COLUMN TRANSFORMER

June 20, 2023

[2]:
```
import pandas as pd
import numpy as np
```

[5]:
```
df=pd.read_csv(r"C:\Users\Rakesh\Downloads\covid.csv")
df
```

[5]:

	age	gender	fever	cough	city	has_covid
0	60	Male	103.0	Mild	Kolkata	No
1	27	Male	100.0	Mild	Delhi	Yes
2	42	Male	101.0	Mild	Delhi	No
3	31	Female	98.0	Mild	Kolkata	No
4	65	Female	101.0	Mild	Mumbai	No
..	…	…	…	…	…	…
95	12	Female	104.0	Mild	Bangalore	No
96	51	Female	101.0	Strong	Kolkata	Yes
97	20	Female	101.0	Mild	Bangalore	No
98	5	Female	98.0	Strong	Mumbai	No
99	10	Female	98.0	Strong	Kolkata	Yes

[100 rows x 6 columns]

[8]:
```
df.isnull().sum()
```

[8]:
```
age           0
gender        0
fever        10
cough         0
city          0
has_covid     0
dtype: int64
```

[9]:
```
#gender-one hot encoding
#cough-ordinal encoding
#fever-simple inputer
#age-min max scaler(but not applying here in this problem)
```

[12]:
```
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OneHotEncoder
```

```
from sklearn.preprocessing import OrdinalEncoder
```

[14]:
```
si=SimpleImputer()
OHE=OneHotEncoder()
OE=OrdinalEncoder()
```

[18]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(df.
  ↪drop(['has_covid'],axis=1),df['has_covid'],test_size=0.2)
x_train
```

[18]:

	age	gender	fever	cough	city
61	81	Female	98.0	Strong	Mumbai
21	73	Male	98.0	Mild	Bangalore
52	47	Female	100.0	Strong	Bangalore
5	84	Female	NaN	Mild	Bangalore
92	82	Female	102.0	Strong	Kolkata
..	…	…	…	…	…
81	65	Male	99.0	Mild	Delhi
18	64	Female	98.0	Mild	Bangalore
23	80	Female	98.0	Mild	Delhi
73	34	Male	98.0	Strong	Kolkata
46	19	Female	101.0	Mild	Mumbai

[80 rows x 5 columns]

[21]:
```
x_train.shape
```

[21]:
```
(80, 5)
```

[24]:
```
from sklearn.compose import ColumnTransformer
ct = ColumnTransformer(transformers=[
    ('n1',SimpleImputer(),['fever']),
    ('n2',OneHotEncoder(),['gender']),
    ('n3',OrdinalEncoder(),['cough'])
],remainder='passthrough')
```

[26]:
```
ct.fit_transform(x_train).shape
```

[26]:
```
(80, 6)
```

[28]:
```
x_trainF1=ct.fit_transform(x_train)
x_trainF1
```

[28]:
```
array([[98.0, 1.0, 0.0, 1.0, 81, 'Mumbai'],
       [98.0, 0.0, 1.0, 0.0, 73, 'Bangalore'],
       [100.0, 1.0, 0.0, 1.0, 47, 'Bangalore'],
       [100.83561643835617, 1.0, 0.0, 0.0, 84, 'Bangalore'],
```

