ASSIGNMENT- REGRESSION

1. PROBLEM STATEMENT-

Stage 1- Supervised learning

Stage 2- Machine learning

Stage 3- Regression

2. Total number of rows- 1338

Total number of columns- 6

3. Changed string to Number- Smoker, Sex

4. R2_score-Best model- Random forest

Machine learning regression- using R_Score valuation.

A. MULTI LINEAR REGRESSION- R2 Value- 0.7767

B. SUPPORT VECTOR MACHINE-

	R_SCORE	C=0.01	C=1	C=10	C=100
LINEAR	-0.1116	-0.0797	-0.1116	-0.0016	0.5432
RBF	0.0884	-0.0896	-0.0884	-0.0819	-0.1248
POLY	-0.0642	-0.0893	-0.0642	-0.0931	-0.0997
SIGMOID	-0.0899	-0.0897	-0.0899	-0.0907	-0.1181

C.DECISION TREE-

CRITERION	SPLITTER	R_SCORE
Squared error	Best	0.6895
Friedman_Mse	Best	0.7124
Absolute_error	Best	0.6732
Poisson	Best	0.7162
Squared error	Random	0.7294
Friedman_Mse	Random	0.6871
Absolute_error	Random	0.7264
Poisson	Random	0.6425

D.RANDOM FOREST-

N_ESTIMATORS	R_SCORE
10	0.8330
20	0.8468
30	0.8509
40	0.8539
50	0.8498

60	**0.8499**
70	**0.8520**
80	**0.8534**
90	**0.8534**
100	**0.8538**

As a result of comparison between above algorithms, the best model is considered as RANDOM FOREST for this particular dataset : 0.8539