

REGRESSION-ASSIGNMENT

1.Problem Statement:

By using the given dataset **Insurance_pre.csv** we are going to predict, when the customer needs insurance amount and we give the insurance charges for the needed customers.

2. Information:

The given dataset is in Excel sheet and it is number so we choose the domain Machine Learning. Input and output are also clear so we use Supervised Learning. Output (Insurance Charges)is in Numerical form so we use Regression.The dataset contains 1338 Rows and 6 Columns.

Machine Learning→**Supervised Learning**→**Regression**

3. Preprocessing Method:

In this dataset sex column and smoker column is in string type. This name will not be comparable so we use **nominal method** to change the string into numerical data.

4. Good Model:

I have used all the machine learning algorithm like Multiple Linear Regression, Support Vector Machine, Decision Tree and Random Forest. I got the best model creation in **Support Vector Machine** for this dataset.

5. Research Values:

MACHINE LEARNING-REGRESSION

1.Multiple Linear Regression

$r_score=0.7894790349867009$

2.Support Vector Machine

S.NO	HYPER PARAMETER	LINEAR (r value)	RBF(NON LINEAR) (r value)	POLY (r value)	SIGMOID (r value)
1	C10	0.462468414	-0.0322732939	0.0387162227	0.033071437
2	C100	0.628879285	0.32003178320	0.6179569621	0.527610354
3	C3000	0.741423659	0.86633939509	0.8598930084	-2.12441947
4	C4000	0.741419880	0.87174078697	0.8600049580	-5.5103335471
5	C5000	0.741417930	0.87477781749	0.8595656394	-7.530043237
6	C7000	0.741422348	0.87769236904	0.8596647468	-17.548070147
7	C10000	0.741423013	0.87799524262	0.8591715079	-34.151535978

$r_score=0.87799524262$

3.Decision Tree

S.NO	CRITERION	MAN FEATURES	SPLITTER	r_VALUE
1	mse	auto	best	0.69022390
2	mse	auto	random	0.74844941
3	mse	sqrt	best	0.54871640
4	mse	sqrt	random	0.66675629
5	mse	log2	best	0.73577218
6	mse	log2	random	0.66944211
7	mae	auto	best	0.68047480
8	mae	auto	random	0.71274367
9	mae	sqrt	best	0.66454910
10	mae	sqrt	random	0.64770358
11	mae	log2	best	0.72040898
12	mae	log2	random	0.69479663
13	friedman_mse	auto	best	0.69863903
14	friedman_mse	auto	random	0.67241959
15	friedman_mse	sqrt	best	0.71634865
16	friedman_mse	sqrt	random	0.71254463
17	friedman_mse	log2	best	0.65448535
18	friedman_mse	log2	random	0.73748759

r_score=0.74844941

4.Random Forest

r_score=0.8498823801108987