Assignment - Baby Step 3

Requirement:
Predict the insurance charges.

1.Identify Problem statement:

Predict insurance charges

2.Tell basic info about dataset

Rows-1338, columns-6

3.Pre-processing method

Dataset has nominal dataset, using of get_dummies method, have changed to string to integer.

4. Models

4.1 Multiple Linear Regression
R^2 Score = 0.78947903498

4.2 Support Vector Machine

SL. No	Hyper Parameter	Linear	RBF (non-linear)-r value	POLY (r value)	SIGMOID (r value)
1	C=.1	-0.0809599	-0.089074	-0.0883023	-0.088269
2	C=10	0.4624684	0.3200317	0.038716	0.0393071
3	C=100	0.62887928	0.6642984	0.617956	0.5276103
4	C=500	0.763105	0.664298	0.8263683	0.4446061
5	C=1000	0.7649311	0.810206	0.8566487	0.287470
6	C=2000	0.744041	0.8547766	0.860557	-0.593950
7	C=3000	0.7414236	0.8663393	0.8598930	-2.12441
8	C=5000	0.7414179	0.874777	0.8595656	-7.530043

The SVM use R^2 value (rbf) and hyper parameter (C5000)) =0.874777

4.3 Decision Tree

SL. NO	CRITERION	MAX FEATURES	SPLITTER	R Value
1	*squared_error*	None	auto	0.6896639
2	*squared_error*	None	best	0.6802739
3	*squared_error*	None	random	0.7490586
4	*squared_error*	sqrt	auto	0.74516469
5	*squared_error*	sqrt	best	0.7721569
6	*squared_error*	sqrt	random	0.58916340
7	*squared_error*	Log2	auto	0.6622829
8	*squared_error*	Log2	best	0.70448358
9	*squared_error*	Log2	random	0.6620452
10	*friedman_mse*	None	auto	0.6867127
11	*friedman_mse*	None	best	0.688167
12	*friedman_mse*	None	random	0.7215396
13	*friedman_mse*	sqrt	auto	0.67276
14	*friedman_mse*	sqrt	best	0.704074
15	*friedman_mse*	sqrt	random	0.7841867
13	*friedman_mse*	Log2	auto	0.684577
14	*friedman_mse*	Log2	best	0.6563524
15	*friedman_mse*	Log2	random	0.738002
16	*absolute_error*	None	auto	0.68480736
17	*absolute_error*	None	best	0.68787262
18	*absolute_error*	None	random	0.6895067
19	*absolute_error*	sqrt	auto	0.720446
20	*absolute_error*	sqrt	best	0.693450
21	*absolute_error*	sqrt	random	0.7244735
22	*absolute_error*	Log2	auto	0. .68487
23	*absolute_error*	Log2	best	0.747989
24	*absolute_error*	Log2	random	0.541372
25	*poisson*	None	auto	0.7182793

26	*poisson*	None	best	0.71777
27	*poisson*	None	random	0.6977028
28	*poisson*	sqrt	auto	0.6813093
29	*poisson*	sqrt	best	0.6956314
30	*poisson*	sqrt	random	0.66000
31	*poisson*	Log2	auto	0.709384
32	*poisson*	Log2	best	0.79875
33	*poisson*	Log2	random	0.7335870

4.4 Random Forest

Random forest regression R^2 score(estimators=1000) = 0.8541778

I choose the best model as **Support Vector Regression** since it has **0.874777** r2_score value which is near to 1 compare with other models.