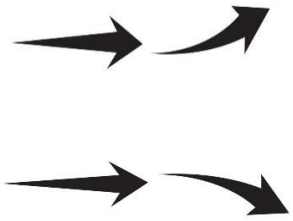
Data Preprocessing

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

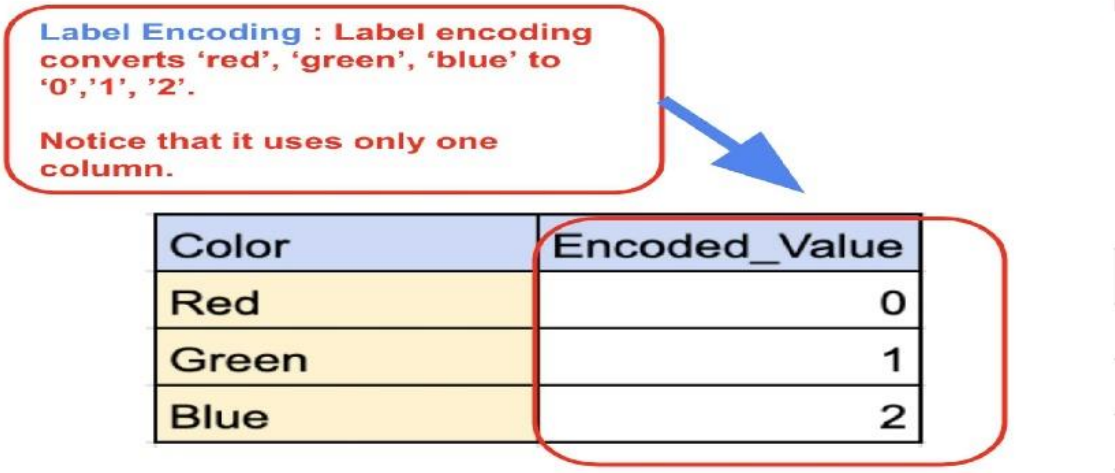
Labels
Sunny
Rain
Wind

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

1 means yes it is sunny, rain or wind

0 means no it is not

Label encoding



Differences

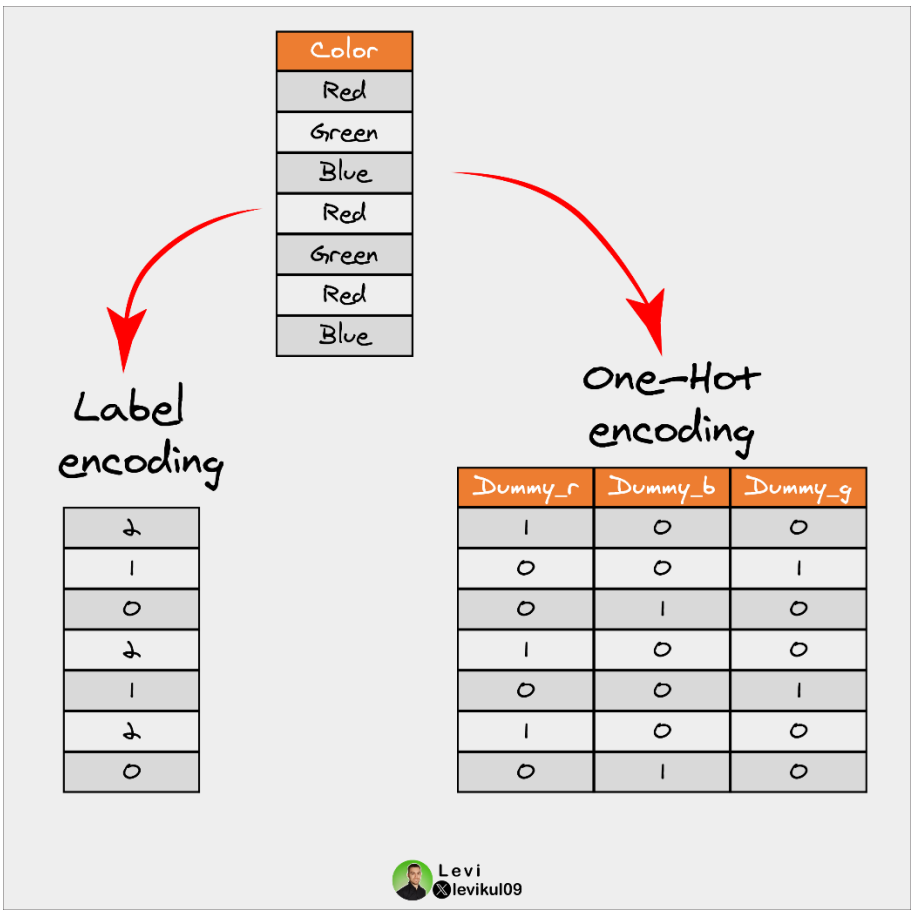
Label Encoding

Food Name	Categorical #	Calories
Apple	1	95
Chicken	2	231
Broccoli	3	50

→

One Hot Encoding

Apple	Chicken	Broccoli	Calories
1	0	0	95
0	1	0	231
0	0	1	50



For Loop

Used to automate the process