Shot on Y12
Vivo AI camera

Maching learing

* Types of Machine Learning

1] Supervised ML
2] Unsupervised ML
3] Semi- supervised ML
4] Reinforcement Learning

1 Supervised ML
- → Classificaition
- → Regression

Dataset → Specific o/p → Model → o/p

Dateset → you know the o/p (output) feature of the Dateset

A] Classification → o/p → categorial value होती है।

There are three feature

Independent feature

Played hours	Study hours	Pass / Fail → out ~~ecored~~ feature
8	2	Fail
7	3	Fail
6	4	Fail
5	5	Pass
4	6	Pass

o/p feature → dependent feature

इसमें output दिया होता है, जबकि input देने पर output predict करता है, model train होता है।

B] In Regression → o/p ⇒ continous value होती है।

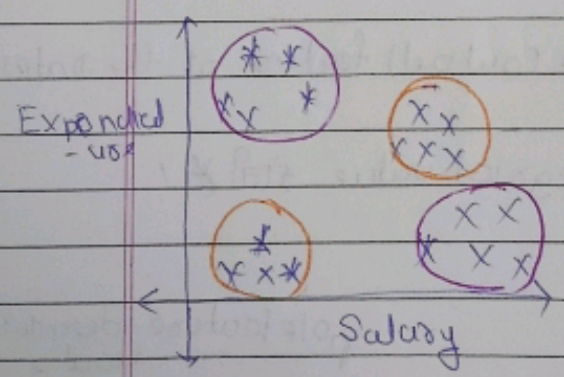
size of House | No. of Room | Price of House → continuous

2] UnSupervised ML :- (clusters or similar (Groups)

There are no output // समान features होते हैं // Group of similar Data

Ex:- if we have a Mall and data of customer found look out graph, then if we launch a watch The we hit a specifics adience with discount coupen (- or high expedutors persons so that make more profit

Ex:- Customer Segmentation



Poudst → clusters
↓
Target the audience
↓
Sol. Discount

3] Semi Supervised :-

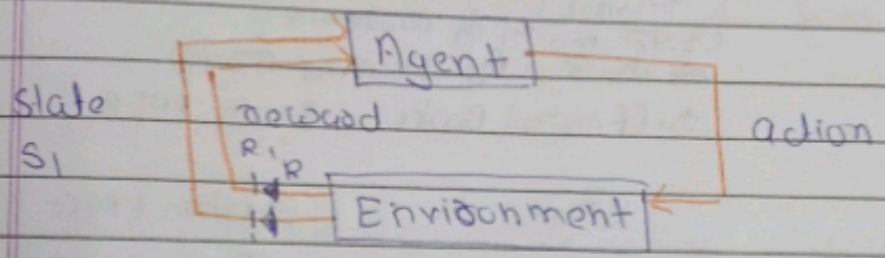
Supervised + Unsupervised

In these basically first we trained the model so that we collect the supervised data later on we used that data for custom clustering or custom segmentation in Unsupervised way There are combination of Both

Shot on Y12
Vivo AI camera

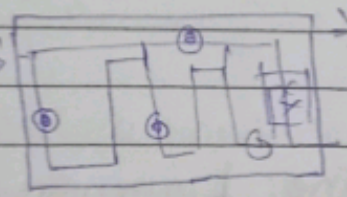
4] Reforcement Learning:

Reinforcement learning is an area of machine learning concerned with how intelligent agents ought to take actions in an enviornment in order to maximize the nation of cumulative reward Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.



For Ex

Agent →



→ Reforence learning मे Agent होता है जो सारे path को pass (कराएगा) करेगा At the same time वे collect करना जाएगा Rewards को और Major कोशिश की कौनसा path मे जादा Reward मिल रहे है।

Note all the combination उसको मे समझ आ जाएगा की कौनसे path मे जादा Reward मिल रहे है। Reward → Point OR Best way to pass the line. (या Best)

Ans Agent अपने environment, Rewards state उसके हिसाब से अपना action लेता जायेगा और slowly-2 मे समझता है की कि है अगर मे इस path को choose करना तो मेरा live Bot होगा क्योंकि मेरे को मेरा जादा Rewards मिलेगा.

