4-dars (Report)

data frame

1) import pandas as pd
2) pip install pandas
3) df = pd.read_csv('Copy Path')

☆ df.drop('column nomi', axis=1, inplace=True)

ML project ketma ketligi (1)

- Data Collection (Malumot toplash)
  - Sorovnoma
  - Web sitelardan olish (web scraping)
  - Tayyor datasetni topish (github, Kaggle)
  - Misol
- Data bilan tanishuv
  - Faylva datasetni bir xil joyga qoyish (yonma-yon)
  - Pandas install
  - Pandas import
  - Datani ovg tomonga olish
  - head/tail
  - Input/output
  - Keraksiz ma'lumotlarni tashlab yuborish.

Keraksiz ustunlarni tashlab yuborish.

df['customerID'] ⇒ (df.drop('customerID')) ⇒ ma'lum bir qatorni tashlab yuborish u/n("axis=0")
axis ⇒ bu ustundan birini tashlab yuborish

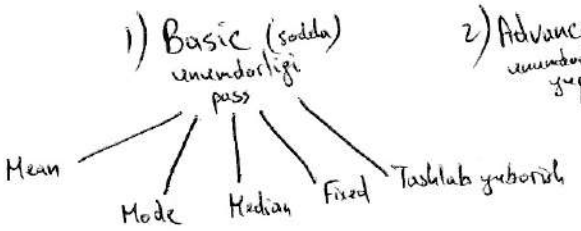
df.drop('customerID', axis=1, inplace=True)
⇒ (kodni) tepadagi originali bilan yangisini ozgartirish.
inplace → ni True qilmasak ozgarmasdan qolib ketadi.

ML project ketma ketligi (2)

- Data Preprocessing (Malumot tayyorlash)
- Model selection (Model tanlash)
- Model training (datani ogitish) -- 70%
- Model prediction (bashorat qilish)
- Model evaluation (Baxolash) -- 15%
- Model testing (test qilish) -- 15%
- Model deployment
- Model maintaining / Monitoring

Data Preprocessing.

1) Tushirib qoldirilgan qiymatlarni toldirish
   - Mean (orta arifmetik qiymat bn toldiring) (raqamli ustunlar uchun)
   - Mode (eng koptakrorlangan element bilan toldirish) (kategorial xarakterdagi)
   - Median (eng ortasidagi element bn toldirish) (raqamli)
   - Fixed qiymat bn (ozimiz xoxlagan qiymat bn toldirish)
   - Tashlab yuborish.



1) Mean:

	Age
x	10
y	20
10 ⇔ z	

orta arifmetik qiymat

2) Mode: eng kop takrorlangan son

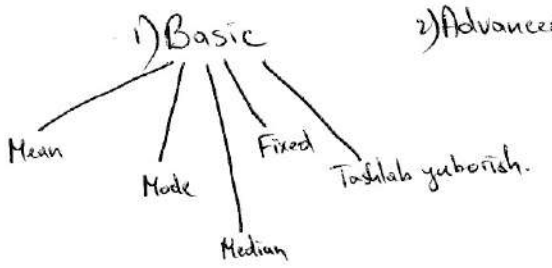
Age
10
20
30
10
10 ⇔ x

$\frac{x+y}{3} = z$

$\frac{10+20}{3} = 10$

Data Preprocessing:

ctrl + F2 ⇒ tanlanganni tanlab beradi.



1) Mean ⇒ o'rta arifmetik qiymat;
(Num)

Age
x = 10
y = 20
z = 10

⇒ $\frac{x+y}{3} = z$ ⇒ $\frac{10+20}{3} = 10$

#Mean. // o'zini ichidagi o'rta arifmetik 4/11

df['SeniorCitizen'].fillna(df['SeniorCitizen'].mean(), inplace=True)

2) Mode ⇒ eng ko'p takrorlangan qiymat:
(Num, Catogory)

Age
10
20
30
10
10 ⊘ x

1) #Mean (float) ⇒ df['Ustun nomi'].fillna(df['Ustun nomi'].mean(), inplace=True)

2) #Mode (object) ⇒ df['Ustun nomi2'].fillna(df['Ustunnomi2'].mode()[0], inplace=True)

// ifdagi [0] qo'shiladi.

// x=['ali', 'vali']
x[0]
// 'ali'

3) Median: ⇒ o'rtada turgan qiymat bilan to'ldiradi.

Name
Ali
Vali
Gani
Vali ⇔ x

(Num) statistic
~~x object x~~ ?
~~x Catogory x~~

#median
df['Ustunturi3'].fillna(df['Ustunnomi3'].median(), inplace=True)
df.isnull().sum()

#df.describe()

4) Fixed: ⇒ so'zni bilasiz, o'zingiz aniq bilib to'ldirish.

#Fixed
df['PhoneService'].fillna('Yes', inplace=True)

5) Drop: ⇒ tashlab yuborish. (qatorni tashlab yuborish) < qator bo'yicha / ustun bo'yicha
1) Qatorlar ko'p tushirib qo'yilgan qiymatlar juda oz: df.dropna(inplace=True) (hammasi 4/4)

5) Drop (Num, Cat)

Qator Ustun

5.1) #drop
#qatorlar bo'yicha.
#1-holat → qatorlari juda ko'p tushirib yuborilgan qiymat o'z.

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

#ustun bo'yicha, ma'lum bir ustunda juda ko'p (50%dan ko'p qiymatla tushib qolganda)

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

Library (Kutubxona)

Kutubxonalar:
- Numpy (Matematic statistic)
- Pandas (data)
- Matplotlib (grafik)
- Seaborn (documentation)
- Scikit-learn (Vis)
- Kutubxonalar (doc)

1) Pandas
Ma'lumotlarni kirgazish.
Ma'lumotlar bilan tanishish.
Ma'lumotlarni o'zgartirish, tozalash
Ma'lumotlarni taxlil qilish.

2) Numpy
Statistika va matematik boshqaruv
Matematik taxlil va xisob kitoblar

3) Matplotlib
grafiklar
Dizaynlar

4) Seaborn
Statistik ma'lumotlarni turli xil yo'llar bn ko'rsatish
(visual ko'rinishda)

5) Scikit-learn → Supervised ML yaratish
ML o'qitish
ML project bosqichlarini bajarish.

Data

Data:
Structured
To'liq
Katta
Soxaga aloqador
Ishonchli.