

Data preprocessing bosqichilari

Datasetimizda keraksiz ustunlar bolsa ularni tushirib qoldirish kerak.

Datasetda tushurib qoldirilgan qiymatlarni toldirish.

- **MEAN** => orta arifmetik qiymat bilan toldiriladi (number)

- `df["ustun nomi"].fillna(df["ustun "].mean(), inplace=True)`

- **MODE** => eng kop takrorlangan qiymatlar bilan toldiriladi

- `df["ustun nomie"].fillna(df["ustun "].mode()[0], inplace=True)`

- **MEDIAN** => oraliqdagi qiymat bilan toldiriladi.

- `df["ustun nomi"].fillna(df["ustun "].median(), inplace=True)`

`df.describe()` => bu kod orqali number qiymatga ega ustunlarning **mean**, **min**, **max** qiymatlarini chiqarib beradi.

- **FIXED** => biz xoxlagan qiymat bilan

- `df["ustun nomi"].fillna("fixed qiymat", inplace=True)`

-

- `df.drop("ustun nomi", axis=1, inplace=True)` => ustunni tashlab yuborish

`axis=` - ustun bo'yicha tashlab yuborilayotganini anglatuvchi kod

`inplace=True` - ustunni tashlab yuborganimizdan so'ng

- `df.dropna(inplace=True)` Qatorda ma'lumotlar tushib qolgan bolsa o'sha qatorni tashlab yuboradi.

Library => Kodlar kutubxonasi

- Numpy,

- Pandas,

- Matplotlib,

- Seaborn,

- Scikit-learn => Eng katta kutubxona

`df['costumer_id']`

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

data preprocessing

- fushirib qoldirilgan qiymatlar ni toldirish

`df.isnull().sum()`

data preprocessing

1. mean (o'rtacha arifmetik qiymat)
2. mode (eng ko'p takrorlangan)
3. median (eng o'rtasidagi)
4. Fixed | O'zimiz xohlagan qiymat
5. tashlab yuborish

Mean

$$\begin{array}{r|l} \text{age} & \\ \hline 10 & \\ 20 & \\ \hline & \times 10 \end{array} \quad \begin{array}{l} 10+20 = \\ \hline 30 \\ \hline 3 \end{array}$$

Median (орта қымет)
numerical

```
# Median  
df['tenure'].fillna(df['tenure'].mode()[0],  
inplace = True)
```

```
df.isnull().sum()
```

```
df.describe()
```

statistika
ma'lumot

Fixed (to'g'ri)

```
# Fixed  
df['phone service'].fillna('yes',  
inplace = True)
```

Drop

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

qator.b

ustun.b.

drop

qatorlar bo'yicha

1-holat

```
df.dropna(inplace=True)
```

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

Kutubxona

scikit-learn - eng hattaki

Mode

age
 10
 20
 30
 10
 X
 mode
 eng hóp tashkilot
 gan qiymatlar
 df['dependents']
 fillna(df['dependents']
 .mean(), inplace=
 True)
 0
C2H1 + P2

Mean

df['senior citizen'].fillna(0)
 # mean
 df['senior citizen'].fillna(0).mean()
 df['senior citizen'].fillna(df['senior citizen'].mean(),)