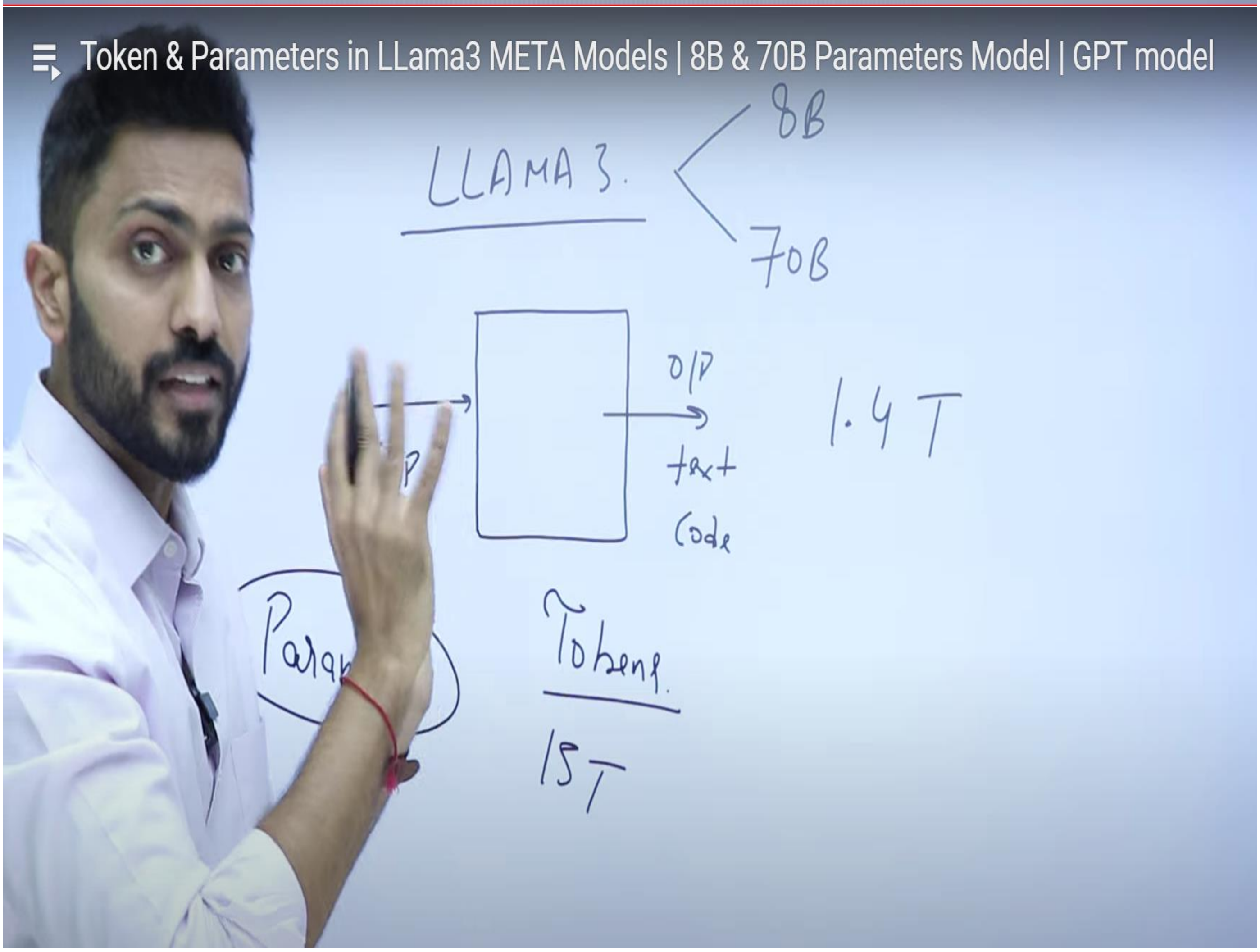
Match Number	Pitch	Format	Winner (Target)
1	S	T	Green
2	S	T	Blue
3	F	O	Blue
4	S	O	Blue
[illegible]	F	T	Green
[illegible]	F	O	Blue
[illegible]	S	O	Green
8	F	T	Blue
9	F	O	Blue
10	S	O	Green

