Problem Identification:

- **The i/p are <u>numerical</u> so Machine Learning**

- **The o/p is given so Supervised Learning**

- **The o/p value is <u>numerical</u> so regression**

1) Identify the problem statement

   To find the charges based on the several parameters

2) Tell the basic info about the dataset

   Input columns : age,sex,bmi,children,smoker

   Output columns: charges

3) Pre-processing

   The nominal data sex & smoker is converted

- Multiple Linear regression the R2 score is **78%**

svm

s.no	C parameter	linear r value	poly r value	rbf r value	Sigmoid r value
1.	C=1.0	−0.010102	−0.07569	−0.083382	−0.075429
	C=10	0.462468	0.03871	−0.032273	0.039307
	C=100	0.62887	0.617956	0.320031	0.527610354
	C=1000	0.7649311	0.85664	0.81020648	0.2874706
	C=10000	0.741423	0.85917	0.87799	−34.1515

- Svm when c=10000,rbf the r2 score is **87%**

Decision Tree

	criterion	splitter	r value
1	*squared_error (default)*	*Best(default)*	0.697361
		random	0.687923
2	*friedman_mse*	*best*	0.695353
		random	0.694586
3	*absolute_error*	*best*	0.707805
		random	0.72096
4	*poisson*	*best*	0.725017
		random	0.70378

- Decision tree poisson,best r2 value is 73%

Random Forest

s.no	criterion	n_estimators	r value
1	*squared_error*	10	0.83369
		50	0.85091
		100	0.85495
2	*absolute_error*	10	0.83628
		50	0.85412
		100	0.85318
3	*friedman_mse*	10	0.83373
		50	0.85111
		100	0.85111
4	*poisson*	10	0.83221

		50	0.85032
		100	0.85358

Random forest *absolute_error,50 r score is 85%.*

1.The best model svm 87%

2.random forest 85%