

Data Preprocessing

Handling missing values

- Mean

Basic

- Mode



- Median



- Fixed

Advanced

- Drop

- Encoding

- Scaling

Encoding

Transferring data from one type to another

1. One-hot

2. Label

Categorical-> Numerical

- Types

- One-Hot
- Label Encoding
- Frequency encoding
- Target Encoding

- Ordinal Encoding

One hot -Encoding

Labels	Sunny	Rain	Wind
Sunny	1	0	0
Rain	0	1	0
Wind	0	0	1

1 means yes it is sunny, rain or wind

0 means no it is not

Label encoding

Label Encoding : Label encoding converts 'red', 'green', 'blue' to '0', '1', '2'.

Notice that it uses only one column.

Color	Encoded_Value
Red	0
Green	1
Blue	2

Differences

Label Encoding			One Hot Encoding			
Food Name	Categorical #	Calories				
Apple	1	95				
Chicken	2	231				
Broccoli	3	50				
→			Apple	Chicken	Broccoli	Calories
			1	0	0	95
			0	1	0	231
			0	0	1	50

The diagram illustrates the conversion of a categorical 'Color' column into two different encoding formats: Label encoding and One-Hot encoding.

Original Data:

Color
Red
Green
Blue
Red
Green
Red
Blue

Label encoding: A vertical vector where each element corresponds to the color in the original list. Red is 1, Green is 0, and Blue is 2.

2
1
0
2
1
2
0

One-Hot encoding: A matrix where each row represents a color from the original list. The columns are labeled 'Dummy_r', 'Dummy_b', and 'Dummy_g'. A value of 1 indicates the presence of the color, while 0 indicates absence.

Dummy_r	Dummy_b	Dummy_g
1	0	0
0	0	1
0	1	0
1	0	0
0	0	1
1	0	0
0	1	0

 Levi
levikul09

For Loop

Used to automate the process