MACHINE LEARNING REGRESSION ASSIGNMENT

1. **Problem Statement:**

   a. Machine Learning
   b. Supervised Learing
   c. Regression

2. **Basic info of Dataset:**

   Inputs = Age, bmi, children, sex, smoker
   Outputs = Charges
   No of Rows = 1338
   No of Columns = 06

3. **Pre-Processing Method:**

In this Dataset Sex and Smoker column is a Categorical Data. Here this Sex and Smoker column considering as a comparing Data. So It is **Odinal data.**

4. **Final model:**

   Random Fores

5. **r2_score of the models:**

A).Multiple Linear Regression (R² value) = 0.789


```
Out[18]: array([-12057.244846])

In [19]: y_pred=regressor.predict(X_test)

In [20]: from sklearn.metrics import r2_score
         r_score=r2_score(y_test,y_pred)

In [21]: r_score
Out[21]: 0.7894790349867009
```

B).Support Vector Machine:

S.No	Hyper Tuning parameter	linear	rbf	poly	sigmoid
1	without Standardized	-0.111	-0.088	-0.064	-0.089
2	C 10	0.462	-0.03	0.038	0.039
3	C 100	0.628	0.32	0.617	0.527
4	C1000	0.764	0.81	0.856	0.287
5	C2000	0.744	0.854	0.86	-0.593

SVM linear (R² value)= 0.860

MACHINE LEARNING REGRESSION ASSIGNMENT

C).Decision Tree:

S.No	creterion	Max feature	spleter	r_score
1	Squared_error	auto	best	0.71
2	Squared_error	sqrt	best	0.652
3	Squared_error	log2	best	0.671
4	Squared_error	auto	random	0.714
5	Squared_error	sqrt	random	0.658
6	Squared_error	log2	random	0.713
7	friedman_man	auto	best	0.712
8	friedman_man	sqrt	best	0.708
9	friedman_man	log2	best	0.716
10	friedman_man	auto	random	0.768
11	friedman_man	sqrt	random	0.668
12	friedman_man	log2	random	0.695
13	absolute_error	auto	best	0.661
14	absolute_error	sqrt	best	0.729
15	absolute_error	log2	best	0.664
16	absolute_error	auto	random	0.752
17	absolute_error	sqrt	random	0.679
18	absolute_error	log2	random	0.677
19	poisson	auto	best	0.681
20	poisson	sqrt	best	0.51
21	poisson	log2	best	0.702
22	poisson	auto	random	0.7
23	poisson	sqrt	random	0.67
24	poisson	log2	random	0.7

DT (friedman_man,auto,random) **(R² value)= 0.768**

D).Random Forest

S.No	creterion	Max feature	n_estimators=100, r_score	n_estimators=200, r_score
1	Squared_error	sqrt	0.871	0.871
2	Squared_error	log2	0.871	0.871
3	friedman_man	sqrt	0.871	0.871
4	friedman_man	log2	0.871	0.871
5	absolute_error	sqrt	0.871	0.872

6	absolute_error	log2		0.871	0.871
7	poisson	sqrt		0.829	0.83
8	poisson	log2		0.829	0.83

Random Forest (R² value)= 0.871

6. Justification of Final model:

In this Requirement i created many models and R^2 values. MLR and Decision Tree not giving near 1. but SVM and Random forest giving little bit nearer to 1 compared to other. so i choosed Random forest is a good model for this requirement. Hence I taken into final model.