

# 17-dars (Report)

## Data Preprocessing:

### • Handling missing values

- mean
- mode
- median
- fixed
- drop

} Basic

### • Encoding (Sınıflaşım)

- One-hot
- Label Encoding

- Frequency
- Target
- Ordinal

### • Scaling

## \* Frequency encoding

- Tarroqlanıshlar sonini unutmaq elementler soniga nisbati bilan almastırılıadi.

Column	Animal	Animal - freq
Num	cat	(cat) 0.5
Num	cat	(cat) 0.5
Num	dog	(dog) 0.25
Num	crocodile	(crocodile) 0.25

(Ninə üçün) Chunki qatorlarsoni 4ta va cat 2ta, crocodile 1ta, dog 1ta telgəni  
mənənətə tutqulğan ; dog  $\rightarrow \frac{1}{4}$ ; crocodile  $\rightarrow \frac{1}{4}$   
cat  $\Rightarrow \frac{2}{4} \rightarrow$  qatorlarsoni

## \* Target encoding (Output)

Column	Animal	Column2
1.5	cat	0
3.6	cat	1
4.2	dog	1
7.1	crocodile	1

Column1	Animal	Column2
1.5	0.5	0
3.6	0.5	1
4.2	1	1
7.1	1	1

Target encodingda hərffli Columnda əylənməsi columnun hissələydi.

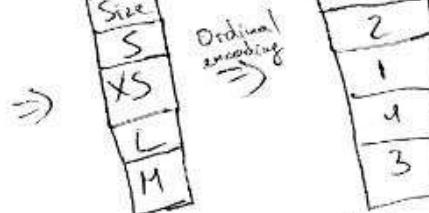
Animal	Column2
cat	0
cat	1
dog	1
crocodile	1

2ta catning rəparasidən gəlsək ( $0+1=1$ )  $\Rightarrow$  uni (2ta) katlaş  $\Rightarrow 0.5$   
dogning rəparasida (1) və dog (1) tələgi üçün  $\Rightarrow (1)$   
crocodile rəparasida (1) və crocodile (1) tələgi üçün  $\Rightarrow (1)$

## \* Ordinal encoding

- Ketmə sıtılığı nisba sənətlər bilan almastırılıadi. onu nüvviyyətli bəsləmə kətəməngə.

Column1	size	Column2
50	large	1.0
35	small	3.0
75	extra large	0.0
42	medium	2.0
54	large	1.0
71	extra large	0.0



Size sıtılığı kətəməngi:  
XS  $\rightarrow$  S  $\rightarrow$  M  $\rightarrow$  L  $\rightarrow$  XL(s)

LABEL  $\Rightarrow$  alfavit böyüməş repəntəltirildi  
ketmə sıtılığı yox

Ordinal  $\Rightarrow$  ketmə sıtılığı uzmaslık rəsəd.  
kətəməng  $\rightarrow$  kətəməng/yox / kətəməng  $\rightarrow$  kətəməng

Ita datasetda fəzət birləşdirilmiş kərəv.

# For loop

① Data types

② Operations

③ Conditions (if/else)

④ For loop (Datalar ait olmasi shart)  
- Jarayonni ait olmasi shart uchun ishlatalinadigan kodlar bloki.

a = [1, 2, 3]

print(a[0])

print(a[1])

print(a[2])

Value:

1  
2  
3

for i in [1, 2, 3]  
 print(i)

Value:

1  
2  
3

II) a = [1, 2, 3]

for i in a:  
 print(i)

1  
2  
3

for i in a:  
 print(a[i])

2  
3  
4

for element in a:  
 print(element)

1  
2  
3

b = ['Hello', 'Hi', 'Bye']

for i in b:  
 print(len(i))

5  
2  
3

len() → nekta harfdan iboratligini korish.

for i in range(1, 10):  
 (oralig) → 1 dan 9 gacha sonni oлади.

for i in range(1, 11):  
 (oralig) → 1 dan 10 gacha sonni oлади.

1) 1 dan 10 yurcha foydali surʼunni xiritish uchun?

for i in range(1, 11):

if i%2 == 0:  
 print(i)

2  
4  
6  
8

⇒

2) 1 dan 10 gacha surʼunni oлади:

for i in range(1, 11):  
 if i%2 == 0:  
 print(i)

1  
3  
5  
7  
9

3) for i in range(1, 11, 2):

print(i)

1  
3  
5  
7  
9

⇒ range(nima? oralig degan?)

range(start, stop, step)

1 dan 2 dan tashla deganmiz  
oldingi misolda:

# DATA PREPROCESSING + FOR LOOP

## 1) Missing + for loop

df.columns # үстендеңі барча үстүндер номи  
print(df.dtypes)

## ★ for col in df.columns:

```
if df[col].isnull().any():
    if df[col].dtype == 'string':
        df[col] = df[col].fillna(df[col].mode()[0], inplace=True)
```

```
else:
    df[col] = df[col].fillna(df[col].mean(), inplace=True)
```

⇒ Agar "mean"дан один төң (:) ishorasi  
білса "mean" пайтда "inplace=True" ғына сактама  
бләди.

## 2) Encoding + for loop

⇒ from sklearn.preprocessing import LabelEncoder

```
encoder = LabelEncoder()
```

```
encoder
```

```
df['Ustunnomi'] = encoder.fit_transform(df['Ustunnomi'])
```

## 3) df.unique

```
dummies = pd.get_dummies(df['Ustunnomi'], prefix='Ustunnomi', dtype=int)
```

dummies (Shift+Enter)

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

## ★ for loopنى encodingde табиги жиғіш

### for col in df.columns:

```
if df[col].dtype == 'string': # string yoki object болады
```

```
if df[col].nunique() <= 5:
```

```
    dummies = pd.get_dummies(df[col], prefix=col, dtype=int)
```

```
    df = pd.concat([df.drop(columns=[col]), dummies], axis=1)
```

```
else:
```

```
    df[col] = encoder.fit_transform(df[col])
```