

feature engineering categorical data handling

May 10, 2021

1 Feature Engineering Handle the catogorical Data

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

[2]: df=pd.read_csv('train.csv')

[3]: df['Sex'].head()

[3]: 0    male
1    female
2    female
3    female
4     male
Name: Sex, dtype: object
```

2 pd.get_dummies

```
[ ]: #it only handles catogorical variables only
pd.get_dummies(df['Sex'],drop_first=True).head()
```

3 OneHot Encoding

```
[7]: #its helps to handle multiple catigorical variable
from sklearn.preprocessing import OneHotEncoder

[8]: oe=OneHotEncoder()

[ ]: oe.fit_transform(df['x_variable'].toarray)
```

4 LabelEncoder

```
[10]: #its helps when we deal with the multiple labels in variable  
from sklearn.preprocessing import LabelEncoder
```

```
[11]: le=LabelEncoder()
```

```
[14]: df['Sex']=le.fit_transform(df['Sex'])
```

```
[15]: df['Sex'].head()
```

```
[15]: 0    1  
      1    0  
      2    0  
      3    0  
      4    1  
      Name: Sex, dtype: int32
```

```
[ ]:
```