Shot on Y12
Vivo AI camera

Famous
Page No.:
Date.:

DATA Set

Training Dataset	Validation Dataset	Test Dataset
We will Train the Model	Hyper parameter Tuning of the Model (इसमें model के parameter को तोड़ने से और Accuracy बढ़ाने)	Test the Model Accuracy??

Ex:-

Books	Different Books	Exam PPER
Q&A + TRAIN	↓ TRAIN	Question paper Train test ↓ 86%

Machine Learning

1] Model Performance - Accuracy ↑↑ → High
2] Overfitting, Underfitting
3] Bias vs variance

1] Overfitting Underfitting

DATASET

1] Books → Train → Model is Trained → Accuracy ↑↑ = 95%
Exam → Test → Model is Tested → Accuracy ↓↓ → 60%

{ Low Bias
High variance }

Shot on Y12
Vivo AI camera

2] Train → Accuracy ↓↓ - 55%
Test → Accuracy ↓↓ → 50%

} Underfitting
⇓
{ High Bias
High variance }

Generalized Model
Train → Acc ↑↑
Test → Acc ↑↑
} ⇒ { Low Bias
Low variance }

* Interview

Bias is Related to training Accuracy if,
- if Training Accuracy ↑↑ ⇒ Low Bias
- if Training Accuracy ↓↓ ⇒ High Bias

variance also Related to Test Accuracy
- if Testing Accuracy low ↓↓ ⇒ High variance
- if testing Accuracy ↑↑ ⇒ Low variance.

* Handling Imblanced Dataset :-

Classification Problem ⇒ Supervised ML Technique
↓
o/p Categorical Featurel
↓
let, 2 categorical feature → Yes/No

total data point
900 100

Shot on Y12
Vivo AI camera

f_1	F_2	O/P
–	–	Yes } 900 Yes, 100 No
–	–	No
–	–	Yes

1] 900 : 100 ⇒ 9 : 1 [Imbalanced dataset]
{ Becoz, No. of Yes is very more than No. of "No" }

2] 500 : 500 ⇒ Balanced dataset

900 : 100 [Imbalanced dataset]

900 Yes | 100 No

⇓

we create Dume model that always

⇓

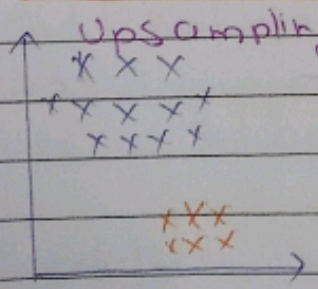
$\boxed{\text{Yes}}$

{ by this dumb dataset Accuracy → 90% }

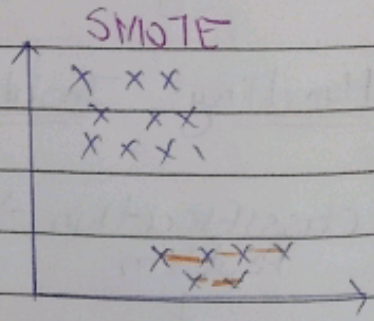
* Imbalanced dataset Problem solved by 2 techniques
1] Upsampling 2] Down sampling

$\boxed{\text{check out → In Jupyter Notebook}}$

Upsampling



SMOTE



Upsampling मे low data की Quantity increase कर देते हैं। पर इस professional मे सही तरीका नहीं होता।
Best तरीका होता है 'Imblanced को cover up करने के लिए, जिसे $\boxed{\text{Interpolation}}$ कहते हैं

