

8 xil yollar bñ k rsatish
9 (visual k rinishda)

10 Scikit-learn

11 ML oqitish
12 ML project bosqichlarini
13 bajarish
14 ML ni yuragi bu "

15 Data

16 Data:
structured

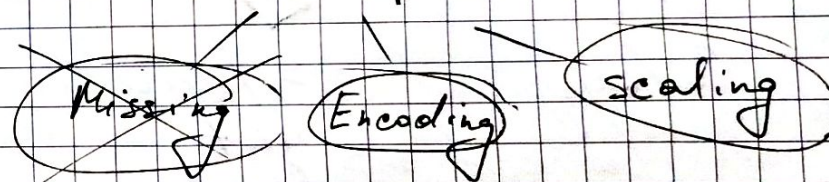
17 T liq
katta

18 soxaga ologaden

8 ishonchli:
9 b tishi kerak.

10 5-Dars report

11 Data pre.



14 Data preprocessing: Review

15 Handling missing values

• Mean

• Mode

• Median

• fixed

• Drop

Basic

Encoding
Scaling

Encoding

Categorical \Rightarrow Numerical
Object \Rightarrow int
float

Компьютер ишлаётган
t'isla qilish yoki o'qish
"Encoding" deyiladi.

Encodingni turlari:

1) One - Hot

2) Label Encoding

3) Frequency Encoding
4) Target Encoding
5) Ordinal Encoding

One - Hot Encoding

org
↓
Colour

one-hot encoding

Colour	Red	Yellow	Green
red	1	0	0
red	1	0	0
Yellow	0	1	0
Green	0	0	1
Yellow	0	1	0

Har 1 class uchun 1 ta
qizmat ishlatiladi. Taqat
0/1 qoyiladi. Lekin biz
har daim ham one-hot ni

ishlata olmaymiz. Bu dega-
ni boshqa encoding tur-
larini ham ishlatishimiz
kerak. Har bir kategoriya-
ga alohida ustun (column)
ochadi, va Model chalgash-
maydi. Oqcham ishlatamiz
Agar kategoriyalar tartibsiz
(nominal) bo'lsa. Masalan:
- Ranglar
- Davlatlar nomi
- Keshlar nomi

Asosiy farq:

- 1) Ko'p ustun
- 2) tartib yo'q

3) Ko'p joy ajallaydi
4) Nominal uchun yaxshi
biz ob'ektlarni "Type" ga
e'lib bor beramiz, va klasslar-
ga ham, klasslarni topish
uchun yana sonini biz bu
cod dan foydalanamiz =>
df.unique()

One-Hot encodingni ishlatish
uchun ustunlar => "dummies"
elementlar yaratib olishimiz
kerak! Cod:

dummies = pd.get_dummies(df['ustun'], prefix =
'col', dtype = int) keyin
dummies ni topamiz

8 endi original data ga
 9 o'tkazib olamiz. code:
 10 # df = pd.concat([df.drop(columns=
 11 => [df.drop(columns=['ustun']), dummies], axis=1)

Label encoding

13 Har bir kategoriyaga 1 ta
 14 raqam beradi. Misol:

Rang:

Qizil label → 0

16 Yashil → 1

17 Ko'k → 2

18 Bu yerda 2 > 1 > 0 deb o'ylashi
 mumkin. Yani =>

8 Ko'k > Yashil > Qizil degan
 9 noto'g'ri tartib paydo bo'la-
 10 di. Bu label encoding Alifbo
 11 tartibida bo'ladi, yani data-
 12 setni Alifbo tartibida to'g'ir-
 13 laydi. Birdaniga object dan
 14 numerical ga o'tkazib ola-
 15 di. Biz birinchi brincha partas-
 16 dan tashqari boshqa kutub-
 17 xona da ishlaymiz. code:
 18 # from sklearn.preprocessing import =>
 => LabelEncoder

encoder = LabelEncoder

18 # encoder deb saqlaymiz!

Biron ta data type ni o'z-
gartirish uchun cod!

df['ustun'] = encoder.fit_transform =>

=> (df['ustun'])

keyin "info" b'la tekshirib
olamiz fit_transform = dataset

ni ichidagi ma'lumotlarni o'z-

garigan holda shifro lar-
tibida belab joylashtirib
beradi!

Encoding

- Data ni 1 kundan 2-kunga
o'tkazadi!