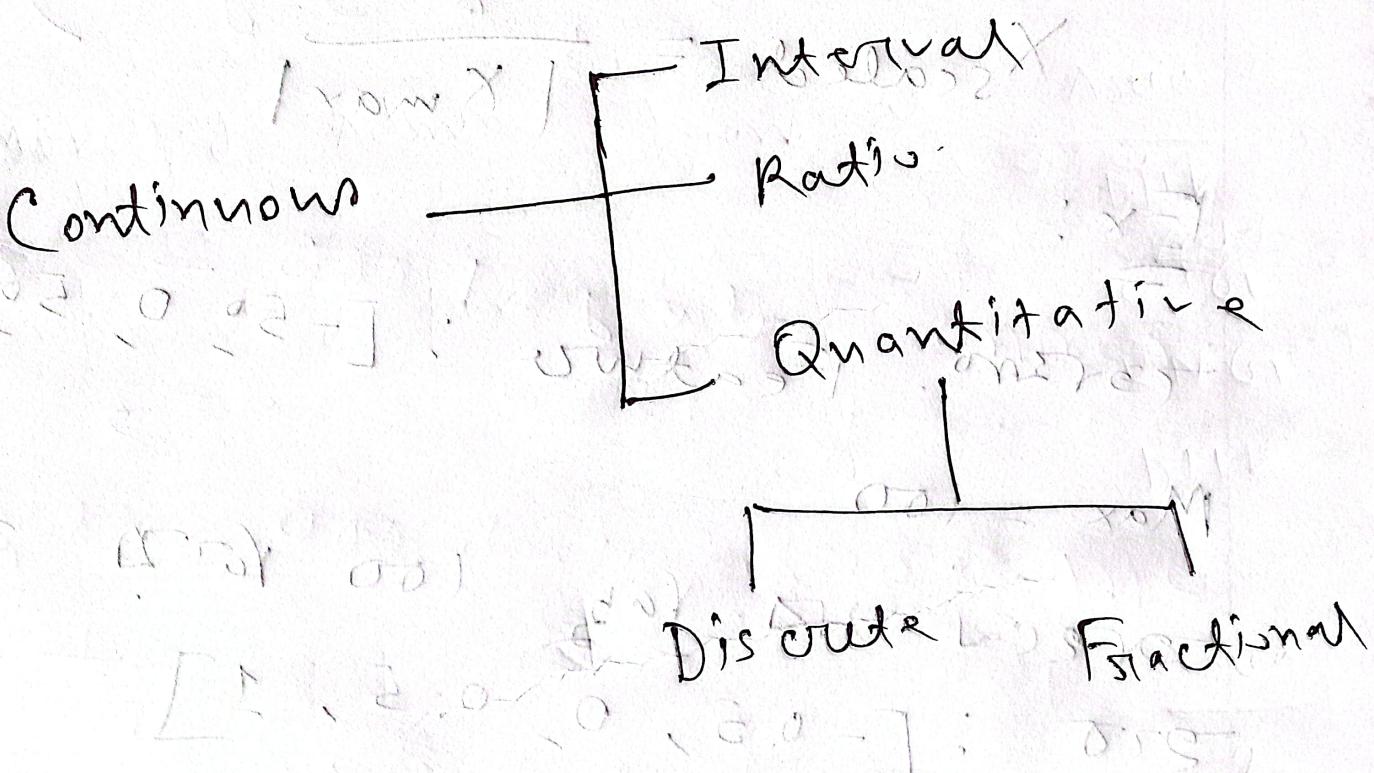
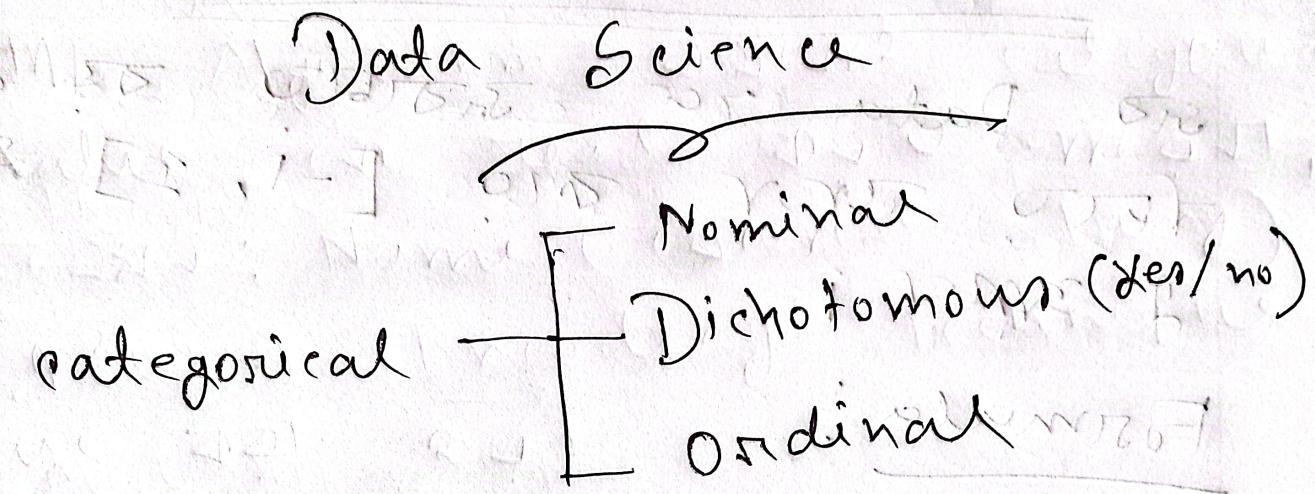


