REGRESSION ASSIGNMENT 12/08/2023

1.problem statement:

The given insurance dataset is

machine learning :supervised learning:regression

2.given dataset of insurance prediction 1338rows *6 columns

3.preprocessing method

the given dataset of two input catagorical column, so its converted into numerical numbers

4. develop various model of machine learning of regression algorithm

Multiple Linear Regression

R Score Value =0.7894

SVM R score value of various parameters

SL NO	hyper parameter	linear r value	RBF R value	POLY R value	SIGMOID R value
1	C10	0.4624	-0.032	0.038	0.039
2	C100	0.6288	0.32	0.617	0.527
3	C500	0.7631	0.664	0.826	0.444
4	C1000	0.7649	0.81	0.8566	0.2874
5	**C2000**	0.744	0.8547	**0.86**	-0.5939
6	C3000	0.7414	0.8663	0.859	-2.124

R score value =0.116(before standardization)

R score value=-0.010(after standardization)

without input of 'C' VALUE

Best R Score value of SVM =0.86

(poly,c=2000)

DECISION TREE r score value of various parameters

SL NO	CRITERION	MAX FEATURES	SPLITTER	R VALUE
1	Mse	auto	best	0.6899
2	Mse	auto	random	0.6898
3	Mse	sqrt	best	0.6981
4	Mse	sqrt	random	0.6905
5	Mse	log2	best	0.6836
6	Mse	log2	random	0.6525
7	Mae	auto	best	0.6864
8	Mae	auto	random	0.7123
9	Mae	sqrt	best	0.6955
10	Mae	sqrt	random	0.6292

11	Mae	log2	best	0.6775
12	**Mae**	**log2**	**random**	**0.7492**
13	Friendman_mse	auto	best	0.7099
14	Friendman_mse	auto	random	0.7463
15	Friendman_mse	sqrt	best	0.7012
16	Friendman_mse	sqrt	random	0.677
17	Friendman_mse	log2	best	0.6933
18	Friendman_mse	log2	random	0.7474

Best R score value of DECISION TREE=0.7492(mae,log2,random)

RANDOM FOREST

SL NO	CRITERION	MAX FEATURES	_ESTIMATOI	R VALUE
1	Mse	auto	10	0.8331
2	Mse	auto	100	0.8539
3	Mse	sqrt	10	0.852
4	Mse	sqrt	100	0.8709
5	Mse	log2	10	0.852
6	Mse	log2	100	0.8709
7	Mae	auto	10	0.8355
8	Mae	auto	100	0.8521
9	Mae	sqrt	10	0.8574
10	**Mae**	**sqrt**	**100**	**0.8717**
11	Mae	log2	10	0.8574
12	**Mae**	**log2**	**100**	**0.8717**

Best R score value of RANDOM FOREST=0.8717((mae,sqrt,estimator-100,mae) mae,log2,estimator100)

6.Final model :

Best model of this insurance dataset is **RANDOM FOREST**
because of High **R Score** value
0.8717 obtained
comparatively all other machine learning algorithm