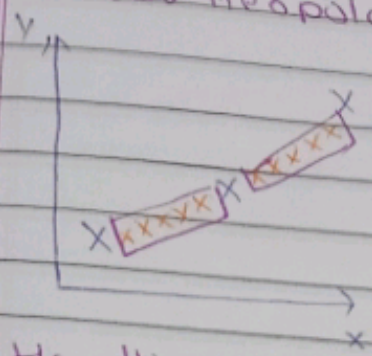
↳ In SMOTE सभी के values मे line draw connection बनाता है और फिर उसे Interpolation through value fill up होती है।

$\boxed{\text{look out → Jupyter file}}$

Shot on Y12
Vivo AI camera

* Data Interpolation

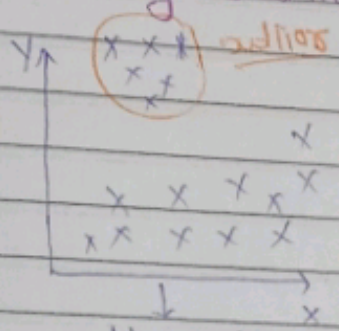
1 Linear interpolation:



→ Polynomial से curve बनता है। फिर उससे से value generate करता है।

In Jupter

2] Handling outliers:



outliers

{1, 2, 3, 4, 5, 100, 125}

outliers impact Algorithms

* Feature Extraction

Feature Extraction is process of selecting and extracting the most importance features from raw data.

ML Application → 1000 features
↓
Most Important features
↓
Machine Learning Algo.

1] Feacture Scaling

(Y) Age	(kg) Weight	(cm) Height	BMI
32	70	160 cm	—
28	75	160 cm	—
35	80	155 cm	—

Normalize or Standardize
↓
$Z = \text{code} = \boxed{\mu = 0 \quad \sigma = 1}$

$$\boxed{\text{Zscore} = \frac{x_i - \bar{x}}{\mu}}$$
↓
Standardization

Normalization [Min Max Scaler]
(0, 1)

unit vector

Feature scaling एक ऐसा Technique है जिसमें जगह आप specifically features को scale down कर रहे हैं. Both some range of can be between 3 to +3

Shot on Y12
Vivo AI camera

Famous
Page No.:
Date.:

2] Feature selection

→ We just pick the most Important feature

500 features → Top 10 features
↓
ML Model Train

correlation
↓
feature selection

1] filter Method 2] Embeded Method

3] PCA [Principal component Analysis]

** 1] Feature Scaling
1] Standardization → Z score
2] Min Max Scaling [Normalization]
3] Unit vector

1] Standardization

Age
24
25
26
27
28
29

$$\boxed{Z\text{-score} = \frac{x_i - \bar{x}}{\sigma}}$$

$\mu = 0 \quad \sigma = 1$

Age'
$\mu = 0 \quad \sigma = 1$

Shot on Y12
Vivo AI camera

2] Normalization [Min Max Scaler] → 0 to 1

Age
24
25
26
27
28
29
30

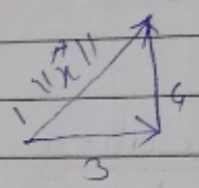
Transmation

$$\boxed{x_{scaled} = \frac{x_i - x_{min}}{x_{max} - x_{min}}}$$

Age' [0-1]

3] ~~PCA~~ Unit vector :-

Magnitue of 1



$\vec{x} = (3, 4)$

$\|\vec{x}\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$

$\hat{u} = \left(\frac{3}{\|x\|}, \frac{4}{\|x\|}\right) = \left(\frac{3}{5}, \frac{4}{5}\right)$

$\|\hat{u}\| = \sqrt{\left(\frac{3}{5}\right)^2 + \left(\frac{4}{5}\right)^2} = \sqrt{\frac{25}{25}} = 1$

