# The Difference between Label Encoding and One Hot Encoding

Our Machine only understands numbers and not text. Hence label encoding and one hot encoding come into the picture.

## Label Encoding –

This converts the objects of a column into numbers for the machine to understand. For example, let’s say we have a dataset where it’s given which people bought iPhones and which people didn’t buy. We can display the same as below,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Before Label Encoding** | |  | **After Label Encoding** | |
| **Name** | **Bought iPhone** | **Name** | **Bought iPhone** |
| Mr X | Yes | Mr X | 1 |
| Mr Y | No | Mr Y | 0 |
| Mr Z | Yes | Mr Z | 1 |

## One Hot Encoding –

In case a dataset available has regions in the dataset, if we use label encoding, for 4 regions our encoder will generate a value of 0,1,2,3,4. Which will eventually create a benchmarking for the regions for that the machine will think of the region with value 4 as more important than the region with value 0, which is not accurate. Hence we use a one-hot encoding. Which creates an additional column for each region and assign them with binary values for the machine to provide accurate prediction/classification

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Before One Hot Encoding** | | | |  |
| **Name** | | **Region** | |
| Mr. A | | North | |
| Mr. B | | South | |
| Mr. C | | East | |
| Mr. D | | West | |
|  | | | | |
| **After One Hot Encoding** | | | | |
| **Name** | **Region\_North** | **Region\_South** | **Region\_East** | **Region\_West** |
| Mr A | 1 | 0 | 0 | 0 |
| Mr B | 0 | 1 | 0 | 0 |
| Mr C | 0 | 0 | 1 | 0 |
| Mr D | 0 | 0 | 0 | 1 |