A client's requirement is, he wants to predict the insurance charges based on the several parameters. The Client has provided the dataset of the same.

1.) Identify your problem statement

- Insurance Charges Prediction
  Domain selection – stage1: Machine Learning (due to excel data set)
  Stage2: Supervised Learning (both input and output are present in dataset)
  Stage3: Regression (Output is charges value)

2.) Tell basic info about the dataset (Total number of rows, columns)

Rows = 1338

Columns = 6

3.) Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

Sex and smoker columns are nominal data, we have convert to binary data by using one hot encoding algorithm.

R2_score value Tabulation

1.Multiple Linear Regression

R2_score=0.7894790349867009

SVM: To found the best model in SVM algorithm by passing different parameter.

S.No	kernel	C	R2_score
1	rbf	0.001	-0.08969568284239293
3	rbf	100	-0.12480367775039669
4	rbf	10000	-0.017278762879702203
6	linear	0.001	-0.08788158489782139
7	linear	100	0.5432818196692804
8	linear	10000	0.7414230132360546
11	poly	0.001	-0.08966836238714304
12	poly	100	-0.09976172333666167
13	poly	10000	0.3529024061460666
16	sigmoid	0.001	-0.08966836238714304
17	sigmoid	100	-0.11814554828411405
18	sigmoid	10000	-119.51859539495995

Decision Tree:

To found the best model in decision tree algorithm by passing different parameter.

S.No	Criterion	Splitter	R2_score
1	squared_error	best	0.7111575033301749
2	squared_error	random	0.677332837047256
3	friedman_mse	best	0.7001826527448395
4	friedman_mse	random	0.7143057135504312
5	absolute_error	best	0.6660549033579489
6	absolute_error	random	0.7233912231246287
7	poisson	best	0.7259351571473129
8	poisson	random	0.7740597254172398

Random Forest:

To found the best model in random forest algorithm by passing different parameter.

S.No	Criterion	max_features	R2_score
1	squared_error	sqrt	0.8710271903471005
2	squared_error	log2	0.8710271903471005
3	squared_error	None	0.8538307913484513
4	absolute_error	sqrt	0.8710685856341518
5	absolute_error	log2	0.8710685856341518
6	absolute_error	None	0.8520093621081837
7	friedman_mse	sqrt	0.8710544015500664
8	friedman_mse	log2	0.8710544015500664
9	friedman_mse	None	0.8540518935149612
10	poisson	sqrt	0.8680156984764337
11	poisson	log2	0.8680156984764337
12	poisson	None	0.8526334258892607

All models R2_score:

S.NO	Algorithm	R2_Score
1	Multiple Linear Regression	0.7894790349867009
2	SVM	0.7414230132360546
3	Decision Tree	0.7740597254172398
4	Random Forest	0.8710685856341518

5.Best Model

Random Forest algorithm is best model because r2_score is validating parameter, which is closer to 1 in random forest compare to some other algorithm.