

**Data preprocessing** : 3 ga bolinadi

**Missing values**

**Encoding**

**Scaling**

**Encoding** => Categorical qiymatlarni numerical ga o'tqazish. Ya'ni object data type dan integer yoki float ga o'tish.

**Encoding 5 ga bolinadi**

- one-hot encoding

- label encoding

- frequency encoding

- target encoding

- ordinal encoding

**one-hot encoding** => datasetdagi har bir classlar uchun alohida ustun yaratib ajratadi.

-misol uchun color ustuni bor va unda 5 xil rang bor one-hot encoding har biriga ustun yaratadi

**Label encoding** => one-hot encoding kabi qo'shimcha ustunlar yaratmaydi. o'sha ustun saqlanib qoladi faqat qiymatlari alifbo tartibida numerical qiymatlarga o'zgartiriladi. Misol uchun tepadagi ranglarni son korinishiga olib otadi

```
from sklearn.preprocessing import LabelEncoder
```

- LabelEncoder funksiyasini sklearn.preprocessing shundan import qilib olamiz.

## data processing.

Drop - tashlab yuborish

Qator bo'yicha tashlab yuborish

`df.drop('size', axis=1, inplace=True)`

tashlab yubordih

Ustun bo'yicha tashlab

`df.dropna(inplace=True)`

## Data processing

missing

encoding

scaling

mean  
mode  
median  
freq  
drop

Basic

tashlab qoldirib  
qiymatlarni  
oldirish usuli

## Encoding

one hot coding

ham bir class  
ustun ustun  
yaratol

color

3 x 1 rang bar

demet  
class = 3

Red

Red

yellow

green

yellow

Red

gel

gr

1

0

0

1

0

0

0

1

0

0

0

1

## One Hot encoding

dummies pd.get\_dummies(  
df['size'], prefix='col', dtype  
e=int)

ustun nom

ustun  
debyoz  
ham botat

df = pd.concat([df.drop(  
columns=['size'], dummies],  
axis=1)

↓  
tashlab  
yuborid

`df = pd.concat(df_1, df_2)`

`df.unique()` = qilib  
olish kinish

Label encoder

darajalash uchun ishlatiladi

from sklearn.preprocessing  
import LabelEncoder

`encoder = LabelEncoder()`

`encoder`