Types of variable in Data Science



Mr. Neelkanth Bribawar  
Lokam J.M.

Nominal	Dichotomous	Ordinal
1. colors names	1. Yes / No	① Good $\rightarrow$ Avg $\rightarrow$ Below
2. Location names	2. Male / Female	② High $\rightarrow$ Mid. $\rightarrow$ Low
3. Car names	3. Left / Right	

### Measure of Central Tendency

- ① Mean
- ② Median
- ③ Mode

# Encoding Techniques in ML Model

ML Algorithm କାହାରେ categorical value ରେ read କରିବାରେ ନେମାନ୍ ,  
ନାମ : Name (character string)  
ସ୍ଟ୍ରିଙ୍ଗ Encoding Technique  
use କଥା କଥା ।

```

graph TD
    MS[Marketing speed] --- B1[Branch 1]
    MS --- B2[Branch 2]
    Admin[Admin] --- B3[Branch 3]
    Transport[Transport] --- B4[Branch 4]
    Area[Area] --- B5[Branch 5]
    Area --- B6[Branch 6]
    Area --- B7[Branch 7]
    B4 --> B8[Branch 8]

```

The diagram illustrates a company's organizational structure with four main departments:

- Marketing speed** oversees two branches.
- Admin** oversees one branch.
- Transport** oversees one branch, which further oversees a branch labeled **Branch 8**.
- Area** oversees three branches.

① Without use any  
Encoding Techniques

② Label Encoding

③ One-hot Encoding

④ Ordinal Encoding

⑤ Hash Encoding

## Without using Encoding Techniques

Height

Tall

Medium

Short

Height

2

4

6

Replace function  
manually encode

use

32(3)

37(3)

error

## Label Encoding

Alphabetic Order follow onto unique object to map unique name set into numbers.

