Machine Learning – Assignment

1. To predict the insurance chargers based on given parameters.

Stages: Machine Learning -> Supervised Learning as the input and output are clear -> Regression because the expected output is numeric.

2. Total no of Rows: 1338 Columns -6
3. Sex and Smoker columns are string so we need to convert them as numeric.
4. Finding best algorithm using R^2 value in Machine Learning:

I. Multiple Linear Regression - R^2 value = 0. 78947

II. Support Vector Machine:

S.No	Regularization parameter	kernel	R^2 value
1.	C1000	linear	0.76493
2.	C1000	rbf	0.81020
3.	C1000	poly	0.85664
4.	C1000	sigmoid	0.28747
5.	C5000	linear	0.74141
6.	C5000	rbf	0.87477
7.	C5000	poly	0.85956
8.	C5000	sigmoid	-7.5300
9.	C10000	linear	0.74142
10.	C10000	rbf	0.87799
11.	C10000	poly	0.85917
12.	C10000	sigmoid	-34.151

III. Decision Tree:

S.No	Criterion	Splitter	R^2 value
1.	friedman_mse	random	0.69402
2.	friedman_mse	best	0.69475
3.	squared_error	best	0.69380
4.	squared_error	random	0.72753
5.	absolute_error	random	0.71725
6.	absolute_error	best	0.68218
7.	poisson	best	0.71869
8.	poisson	random	0. 73484

iv. Random Forest:

S.No	Criterion	n_estimators	R^2 value
1.	friedman_mse	100	0.85398
2.	absolute_error	100	0.85259
3.	squared_error	100	0.85374
4.	poisson	100	0.85417
5.	friedman_mse	150	0.85284
6.	absolute_error	150	0.85231
7.	squared_error	150	0.85286
8.	poisson	150	0.85320

5. Uploaded code in repository.
6. Th best Model is SVM with C10000 and rbf parameter value and the R^2 value is 0.87799.