Preprocessing

LabelEncoder

- LabelEncoder() : Encode labels with value between 0 and n_classes-1.

In [1]:
```
from sklearn import preprocessing
import numpy as np
```

In [2]:
```
data = ["paris", "paris", "tokyo", "boston", "amsterdam", "boston"]
```

In [3]:
```
np.unique(data)
```

Out[3]:
```
array(['amsterdam', 'boston', 'paris', 'tokyo'], dtype='<U9')
```

In [4]:
```
le = preprocessing.LabelEncoder()
```

In [5]:
```
# Fit label encoder

le.fit(data)
```

Out[5]:
```
LabelEncoder()
```

In [6]:
```
le.classes_
```

Out[6]:
```
array(['amsterdam', 'boston', 'paris', 'tokyo'], dtype='<U9')
```

In [7]:
```
# Transform labels to normalized encoding.

le.transform(["tokyo", "tokyo", "paris", "boston"])
```

Out[7]:
```
array([3, 3, 2, 1])
```

In [8]:
```
# Transform labels back to original encoding

le.inverse_transform([3, 3, 2, 1])
```

Out[8]:
```
array(['tokyo', 'tokyo', 'paris', 'boston'], dtype='<U9')
```

In [9]:
```
# Fit label encoder and return encoded labels

le.fit_transform(data)
```

Out[9]:
```
array([2, 2, 3, 1, 0, 1])
```

In [10]:
```
data
```

Out[10]:
```
['paris', 'paris', 'tokyo', 'boston', 'amsterdam', 'boston']
```

In [ ]:
```

```