Problem Statement or Requirement:

1.) Identify your problem statement:

AI USED AT THE PREDICT THE INSURANCE CHARGES

2.)basic info about the dataset (Total number of rows, columns):

1338 rows , 6 columns

3.)Mention the pre-processing method if you're doing any (like converting string to number – nominal data)

	age	bmi	children	charges	sex_male	smoker_yes
0	19	27.900	0	16884.92400	0	1
1	18	33.770	1	1725.55230	1	0
2	28	33.000	3	4449.46200	1	0
3	33	22.705	0	21984.47061	1	0
4	32	28.880	0	3866.85520	1	0
...	...	...	...	...	...	...
1333	50	30.970	3	10600.54830	1	0
1334	18	31.920	0	2205.98080	0	0
1335	18	36.850	0	1629.83350	0	0
1336	21	25.800	0	2007.94500	0	0
1337	61	29.070	0	29141.36030	0	1

4.)Develop a good model with r2_score. You can use any machine learning algorithm; you can create many models. , you have to come up with final model

1.Multiple Linear Regression(R^2-value)=0.7894790349867009

2.Support vector Machine R^2[poly and Hyperparameter(1500)]=0.8580889

3.DECESION TREE (R^2-value)[friedman_mse,log2,best]=0.76619222

4.RANDOM FORESET(R^2-value)[friedman_mse,log2,100]=0.8744112

5.) All the research values (r2_score of the models) should be documented. (You can make tabulation or screenshot of the results.)

1.Multiple Linear Regression(R^2-value)=0.7894790349867009

2.Support vector Machine(R^2-value)

S.NO	HYPER PARAMETER VALUE(C)	LINEAR VALUE (R) V	RBF(non-linear value R)	POLY (R)	SIGMOID(R)
1	C=10	0.4624684	-0.0322732	0.0387162	0.0393071
2	C=100	0.6288792	0.3200317	0.6179569	0.5276103
3	C=500	0.7631057	0.6642984	0.8263683	0.4446061
4	C=1000	0.7649311	0.8102064	0.8566487	0.2874706
5	C=1500	0.7440487	0.8427494	0.8580889	-0.0674411

2.Support vector Machine R^2[poly and Hyperparameter(1500)]=0.8580889

DECESION TREE (R^2-value)

S.NO	CRITIERION	MAX FEATUREAS	SPLITTERS	R VALUE
1	Mse	Auto	Best	0.7000521
2	Mse	Auto	Random	0.6919682
3	Mse	Sqrt	Best	0.6110177
4	Mse	Sqrt	Random	0.7157955
5	Mse	log2	Best	0.7229416
6	Mse	log2	Random	0.7215323
7	Mae	auto	Best	0.6573389
8	Mae	auto	Random	0.7507558
9	Mae	Sqrt	Best	0.6848415
10	Mae	Sqrt	Random	0.7031684
11	Mae	log2	Best	0.6977008
12	Mae	log2	Random	0.676316
13	friedman_mse	auto	Best	0.7033522
14	friedman_mse	auto	Random	0.6861428
15	friedman_mse	Sqrt	Best	0.7519077
16	friedman_mse	Sqrt	Random	0.7088883
17	friedman_mse	log2	Best	0.76619222
18	friedman_mse	log2	Random	0.6962025

3.DECESION TREE (R^2-value)[friedman_mse,log2,best]=0.76619222

RANDOM FORESET (R^2-value)				
S.NO	CRITIERION	MAX FEATUREAS	N_ESTIMATORS	R VALUE
1	Mse	Auto	10	0.8326454
2	Mse	Auto	100	0.8560729
3	Mse	Sqrt	10	0.8405621
4	Mse	Sqrt	100	0.8713959
5	Mse	log2	10	0.8460456
6	Mse	log2	100	0.8744112
7	Mae	auto	10	0.8388896
8	Mae	auto	100	0.8527284
9	Mae	Sqrt	10	0.8559707
10	Mae	Sqrt	100	0.8723463
11	Mae	log2	10	0.8504588
12	Mae	log2	100	0.8723241
13	friedman_mse	auto	10	0.8247351
14	friedman_mse	auto	100	0.8545791
15	friedman_mse	Sqrt	10	0.8542056
16	friedman_mse	Sqrt	100	0.8691738
17	friedman_mse	log2	10	0.8612186
18	friedman_mse	log2	100	0.870177

4.RANDOM FORESET(R^2-value)[friedman_mse,log2,100]=0.8744112

ADA BOOST REGRESSOR(R^2-value)				
S.NO	N_ESTIMATERS	LOSS-LINEAR(R VALUE)	LOSS-SQUARE(R V	LOSS-EXPONENTIAL(R)
1	10	0.8566564	0.7101662	0.8254744
2	50	0.8512062	0.521828	0.6066836
3	100	0.8501721	0.4654836	0.537644
4	500	0.8687976	0.4531099	0.4751371
5	1000	0.8507999	0.4447981	0.4593156

ADA BOOST REGRESSOR(R^2-value)[N-EST=500,LOSS-LINEAR]=0.8687976

XGBOOSTINGREGRESSOR[GradientBoostingRegressor]((R^2-value)				
S.NO	CRITIERION	LOSS	N_ESTIMATOR	R VALUE
1	Mse	HUBER	100	0.89135
2	Mse	QUANTILE	100	0.63105
3	FRIEDMAN_MSE	HUBER	100	0.89147
4	FRIEDMAN_MSE	QUANTILE	100	0.63094

XGBOOSTINGREGRESSOR[GradientBoostingRegressor]((R 0.8914659

6.) Mention your final model, justify why u have chosen the same.

1.RANDOM FORESET

2.R^2-VALUE HIGH BUT COMPARISION ANOTHER MODEL