

NaN

Thursday, November 28, 2024 7:52 PM

- Кога правиме Label encoding важно е да видиме дали некоја NaN вредност се вметнала во label

```
df = pd.read_csv('abalone.csv')
df.Type
```

```
0      M
1      M
2      F
3      M
4      I
..
4172   F
4173   M
4174   M
4175   F
4176   M
Name: Type, Length: 4177, dtype: object
```

```
y = encoder.fit_transform(df.Type)
```

```
df['Type'] = y
```

```
encoder.classes_
```

```
array(['F', 'I', 'M', nan], dtype=object)
```

- Encoder.classes_ ни враќа кои стрингови биле заменети соодветно со 0,1,2,3
Во случајов гледаме дека nan е вметнато во label=3 па мора да ги вратиме назад на nan

1) Пристап: replace()

```
import numpy as np

df['Type'] = df['Type'].replace(3,np.nan)
```

2) Пристап: рачно

```
tmp = []

for value in y:
    if value==3:
        tmp.append(np.nan)
    else:
        tmp.append(value)

df['Type'] = tmp
```