Hope Artificial Intelligence
Assignment-Regression Algorithm

1. Problem Identification:
   Stage 1- Machine Learning
   Stage 2- Supervised Learning
   Stage 3- Regression

2. Total number of rows, columns: 1338 rows × 6 columns

3. Data pre-processing is used to convert the sex and smoker columns (nominal data) to numbers using get_dummies function in pandas.

4. Multiple Linear Regression, Support Vector Machine, Decision Tree and Random Forest algorithms are used to develop models for the given dataset.

5. R2 values from different algorithms are as tabulated below:
   Multiple Linear Regression: 0.789

   Support Vector Machine:

S.No	Standardisation	C	kernel	R2
1	Before	1	rbf	-0.089
2	After	1	rbf	-0.083
3	After	10	rbf	-0.032
4	After	100	rbf	0.320
5	After	1	poly	-0.075
6	After	10	poly	0.038
7	After	100	poly	0.617
8	After	1	sigmiod	-0.075
9	After	10	sigmiod	0.039
10	After	100	sigmiod	0.527

   Decision Tree:

S.No	criterion	splitter	R2
1	friedman_mse	best	0.694
2	friedman_mse	random	0.689
3	squared_error	best	0.693
4	squared_error	random	0.629
5	absolute_error	best	0.672
6	absolute_error	random	0.731

7	poisson	best	0.707
8	poisson	random	0.727

Random Forest:

S.No	criterion	n_estimators	max_features	R2
1	friedman_mse	10	sqrt	0.855
2	friedman_mse	100	sqrt	0.868
3	friedman_mse	10	log2	0.842
4	friedman_mse	100	log2	0.870
5	squared_error	10	sqrt	0.852
6	squared_error	100	sqrt	0.870
7	squared_error	10	log2	0.852
8	squared_error	100	log2	0.870
9	absolute_error	10	sqrt	0.847
10	absolute_error	100	sqrt	0.870
11	absolute_error	10	log2	0.851
12	absolute_error	100	log2	0.871
13	poisson	10	sqrt	0.844
14	poisson	100	sqrt	0.870
15	poisson	10	log2	0.862
16	poisson	100	log2	0.867

6. Final model would be **Random Forest** for the given dataset as the R-square value is nearer to 1 for this algorithm.