

Data preprocessing bosqichilari

Datasetimizda keraksiz ustunlar bolsa ularni tushirib qoldirish kerak.

Datasetda tushirib 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