Area	Area
DHK	1
CTG	0
RNG	1
DHK	0
CTG	1
DHK	0

Label Encoding  
Problems

$$\text{Dhaka} = 1$$

$$\text{CTG} = 0$$

$$\text{RNG} = 2$$

Algorithm use

2<sup>4</sup>, 10<sup>3</sup>

10<sup>3</sup> model

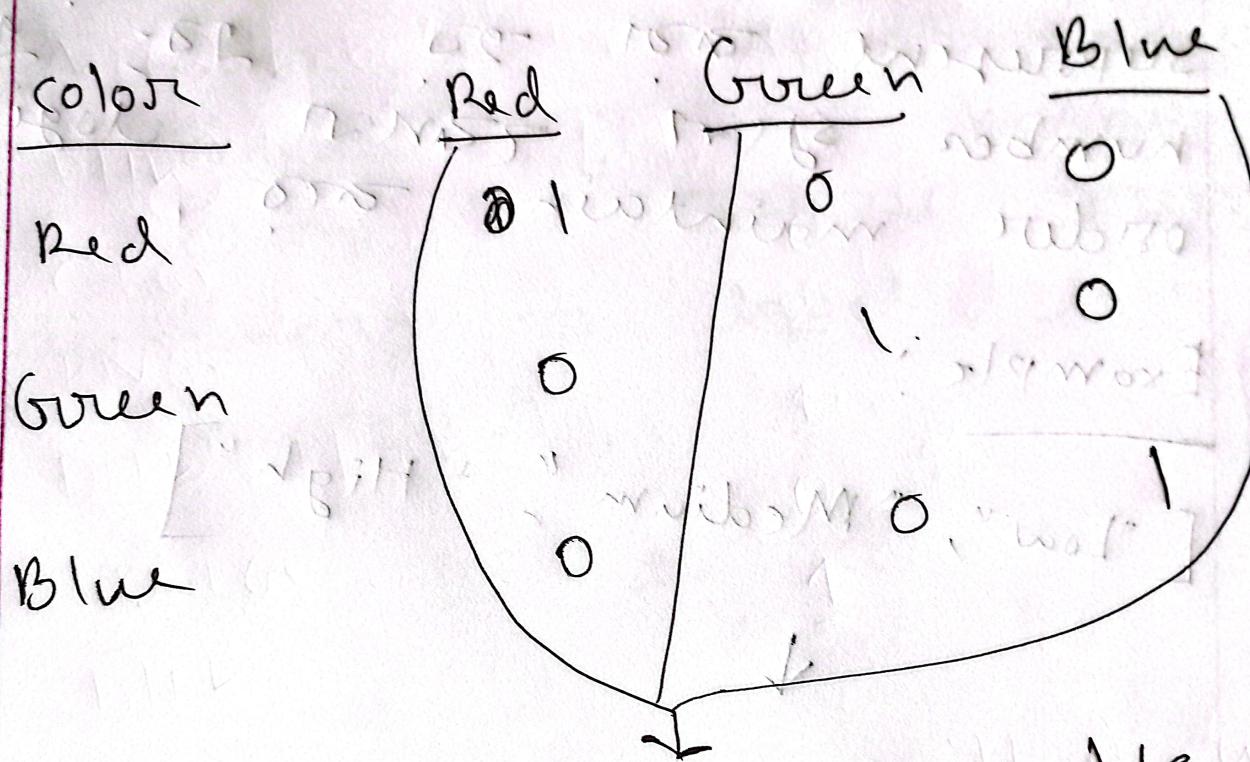
Q12 Distance based Algorithm  
एवं ये problem का तरीका तरीका  
ये, किसे ये problem को  
3 solution करें,

## One-hot Encoding

यह एक तकनीक है (DEMs)  
categorical data को binary  
(0/1) form में convert करती है  
में स्टेप्स को ranking/orden in  
into binary  
एवं ML model confused होता है  
यह category को बदलता है

[Red, "Green", "Blue"]

After one hot encoding.



Dummy variable

inplace use extra column  
drop training time  
model go,

## Ordinal Encoding

Ordinal encoding method jekhane categorical values as number represent 20, 23, 92- number form human logical order maintain

Example :

[ "Low", "Medium", "High" ]



"Low"  $\rightarrow$  1

"Medium"  $\rightarrow$  2

"High"  $\rightarrow$  3

## Hashing Encoder

Hashing Encoder (or Feature Hashing)  
is a technique to convert categorical data to fixed number of numerical features.  
A convert function is used to convert label list to dictionary.  
Hash DS is concept.]

### Drawbacks:

- Hash collision
- Interpretability