

4-Dars uchun report

Demak 4-darsda men ML project uchun muhim ketma-ketliklarni (1) qismini o'rgandim. Bularga; Data Collection (So'rovnoma, Webdan olish ya'ni web scraping qilib, tayyor datasetni topish yokida mijozdan data olish mumkin bo'ladi. Keyingi bosqichda men data bilan ishlash bilan tanishdim bular ichiga; fayl bilan datasetni bir xil joyga qo'yish, pandasni o'rnatish, pasdasi import qilish, datani o'ng tomonga olish, head/tail, input/output va keraksiz ma'lumotlarni tashlab yuborishlardan iborat.

Data preprocessingda biz asosan 5 ta usul orqali tushirib qoldirilgan qiymatlarni to'ldiramiz. Bular asosan: **Mean** (bu o'rta arifmetik qiymat bilan to'ldirish --**numerical** ustunlar uchun), **Mode** (eng ko'p takrorlanadigan elementlar bilan to'ldirish-- **numerical** va **categorical**), **Median**(o'rtasidagi elementlar bilan to'ldirish--**numerical**), Fixed qiymat bilan(o'zimiz xoxlagan qiymat bilan) va Tashlab yuborish usuli bilan (buni 2 hil usul bilan berish mumkin)

Endi esa ularning kodlarini keltirib o'taman:

```
# Tushib qolgan qiymatlarni mean orqali o'rta
arifmetigini chiqaramiz
# df.fillna({'Age': df['Age'].mean()}, inplace=True)
df['Age'].fillna(df['Age'].mean(), inplace=True)
```

```
df['Age'].fillna(df['Age'].mode()[0], inplace=True) # bu
kod ishlaydi lekin hozircha ishlamayotkani sababi bu
Ageda
# bironta ham kam qiymat mavjud emasligi sabaab agar
bitta qiymat bo'lganda ham uni 0 lashtirib olsa bo'lardi
```

```
df['ST_Slope'].fillna(df['ST_Slope'].median(),
inplace=True)
```

```
#df['Oldpeak'].fillna(df['Oldpeak'].median(),
inplace=True)
# bu usul bilan yozsa demak muammo bo'lmas eka pandas 3
da ishlar ekan
df.fillna({'Oldpeak': df['Oldpeak'].median()},
inplace=True)
```

```
df.describe() # Bu bizga qaysi ustunlar numerical
ekanligini ko'rsatib beradi.
```

Fixed

```
df['ExerciseAngina'].fillna('N', inplace=True) # manashu
fixed yani javobini aniq bir bilgan qiymat bilan
```

```
to'ldiriish yani birz bu qatorda tushib qolgan qiyatni  
oldindan bilar edik yoki bizda tajriba orqali ortirilgan  
bilimga asoslanib shu tushib qolgan qiymat bu edi deb  
berak bo'ladi.
```

```
# drop  
# qatorlar bo'yicha  
# 1-xolat
```

```
df.dropna(inplace=True) # agar barcha qatorlarda tushib  
qolgan qiymatlar bilinar bilinmas bo'lsa unda bu koddan  
foydalanamiz.
```

```
# drop malum bir ustudan juda kop qiymat tushib qolgan  
bolsa yani kamida 50% dan koproq
```

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

```
df.isnull().sum() # son orqali tushib qolgan qiymatlarni  
topish
```

Unda keyin esa, kutubxonlar (Library) larni ko'rib chiqdik. Bularga Numpy (matematik amallar ustida ish), Pandas(data manipulyatsiya), Matplotlib va Seaborn statistik ma'lumotlarni turli yo'llar bilan ko'rsatish (vizual ko'rnishda), Scikit-learn (ML o'qitish uchun va ML projectlarni bosqichlarini bajarish uchun--ya'ni tabir joyiz qilib aytganda Scikit-Learn bu Supervised MLning yuragi desak bo'ladi. Eng so'ngida esa Dataning structured xususiyatalri haqida ma'lumotga ega bo'ldik. Bular To'liqlik, Kattalik, Soxaga aloqdor bo'lishlik va Ishonchli bo'lishi lozimligi o'qtirildi.