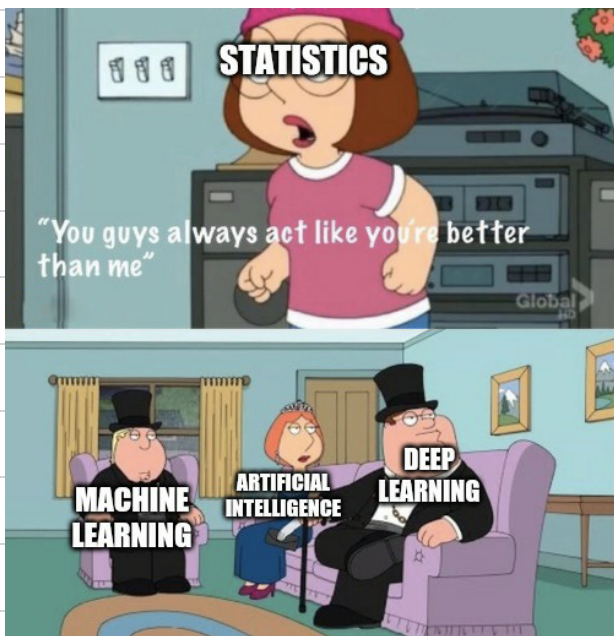
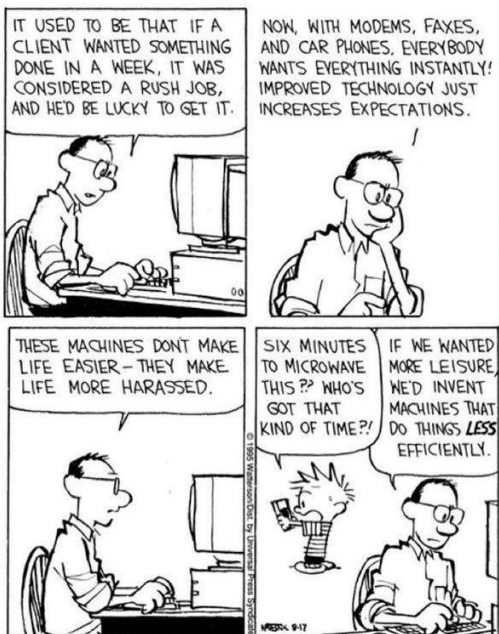


Session -1

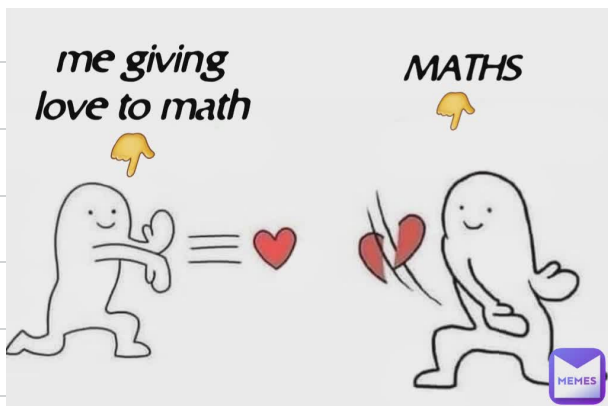
KNN -1

Aug 25, 2025



Agenda

- ① KNN
- ② SMOTE



ML models we're going to learn in this Module

- ① KNN \rightarrow Classification
- ② Decision trees \rightarrow $\begin{matrix} C \\ \hookrightarrow R \end{matrix}$
- ③ Random Forest \rightarrow $\begin{matrix} C \\ \hookrightarrow R \end{matrix}$
- ④ Light GBM / XGBoost $\begin{matrix} C \\ \hookrightarrow R \end{matrix}$
- ⑤ Naive Bayes \rightarrow C
- ⑥ SVM $\begin{matrix} C \\ \hookrightarrow R \end{matrix}$

When to use which ML model?

