

REGRESSION ASSIGNMENT

1) Identified problem statement:

Predict the Insurance charges for people with the given input.

2) Rows and Columns:

1338* 6

3) Preprocessing Method:

Categorical Data converted to numbers:

Column– Sex and Smoker=>Nominal => One hot
Encoding (from pd.get.dummies)

4) Developed Model with R2 score :

Attached with detailed documents below.

TO FIND BEST MODEL FROM R2 SCORE WITH HYPER PARAMETERS

1.MULTIPLE LINEAR REGRESSION

R2 Score

0.78

2.SUPPORT VECTOR MACHINE- REG

S.NO	HYPER.PARAMETER	LINEAR (R2)	RBF (R2)	POLY (R2)	SIGMOID (R2)
1	C =1.0	-0.01	-0.08	-0.07	-0.075
2	C=10	0.46	-0.03	0.038	0.039
3	C=100	0.62	0.32	0.61	0.52
4	C=1000	0.76	0.81	0.85	0.28
5	C=2000	0.74	0.85	0.86	-0.59

SVMR

R2 Score

0.86

(Poly,C=2000)

3.DECISION TREE-REG

S.NO	CRITERION	MAX-FEATURES	SPLITTER	R2 SCORE
1	squared_error	auto	best	0.67
2	squared_error	sqrt	best	0.75
3	squared_error	log2	best	0.73
4	squared_error	auto	random	0.73
5	squared_error	sqrt	random	0.67
6	squared_error	log2	random	0.64
7	friedman_mse	auto	best	0.72
8	friedman_mse	sqrt	best	0.70
9	friedman_mse	log2	best	0.56
10	friedman_mse	auto	random	0.68
11	friedman_mse	sqrt	random	0.68
12	friedman_mse	log2	random	0.62
13	absolute_error	auto	best	0.68
14	absolute_error	sqrt	best	0.71
15	absolute_error	log2	best	0.72
16	absolute_error	auto	random	0.71
17	absolute_error	sqrt	random	0.66
18	absolute_error	log2	random	0.7
19	poisson	auto	best	0.69
20	poisson	sqrt	best	0.58
21	poisson	log2	best	0.63
22	poisson	auto	random	0.75
23	poisson	sqrt	random	0.61
24	poisson	log2	random	0.72

DTR

R2 Score

0.75

(cate:Squared_error, max-feature=sqrt, splitter= best)
(cate:poisson, max-feature= auto, splitter= random)

4.RANDOM FOREST-REG				
S.NO	CRITERION	MAX-FEATURES	n_estimators	R2 SCORE
1	<i>squared_error</i>	sqrt	100	0.87
2	<i>squared_error</i>	log2	100	0.86
3	<i>squared_error</i>	None	100	0.84
4	<i>absolute_error</i>	sqrt	100	0.87
5	<i>absolute_error</i>	log2	100	0.86
6	<i>absolute_error</i>	None	100	0.85
7	<i>friedman_mse</i>	sqrt	100	0.86
8	<i>friedman_mse</i>	log2	100	0.86
9	<i>friedman_mse</i>	None	100	0.85
10	<i>poisson</i>	sqrt	100	0.83
11	<i>poisson</i>	log2	100	0.83
12	<i>poisson</i>	None	100	0.83

RFR	R2 Score	0.87	(MAE,max-feature=sqrt,n-estimator=100)
-----	----------	------	--

BEST MODEL:			
1	Random Forest Regression	0.87(R2 Score)	(MAE,max-feature=sqrt, n-estimator=100)
2	Support Vector Machine Regression	0.86(R2 Score)	(Poly,C=2000)