

# 4-Lesson report

conditions If → THEN  
 misol uchun: yuqorida keltirilganmi.

• if

• elif

• else

4 - Dars report

Keraksiz ma'lumotlarni olib tashlash. Biz umuman katta qo'shma yolg'on features albatta. Endi keraksiz ustunlarni tashlab yuborish uchun bizga cod kerak: Ma'lum ustun-

ni belgilab olamiz va  
 codga copy qilamiz ⇒

#1) df.drop('ustun nomi', axis=1, inplace=True)

Bu codimiz istalgan feature ni "tashlab" yuborish uchun ishlatamiz. Endi ma'nolari

• df. → umumiy dataset

• drop → tashlab yuborish codi

• (''), → qo'shtiruvog' ichiga tashlab yuboriladigan ustun nomi

• axis=1, → ustun / qator bo'yicha aytish. Agar axis=0, balki pastga (qator bo'yicha),



- 8 agar aksis=1, bolsa o'nga  
9 (ustun bo'yicha) ishlatamiz.
- 10 • inplace = True → originali  
11 bilan joyini alishtirish  
12 yani biz co'loga buyruq  
13 berganimizdan keyin biz  
14 info qilsak biz bergan  
15 buyruq ishlagan va "info"  
16 da yog'olgan (tushib qol-  
17 dirilgan) ko'radi.

15 ML project kutma ketligi:

- 16 • Data preprocessing (ma'lumot  
17 tayyorlash)
- 18 • Model selection (model tanlash)

- 8 • Model training (data ni o'qi-  
9 tish) --- 70%
- 10 • Model prediction (bahorat  
11 qilish)
- 12 • Model evaluation (baholash)  
13 --- 15%
- 14 • Model testing (test qilish)  
15 --- 15%
- 16 • Model deployment
- 17
- 18 • Model maintaining / Monitoring



## Data preprocessing

Biz bu orqali datasetni kompyuter tushunadigan shaklda sozlash va chiroyli shaklda taqdim qilishimiz kk. Yana bu bosqich kompyuterga o'rgatishdan oldingi muhim jarayon hisoblanadi. Tayyorgarlik jarayoni hisoblanadi. Bizni bu bo'limda ko'rib chiqadigan muhim ishlarimiz bor ularni ko'rib chiqamiz.

1) Tushib qoldirilgan qiymatlarni to'ldirish.

• Mean (o'rtacha arifmetik qiymat b/n to'ldirish) (raqamli ustunlar)

• Mode (eng ko'p takrorlangan element b/n to'ldirish) (2 row)

• Median (eng o'rtaidagi element b/n to'ldirish) (raqamli)

• fixed qiymat b/n (O'zimiz xohlagan qiymat b/n to'ldirish)

• Tashlab yuborish

basic usulda bularni o'rganamiz !!!



Biz birinchi bo'lib ahami-  
yat beradigan joyimiz:

Tushib qolgan qiymatlarni  
toldirish. Chunki bizni  
datasetimiz ishlamasli-

gi. mumkin va bu qism  
Data preprocessingni "asosiy"

bo'limi hisoblanadi. Biz  
ketma-ketlikda qaramas-

dan shu qismini aytish kerak  
bo'ladi. Qiymatlarni aniqlash  
uchun yana tushib qolganlari-

ni biz bu cod dan use qila-

mi `# df.isnull().sum()`  
Biz tushib qolgan qiymatlar-  
ni toldirganimizdan keyin

keyingi bosqichga o'tamiz.

basic  $\Rightarrow$  statistik qiymatlar  
b'ni toldirish.

Mean

orta arifmetik qiymat b'ni  
toldiradi. for example:

$$\text{Age} \Rightarrow 10 / 20 / 10 \Rightarrow \frac{10+20+10}{3} = 10$$

Bularni biz qolota qilmaymiz  
bu orga jarayon. Mean  $\Rightarrow$   
ustundagi barcha qiymatlar-  
ni hisoblab orta arifmetik  
qiymatlar b'ni topadi, va  
keyingi bosqichga o'tsak  $\Rightarrow$



## Mode

Eng ko'p takrorlangan element b/n topdirish.  
for example:

Age (10 | 20 | 30 | 10 | 10)

bilsa bizda ko'p takrorlangan qiymat "10" va shu orqali kishub qoldirilgan qiymatni topdiradi. Endi Mean va Mode ni bir-biridan farqi 1 → arifmetik, 2 → eng ko'p takrorlangani b/n topdirayapti. Bizda faqat roqamli bilimgani uchun bizda biz muammo bor Numerical va categorical qiymat.

lar bor. Mean faqat (num)

Mode 2 xilani xam (num cat)

Biz kishub qoldirilgan qiymatlarni "Mean" b/n topdirish bizga shu cod tk.

# df['ustun'].fillna() →  
⇒ (df['ustun'].mean(), inplace=True)

codni kuzatish:

1) df['ustun'].fillna(df['ustun']). →

bu ma'lum ustunni o'zini ichidagi qiymati b/n topdirishni anglatadi.

2) mean(). → Mean usuli b/n topdirishimizni aytadi 1 va inplace=True ni qo'yib ketamiz bular fillna() ichida bo'layapti.



va qiymatni tekshirsaq nol  
yani to'ldirilgan bo'ladi;  
Mode => b/n to'ldirish. cod.  
df['ustun'].fillna(df['ustun'].mode()[0],  
inplace = True

## Median

Eng o'rta bo'lgan qiymat  
bilan to'ldiriladi. Such as,  
Name => Ali / (Vali) / X

cod:

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

Median de biz "numerical" dan  
foydalanamiz. yana biz cod:

# df.describe()

bu numerical qiymatlarni  
ko'rsatadi; va biz mean / mode,  
median qilla olamiz.

## Fixed

O'rimiz ko'rilgan qiymat b/n  
to'ldira olamiz. Biz haqiqi  
qatdan tushib qolgan qiymatlarni  
bilsak va u ro'k  
ekaniga ishonchimiz komil  
bilan ishlatishni maqul.

Code:

df['ustun'].fillna('Yes/No',  
inplace = True)



Drop

Tashlab yuborish. Bu 2 xil  
1 → ustun, 2 → qator ni tash-  
lab yuborish. Oqator bo'yicha  
1 - xolat

data setda qatorlar ko'p  
lekin tushib qolgan qiymat-  
lar juda oz bo'lsa biz  
bizdaniga xal qilsak bo'la-  
di yoki hama qatorlarda  
1 tadan qiymat qolib  
kelgan bo'lsa hammasi xal  
bo'ladi. Code:

`df.dropna(inplace=True)`

bu lekin data setimizda

10 ta column bo'lsa 9 ta  
qoladi hammasi qatorlardan  
1 ta dan qiymat olib  
ketadi!

Ustun bo'yicha:

Tushurib qoldirilgan qiymat-  
larni 50% dan ko'p bo-  
lsa ishlatamiz. code:

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

keyin 'info' qilsak hammasi  
xal bo'lib turadi.

Library

Biz code'larni qayta qayta  
yozmas logimiz uchun foyda-







8 xil yollar bñ kōrsatish  
9 (visual kōrinishda)

10 Scikit-learn

11 ML oqitish  
12 ML project bosqichlarini  
13 bajarish  
ML ni yaragi bu i

14 Data

15 Data:  
16 structured  
17 Toliq  
18 katta  
soxaga ologaden

8 ishonchli:  
9 bōlishi kerak.

10

11

12

13

14

15

16

17

18