- ① Size / type of dataset
- ② Latency / Throughput of ML models
- ③ Deployment hardware.
- ④ Thickness of your wallet

2022

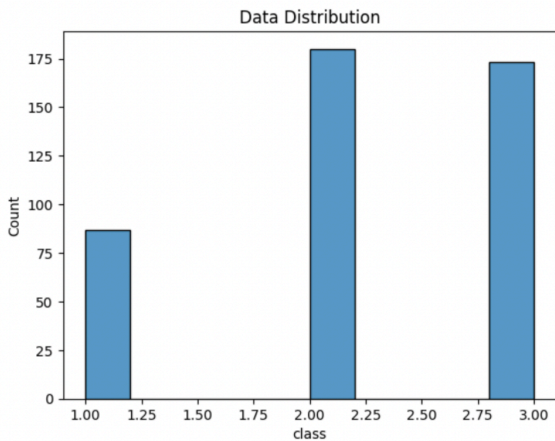
GPT ↓ - decent

2023

↓
GPT? acc↓

Powerful \propto $\frac{1}{\text{Speed.}}$

Latency = time taken for your computer to send request to server + time taken for server to generate response + time taken for their server to send response to your computer.



What can be said about the data ?

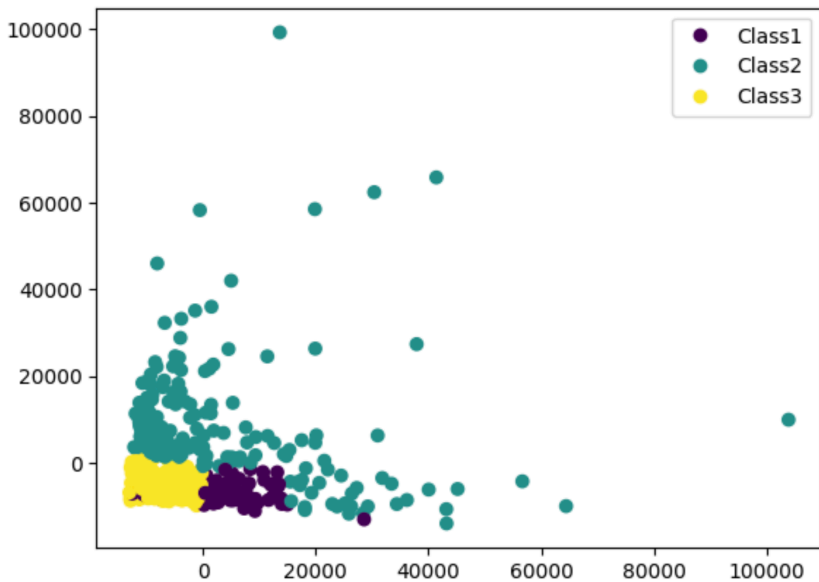
0 users have participated

- ☐ A Multi-class balanced data 0%
- ☒ B Multi-class imbalanced data 0%
- ☐ C Binary-class imbalanced data 0%
- ☐ D Binary-class balanced data 0%

[End Quiz Now](#)

Based on all quizzes from the session

KG 2 Komal Garg 1/1 96.43	K 1 Karthik 1/1 96.49	3 Shoreya gupta 1/1 96.20
4 Mohanakrishna 1/1 95.83	5 Suvaprada Dash 1/1 95.33	6 Praveen 1/1 94.97
7 Aditya Shandilya 1/1 94.53	8 Sri Harsha Nanduri 1/1 94.26	9 OM PRAKASH S 1/1 94.20
10 Umar 1/1 94.00		



How will Logistic Regression handle non-linear, multi-class data ?