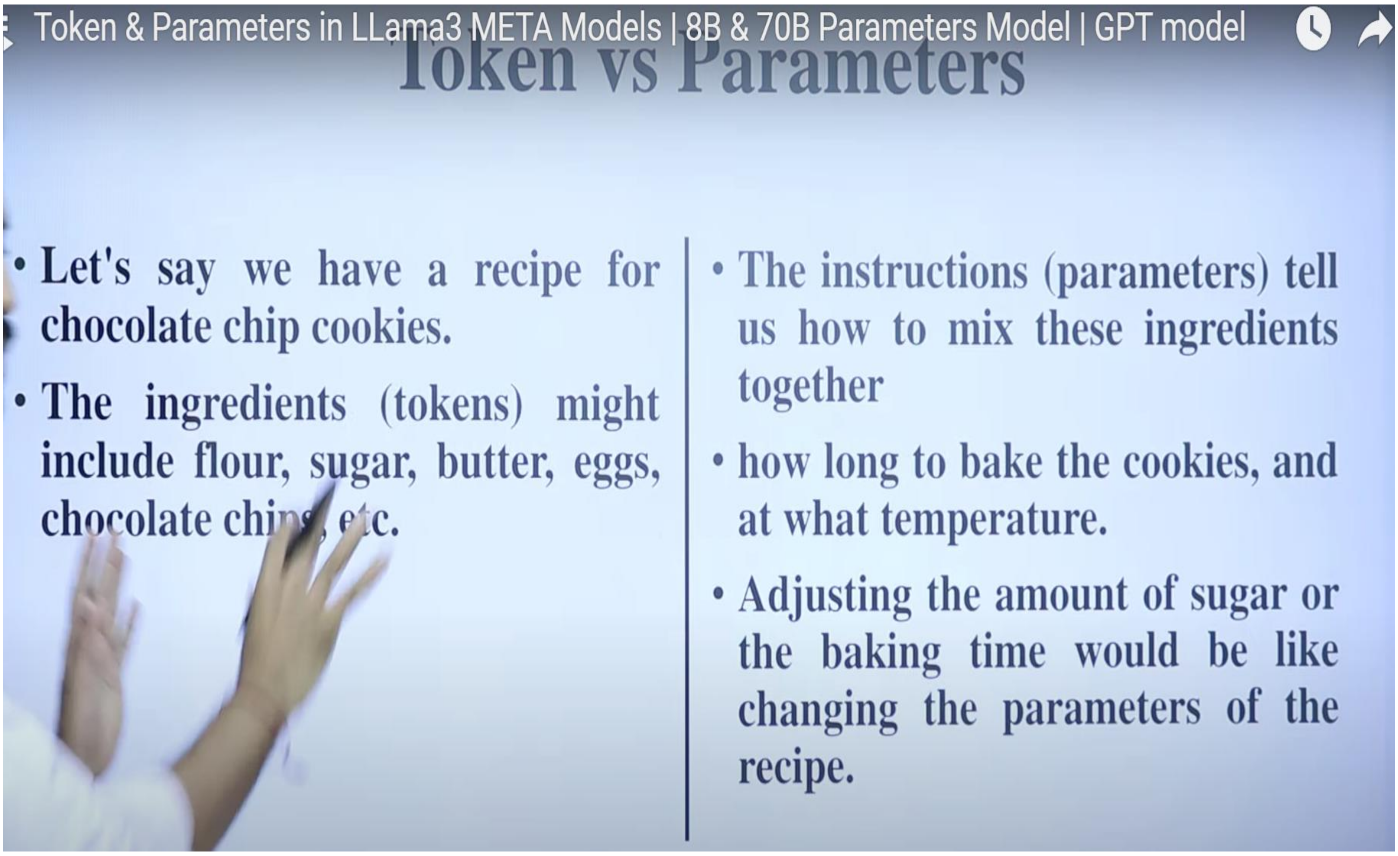




	x1	x2	x3	x4	x5
x1	0	1	4	3	6
x2	1	0	3	5	3
x3	4	3	0	2	5
x4	3	5	2	0	1
x5	6	3	5	1	0


x1
x2
x3
x4
x5


Token & Parameters in LLama3 META Models | 8B & 70B Parameters Model | GPT model
LLAMA 3.
8B
70B
O/P
text
Code
1.4 T
Tokens
15T


Token & Parameters in LLama3 META Models | 8B & 70B Parameters Model | GPT model
Token vs Parameters
• Let's say we have a recipe for chocolate chip cookies.
• The ingredients (tokens) might include flour, sugar, butter, eggs, chocolate chips, etc.
• The instructions (parameters) tell us how to mix these ingredients together
• how long to bake the cookies, and at what temperature.
• Adjusting the amount of sugar or the baking time would be like changing the parameters of the recipe.