```
[102.0, 1.0, 0.0, 1.0, 82, 'Kolkata'],
[100.0, 0.0, 1.0, 0.0, 27, 'Kolkata'],
[98.0, 1.0, 0.0, 0.0, 26, 'Kolkata'],
[101.0, 1.0, 0.0, 0.0, 49, 'Delhi'],
[102.0, 0.0, 1.0, 0.0, 74, 'Mumbai'],
[104.0, 1.0, 0.0, 1.0, 54, 'Kolkata'],
[99.0, 1.0, 0.0, 1.0, 25, 'Kolkata'],
[104.0, 1.0, 0.0, 1.0, 75, 'Delhi'],
[104.0, 0.0, 1.0, 0.0, 51, 'Kolkata'],
[101.0, 1.0, 0.0, 1.0, 34, 'Delhi'],
[100.0, 1.0, 0.0, 1.0, 11, 'Kolkata'],
[99.0, 0.0, 1.0, 0.0, 65, 'Bangalore'],
[100.83561643835617, 0.0, 1.0, 0.0, 23, 'Mumbai'],
[102.0, 0.0, 1.0, 0.0, 64, 'Bangalore'],
[100.83561643835617, 1.0, 0.0, 1.0, 34, 'Mumbai'],
[98.0, 0.0, 1.0, 0.0, 83, 'Delhi'],
[101.0, 1.0, 0.0, 0.0, 8, 'Kolkata'],
[100.0, 1.0, 0.0, 1.0, 13, 'Kolkata'],
[99.0, 1.0, 0.0, 0.0, 14, 'Mumbai'],
[100.0, 0.0, 1.0, 0.0, 10, 'Bangalore'],
[100.0, 0.0, 1.0, 0.0, 27, 'Delhi'],
[102.0, 1.0, 0.0, 0.0, 49, 'Delhi'],
[104.0, 0.0, 1.0, 0.0, 42, 'Mumbai'],
[102.0, 1.0, 0.0, 0.0, 69, 'Bangalore'],
[100.83561643835617, 1.0, 0.0, 1.0, 42, 'Bangalore'],
[101.0, 1.0, 0.0, 1.0, 51, 'Kolkata'],
[103.0, 1.0, 0.0, 0.0, 73, 'Delhi'],
[103.0, 1.0, 0.0, 0.0, 50, 'Kolkata'],
[100.0, 0.0, 1.0, 0.0, 27, 'Delhi'],
[98.0, 1.0, 0.0, 1.0, 71, 'Kolkata'],
[104.0, 0.0, 1.0, 0.0, 25, 'Bangalore'],
[98.0, 0.0, 1.0, 1.0, 23, 'Mumbai'],
[100.0, 0.0, 1.0, 0.0, 11, 'Bangalore'],
[100.0, 0.0, 1.0, 0.0, 55, 'Kolkata'],
[103.0, 0.0, 1.0, 0.0, 83, 'Kolkata'],
[103.0, 0.0, 1.0, 1.0, 46, 'Bangalore'],
[99.0, 0.0, 1.0, 1.0, 66, 'Bangalore'],
[98.0, 1.0, 0.0, 1.0, 5, 'Mumbai'],
[104.0, 1.0, 0.0, 0.0, 12, 'Bangalore'],
[99.0, 1.0, 0.0, 0.0, 22, 'Bangalore'],
[100.0, 1.0, 0.0, 0.0, 5, 'Kolkata'],
[99.0, 1.0, 0.0, 1.0, 59, 'Delhi'],
[102.0, 0.0, 1.0, 1.0, 20, 'Delhi'],
[104.0, 1.0, 0.0, 0.0, 17, 'Kolkata'],
[103.0, 0.0, 1.0, 1.0, 70, 'Kolkata'],
[101.0, 1.0, 0.0, 0.0, 20, 'Bangalore'],
[100.0, 0.0, 1.0, 0.0, 80, 'Bangalore'],
```

```
[103.0, 1.0, 0.0, 0.0, 48, 'Kolkata'],
[101.0, 0.0, 1.0, 0.0, 15, 'Delhi'],
[102.0, 1.0, 0.0, 1.0, 24, 'Bangalore'],
[99.0, 1.0, 0.0, 0.0, 60, 'Mumbai'],
[101.0, 0.0, 1.0, 0.0, 42, 'Delhi'],
[100.83561643835617, 1.0, 0.0, 0.0, 75, 'Delhi'],
[98.0, 1.0, 0.0, 1.0, 10, 'Kolkata'],
[99.0, 1.0, 0.0, 1.0, 49, 'Bangalore'],
[101.0, 1.0, 0.0, 0.0, 65, 'Mumbai'],
[100.0, 1.0, 0.0, 0.0, 19, 'Kolkata'],
[101.0, 0.0, 1.0, 1.0, 14, 'Bangalore'],
[98.0, 1.0, 0.0, 1.0, 69, 'Mumbai'],
[104.0, 0.0, 1.0, 0.0, 16, 'Kolkata'],
[101.0, 1.0, 0.0, 0.0, 38, 'Bangalore'],
[100.83561643835617, 0.0, 1.0, 0.0, 38, 'Delhi'],
[100.83561643835617, 0.0, 1.0, 1.0, 71, 'Kolkata'],
[104.0, 1.0, 0.0, 0.0, 18, 'Bangalore'],
[98.0, 1.0, 0.0, 0.0, 65, 'Mumbai'],
[104.0, 1.0, 0.0, 1.0, 56, 'Bangalore'],
[103.0, 1.0, 0.0, 0.0, 69, 'Kolkata'],
[104.0, 0.0, 1.0, 0.0, 44, 'Mumbai'],
[104.0, 0.0, 1.0, 0.0, 51, 'Bangalore'],
[104.0, 1.0, 0.0, 0.0, 6, 'Kolkata'],
[98.0, 0.0, 1.0, 1.0, 12, 'Bangalore'],
[99.0, 0.0, 1.0, 0.0, 65, 'Delhi'],
[98.0, 1.0, 0.0, 0.0, 64, 'Bangalore'],
[98.0, 1.0, 0.0, 0.0, 80, 'Delhi'],
[98.0, 0.0, 1.0, 1.0, 34, 'Kolkata'],
[101.0, 1.0, 0.0, 0.0, 19, 'Mumbai']], dtype=object)
```

[ ]: