

Categorical data handling (Dummy variable creation)

eg: n

eid	ename	city
1	abc	Mumbai
2	pqr	BUR
3	svr	Mumbai
4	mno	Chennai

Algo of dummy variable:

- ① Take the unique values of the categorical column and represent the same in the form of list

['Mumbai', 'BUR', 'Chennai']

- ② Sort the list in ascending order.

['BUR', 'Chennai', 'Mumbai']
0 1 2

- ③ Replace the data of the categorical colⁿ with the index values from the list

Label Encoding *

eid	ename	city
1	abc	2
2	pqr	0
3	svr	2
4	mno	1

← The problem of this approach is mathematical weightage got introduced which is not appreciable! So we need a technique to remove mathematical weightage.

- ④ create Dummy Variable w.r.t Step 3.

	0	1	2	eid	ename
0	0	0	1	1	abc
1	1	0	0	2	pqr
0	0	0	1	3	svr
0	0	1	0	4	mno

} ONE-HOT ENCODING