R2_Score Values for file "insurance_pre,csv" :

Input:

      Age,bmi,children(0,1),sex_male(0,1),sex_female(0,1),smoker_no(0)
, smoker_yes(1).

Output :
      Charges.

Find the Method :
      1) Input is number .so Machine Learning
      2) Clear requirement to predict Charges .so Supervised learning
      3) Out put is not a catogorical data . so Regression
      4) Multiple inputs – Simple Linear Regression Not applicable.
      5) Models to check :
            a) Multiple Linear Regression
            b) Support Vector Machine
            c) Decision tree
            d) Random Forest

Method 1: Multiple Linear Regression: R2_score Value = 0.7894790349867009 = 0.7895
Method 2 : Support Vector Machine: R2_score Value = 0.8774793560307   = 0.8775

Support Vector Machine			R2_SCORE VALUE		0.8775	
S.NO.	**STANDARDIZATION**	**Hyper Parameter**	**R2_SCORE VALUE**			
			LINEAR	**POLY**	**RBF**	**SIGMOID**	**PRE COMPUTED**
1	NO	C0.01	-0.08	0.8255	-0.09	-0.0897	
2	NO	C10	0.1136	0.8627	-0.087	-0.0897	
3	NO	C1000	0.6893	-	-0.013	-0.0897	
4	NO	C2000	0.7651	-	0.0608	-0.0897	
5	NO	C3000	0.7649	-	0.1323	-0.0897	
6	NO	C4000	0.7442	-	0.1963	-0.0897	
7	NO	C10000	0.7417	-	0.4521	-0.0897	
8	NO	C20000	0.742	-	0.6406	-0.0897	X.shape[0] should be equal to X.shape[1]
9	NO	C50000	0.7425	-	0.741	-0.0897	
10	NO	C100000	0.7438	-	0.7247	-0.0897	
11	YES	C0.01	-0.088	-0.089	-0.09	-0.0895	
12	YES	C10	0.5665	0.1594	-0.018	0.07306	
13	YES	C100	0.636	0.7508	0.3906	0.52756	
14	YES	C1000	0.7441	0.8606	0.8284	0.14378	
15	YES	C2000	0.7414	0.8602	0.8607	-2.584	
16	YES	C3000	0.7414	0.86	0.8685	-6.8262	
17	YES	C4000	0.7414	0.8594	0.8707	-12.298	

18	YES	C10000	0.7414	0.8582	0.8775	–82.19	
19	YES	C20000	0.7414	0.8579	0.8772	-	
20	YES	C50000	0.7414	0.8578	0.8759	-	
21	YES	C100000	0.7414	0.8579	0.8735	-	

Method 3: Decision Tree:R2_score Value = 0.767056088869062   = 0.767

DECISION TREE		**R2_SCORE VALUE**		**0.7671**
S.NO.	**CRITERION**	**SPLITTER**	**MAX FEATURES**	**R2_SCORE**
1	mse	best	auto	0.6815
2	mse	best	sqrt	0.7564
3	mse	best	Log2	0.7221
4	mse	random	auto	0.7254
5	mse	random	sqrt	0.6787
6	mse	random	Log2	0.7846
7	mae	best	auto	0.6507
8	mae	best	sqrt	0.7365
9	mae	best	Log2	0.7190
10	mae	random	auto	0.7660
11	mae	random	sqrt	0.7284
12	mae	random	Log2	0.6496
13	Friedman_mse	best	auto	0.7013
14	Friedman_mse	best	sqrt	0.7671
15	Friedman_mse	best	Log2	0.6425
16	Friedman_mse	random	auto	0.7621
17	Friedman_mse	random	sqrt	0.6383
18	Friedman_mse	random	Log2	0.7494

Method4: Random Forest:

R2_score Value = 0.86988238146114= 0.8699

RandomForestRegression		R2_SCORE VALUE		0.8699
S.NO.	**CRITERION**	**n_estimators**	**MAX FEATURES**	**R2_SCORE**
1	mse	100	auto	0.8554
2	mse	100	sqrt	0.8699
3	mse	100	Log2	0.8699
4	mse	10	auto	0.8392
5	mse	10	sqrt	0.8570
6	mse	10	Log2	0.8570
7	mae	100	auto	0.8534
8	mae	100	sqrt	0.8670
9	mae	100	Log2	0.8670
10	mae	10	auto	0.8346
11	mae	10	sqrt	0.8350
12	mae	10	Log2	0.8350
13	Friedman_mse	100	auto	0.8549
14	Friedman_mse	100	sqrt	0.8681
15	Friedman_mse	100	Log2	0.8681
16	Friedman_mse	10	auto	0.8401
17	Friedman_mse	10	sqrt	0.8595
18	Friedman_mse	10	Log2	0.8595

Result :

      1) The best model Having High r2_score value (0.8775) is Known as

Support Vector Machine , Standardized model ,C=4000.

      2) Next to SVM model ,r2_score value (0.8699) is known as

      Random Forest Reegression,mse,100_estimators,Max_features (sqrt,log2)