0 users have participated

✓

A

Polynomial, OneVsRest

0%

B

Linear, OneVsRest

0%

C

OneVsRest, Polynomial

0%

D

OneVsRest, Linear

0%

End Quiz Now

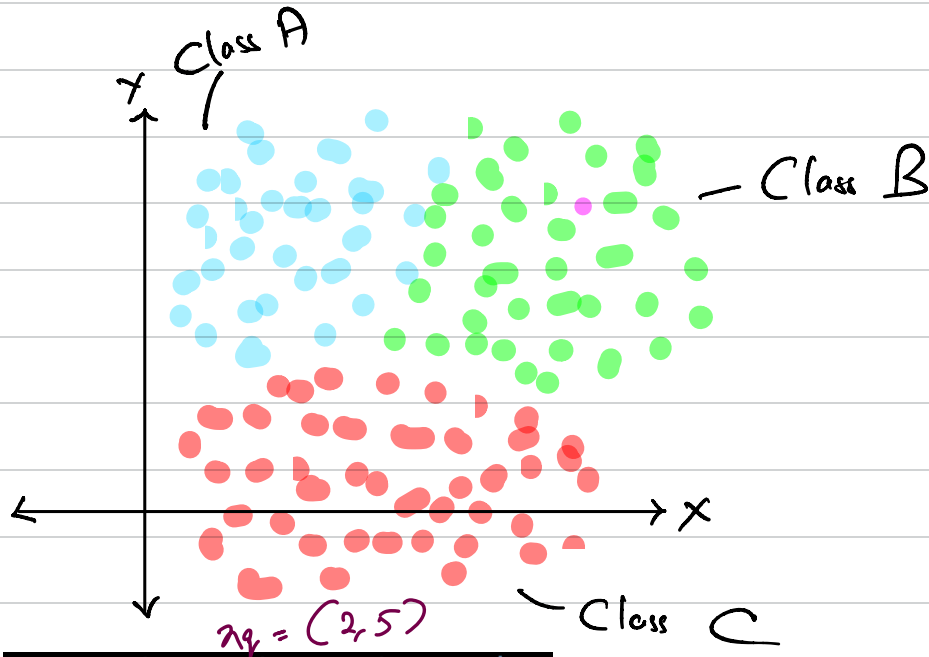
Based on all quizzes from the session

Shoreya gupta	Karthik	Mohanakrishna
2/2 ⚡ 189.47	2/2 ⚡ 233.24	2/2 ⚡ 186.47

4/10	Nikhil Kumar Nigam	2/2 ⚡ 185.53
5/10	OM PRAKASH S	2/2 ⚡ 183.07
6/10	Aditya Shandilya	2/2 ⚡ 182.33
7/10	Abdur Rehman	2/2 ⚡ 182.09
8/10	Sri Harsha Nanduri	2/2 ⚡ 181.40
9/10	Umar	2/2 ⚡ 178.73
10/10	Santhosh	2/2 ⚡ 177.93

How does KNN work??

K nearest neighbours



$$= \sqrt{((3-2)^2 + (6-5)^2)}$$

	f^1	f^2	y
x^1	3	6	1
x^2	6	4	1
x^3	8	2	3
x^4	7	5	3
x^5	1	4	2
x^6	2	2	2

	f^1	f^2	y
x^1	3	6	1.41
x^2	6	4	3.00
x^3	8	2	6.48
x^4	7	5	5.00
x^5	1	4	1.41
x^6	2	2	2.00

	f^1	f^2	y	y
X^1	3	6	1	1.41
X^5	1	4	2	1.41
X^6	2	2	2	2.00
X^2	6	4	1	3.00
X^4	7	5	3	5.00
X^3	8	2	3	6.48

$(1, 1)$

$(2, 1)$

L2 Norm

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

K-value = How many of my neighbors should I look at for deciding the class of my query point?

I'll do a voting on the class of my three most nearest neighbours.

$$\underline{12=3}$$

All machine learning models are wrong, but some of them are useful.

Poll results!

background?

