

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

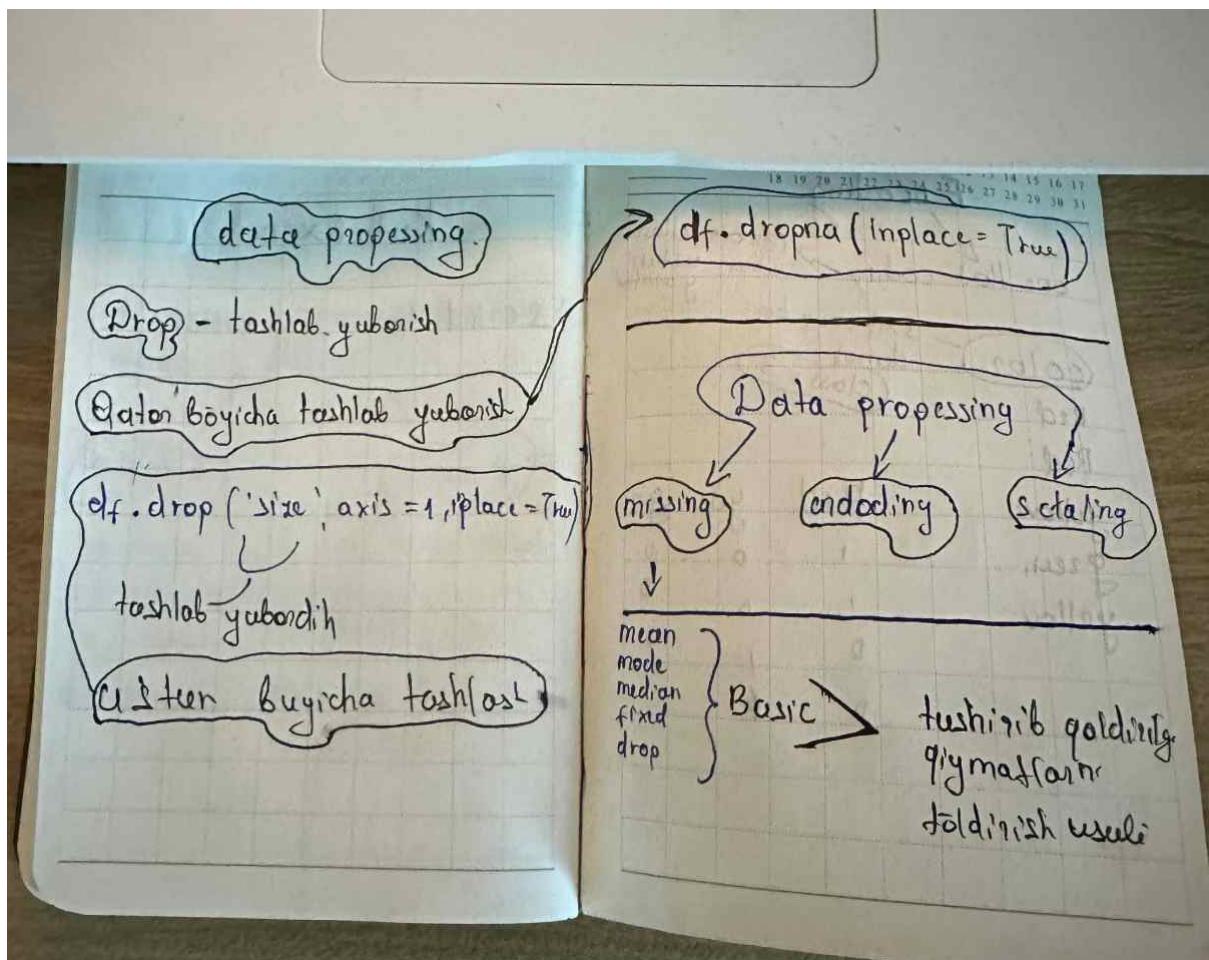
- one-hot encoding
- label encoding
- frequency encoding
- target encoding
- ordinal encoding

one-hot encoding => datasetdagi har bir classlar ucun 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.
```



Encoding

one hot coding → hanbit class
achter ustan yambala

color

Red

Red

yellow

green

yellow

3x1 zong bon
dummek
class = 3

	Red	yu	gr
Red	1	0	0
yellow	0	1	0
green	0	0	1
yellow	0	0	1

OneHot encoding

dummies pd.get_dummies(
df['size'], prefix='col', dtypes
e = int)
ustun nom,

ustun
debyori
ham botal

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

tasklab
yutapilot

$df = pd.concat([df_1, df_2])$

$df.unique()$ = qilib
olish uchun

Label encoding

darajalash uchun ishlash uchun

from sklearn.preprocessing
import LabelEncoder

encoder = LabelEncoder()
()

encoder