

## LabelEncoder and OneHotEncoder

LabelEncoder: `sklearn.preprocessing.LabelEncoder`

OneHotEncoder: `sklearn.preprocessing.OneHotEncoder`

These encoders are used to turn categorical values into numerical for models like regression models.

Example : The size of T-shirts might be (S,M,L,XL,XXL) with respect to other numeric attributes like cost, manufacturing date and so.

LabelEncoder encodes categorical value follows ordinality, example (S,M,L,XL,XXL) --> (0,1,2,3,4).

Whereas, OneHotEncoder encodes in binary i.e (0,1) based on the instance record.

For Example consider one instance of shirt dimension:

[24(length), 12(breadth), S(size)] --> [24, 12, 1, 0, 0, 0, 0]

[30, 18, M] --> [30, 18, 0, 1, 0, 0, 0], Similarly for others

Below are code snippets to explain the same.

In [1]:

```
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
import pandas as pd
X = [[12, 34, 45, "BAN"], [12, 33, 22, "IND"], [11, 11, 10, "PAK"]]
X = pd.DataFrame(X)
X1 = X.iloc[:, 3].values
X2 = X1
X1[:]
```

Out[1]:

```
array([[12, 34, 45, 'BAN'],
       [12, 33, 22, 'IND'],
       [11, 11, 10, 'PAK']], dtype=object)
```

In [2]:

```
labelencoder = LabelEncoder()
X1[:, 3] = labelencoder.fit_transform(X1[:, 3])
X1
```

Out[2]:

```
array([[12, 34, 45, 0],
       [12, 33, 22, 1],
       [11, 11, 10, 2]], dtype=object)
```