39 responses from 39 users

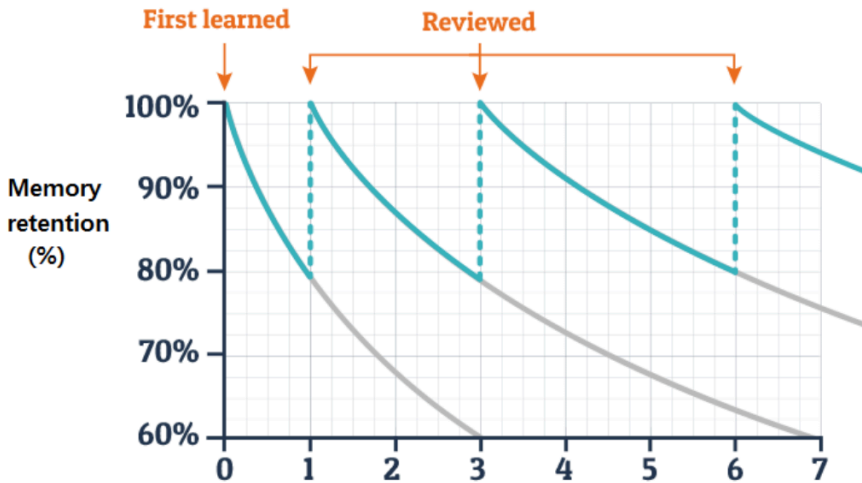
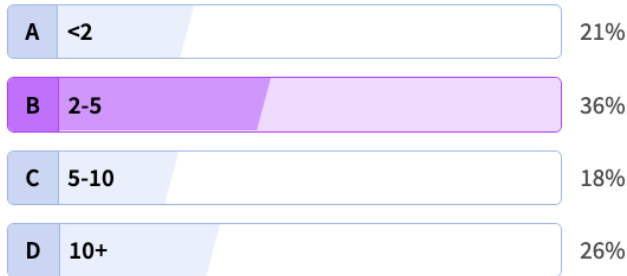
A non-tech 51%

B tech 49%

Poll results!

years of exp

39 responses from 39 users



$$\begin{array}{r} 2^0 \\ 1 \\ 1 \end{array} \quad \begin{array}{r} 2^1 \\ 1 \\ 2 \end{array} \quad \begin{array}{r} 2^2 \\ 1 \\ 4 \end{array} \quad \begin{array}{r} 2^3 \\ 1 \\ 8 \end{array}$$

$$\frac{9:03 - 10 \text{ mi}}{9:15}$$

$$\frac{9:15 - 9:30}{\text{Interval}}$$

Knn is a non-parametric algorithm.

Training phase { Load data into RAM }

Testing phase {
1. Computing L2 norm
2. Sorting.
3. Picking top K Neighbors

Arrange the statements in correct order based on kNN algo

s1- find majority vote

s2- perform euclidean distance

s3- sort and select k datapoints

s4- give class to x_q datapoint

0 users have participated

A s2,s1,s3,s4 0%




B s2,s3,s4,s3 0%








✓ C s2,s3,s1,s4 0%

D s2,s4,s3,s1 0%

[End Quiz Now](#)

Based on all quizzes from the session

		
Shreya Gupta	Karthik	Abdur Rehman
3/3 ⚡ 280.53	3/3 ⚡ 287.74	3/3 ⚡ 272.78

4		Sri Harsha Nanduri	3/3 ⚡ 266.69
5		OM PRAKASH S	3/3 ⚡ 266.00
6		Umar	3/3 ⚡ 265.87
7		SHASHANK JHA	3/3 ⚡ 259.96
8		Santhosh	3/3 ⚡ 258.20
9		Aditya Shandilya	3/3 ⚡ 257.80
10		Harshitha Chowdary Potturi	3/3 ⚡ 255.70

POINTS TO REMEMBER

- kNN is a non parametric algorithm.
- kNN predicts class of test data $[x_q]$ on the basis of neighbourhood.