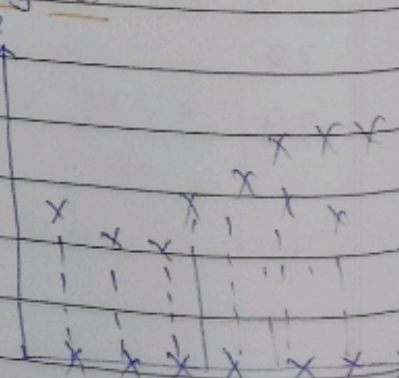
$\boxed{\|\hat{u}\| = 1}$ * checks sucessfully

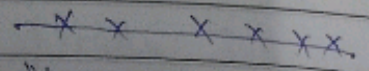
Mean work → scalling the data

* PCA (Principal Component Analysis) :-

Dataset

No. of Rooms	House size	Price
2	800	40lak.
–	–	–
–	–	–
–	–	–





No of Rooms Price

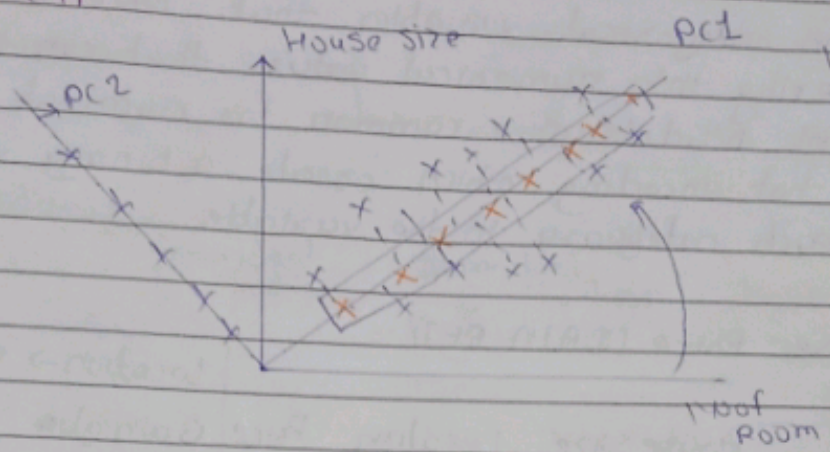
Machine Learning model

2D Plane only take on fix and Make ML model and unit.

Shot on Y12
Vivo AI camera

Problem - Data Loss

* PCA



100D → 2D
2D → 1D
3D → 2D

No of Room | House size | PC1 Price | ⇒ Model Train

DATA ENCODING

I/P feature

Experience	Education	Salary
4	B.E	50k
6	PHD	100k
3	Master	80k
6	PHD	120k

Education → categorical data

Salary → O/P [Dependent feature]

I/P [Independent feature]

* DATA ENCODING :-

AIM :-

convert categorical feature $\xrightarrow{\text{into}}$ Numerical → ML Model → Train

Type of Data Encoding

1) Nominal / OHE (One Hot Encoding)
2) ORDINAL / LABEL ENCODING
3) Target Guided ordinal encoding

} categorical → Numerical

Shot on Y12
Vivo AI camera

1) Nominal OHE :-

Nominal encoding is a technique used to transform categorical variables that have no intrinsic ordering into numerical values that can be used in ML Models One common for nominal encoding one hot encoding, which create a binary vector for each category in the variable.

House Price (DATA SET)

No of Rooms	House Size	Location	Price
		Banglore	
		Noida	
		Delhi	
		Banglore	
		Delhi	

Location → OHE

Banglore	Noida	Delhi
1	0	0
0	1	0
0	0	1
0	0	1

Disadvantage (OHE):

1) Sparse matrix → for ex [100 Location] it will create [many 0 and 1 Binary feature
↳ [overfitting]

2) if [1000 Location] → hey create → 100 feature → Don't use, Because it will Increase the No. of Feature.

$\boxed{\therefore \text{No of features} \uparrow\uparrow = \text{Model Peformace} \downarrow\downarrow}$

Shot on Y12
Vivo AI camera