Problem Statement and Analysis:

We need to predict the Insurance charges based on input parameters.

The dataset consists of 5 inputs and 1 output. Also based on the input parameters we found this requirement can be done in **Machine Learning**.

We have the clear requirement and dataset with historical/expected output. So this can be done in **Supervised learning**.

Output must be given in numbers, then it is **Regression**.

Preprocessing:

Input columns contains two categorical values, those are converted to numeric data by using get dummies method of One Hot Encoding.

Algorithms:

We are using multiple Machine learning algorithms to find better performing model out of it.

Multiple Linear Regression

We have created a Multiple Linear Regression model , where R2_score is 0.789479034986701

No score changes even after standardization. This isn't an efficient model.

Support Vector Machine Regression

In SVM, we have created multiple models by varying Hyper tuning parameters. We will save the best SVMR model out these.

S.No	Kernel	C(penalty Parameter)	Epsilon	R2_score
	Linear	0.1	0.1	-0.122076
	Linear	1	1	-0.111542
	Linear	10	10	-0.001839
	Linear	100	100	0.541979
	rbf	0.1	0.1	-0.089576
	rbf	1	1	-0.088427
	rbf	10	10	-0.081938
	rbf	100	100	-0.123830
	Poly	0.1	0.1	-0.086252
	Poly	1	1	-0.064292
	Poly	10	10	-0.093201
	Poly	100	100	-0.099264
	Sigmoid	0.1	0.1	-0.089743
	Sigmoid	1	1	-0.089941
	Sigmoid	10	10	-0.090638
	Sigmoid	100	100	-0.116150

The model in SVMR gives maximum r2_score after Standardization is 0.6287726.

But this SVMR model isn't a good model for the given dataset. So lets move on with other models.

Decision Tree Regression

We have used standardization technique to get higher accuracy in the algorithm.

S.No	criterion	splitter	r2_score
1	Squared_error	Best	0.704172
2	Squared_error	Random	0.691498
3	Friedman_mse	Best	0.699620
4	Friedman_mse	Random	0.717591
5	Absolute_error	Best	0.685837
6	Absolute_error	Random	0.710194
7	poisson	best	0.733392
8	poisson	random	0.743699

The maximum score from Decision Tree Regression is 0.743699. This isn't a good model.

Random Forest Regression

We have used standardization technique to get higher accuracy in the algorithm.

S.No	Criterion	n_estimators	r2_score
1	Squared_error	50	0.851312
2	Squared_error	100	0.854504
3	Friedman_mse	50	0.845064
4	Friedman_mse	100	0.850692
5	Absolute_error	50	0.854263
6	Absolute_error	100	0.851109
7	poisson	50	0.851470
8	poisson	100	0.853640

There is an R2 score variation everytime we run the same model. The maximum score is 0.854504.

From this we can understand Random forest model predicts better than other model algorithm. We can consider this model for the deployment.