WORKING OF kNN:

- Find distance (x_q and all training data)
- Sort distance
- Pick k nearest neighbors
- Majority rate of class prediction

If x_1 at (4,0) , x_2 at (0,1) and x_3 (5,0) and x_q at (0,0). Which is nearest point to x_q ?

1 user has participated

✓	A	<input type="text" value="x1"/>	0%
	B	<input checked="" type="text" value="x2"/>	100%
	C	<input type="text" value="x3"/>	0%

[End Quiz Now](#)

Based on all quizzes from the session


2
Shreyas Gupta
4/14 366.66


1
Karthik
4/14 379.01


3
Abdur Rehman
4/14 362.69

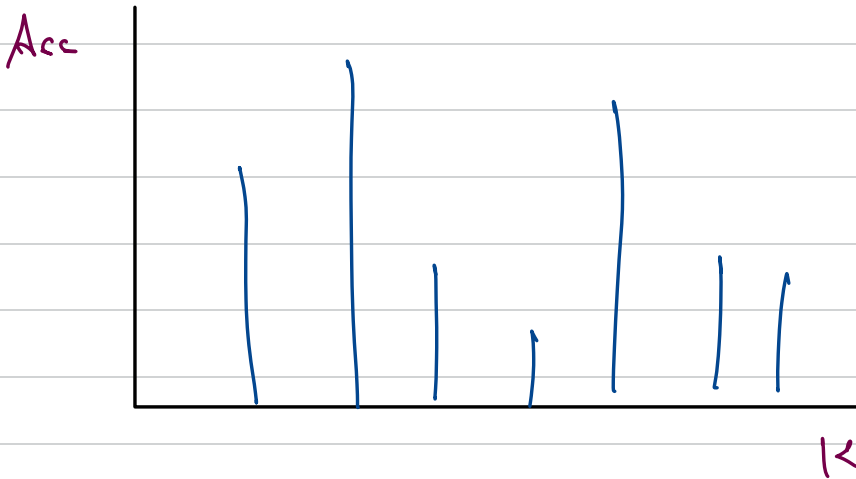
4	Sri Harsha Nanduri	4/14	357.16
5	OM PRAKASH S	4/14	355.33
6	Manjeet Singh	4/14	345.48
7	Santhosh	4/14	344.89
8	Aditya Shandilya	4/14	344.86
9	SHASHANK JHA	4/14	344.56
10	Umar	4/14	334.17

Wherever distance calculation is involved, you should always standardize your dataset.

$(60,000, 4)$
A

$(40,000, 2)$
B

$$\sqrt{(6012 - 4012)^2 + (4 - 2)^2}$$



1. How can the value of k impact underfitting and overfitting?
2. Time Complexity of KNN

How does kNN has good performance on non linear multi class data?

Assume data contains:

(+)

(-)

(o)

& k=5

xq

kNN fails when data has a lot of noise/outliers

kNN assumes neighbourhood as homogenous i.e, characteristics of nearest neighbour and x_q will be same

how kNN is better than logistic regression. Select the correct option

1 user has participated

A kNN has less time complexity

0%

B kNN classifies data better

0%

C kNN handles most noise/outlier

0%

D	kNN handles multi-class problem
---	---------------------------------

100%

[End Quiz Now](#)

Based on all quizzes from the session



 **Shoreya gupta**
5/5 ⚡ 459.53



 **Karthik**
5/5  477.06



OM PRAKASH S
5/5 ⚡ 447.70

4		Sri Harsha Nanduri	5/15	🔥	445.76
5		Manjeet Singh	5/15	🔥	428.62
6		Santhosh	5/15	🔥	415.89
7		RAHUL	5/15	🔥	411.76
8		Mohanakrishna	4/15	🔥	372.13
9		Abdur Rehman	4/15	🔥	362.69
10		Tanvi Singh	5/15	🔥	361.37

Closed Form \rightarrow Linear Reg