Assignment- Regression Algorithm

1. Problem statement identification:

   Stage1: ML
   Stage2: Supervised Leaning method
   Stage3: Regression

2. Dataset

   $1338 \text{ rows} \times 6 \text{ columns}$

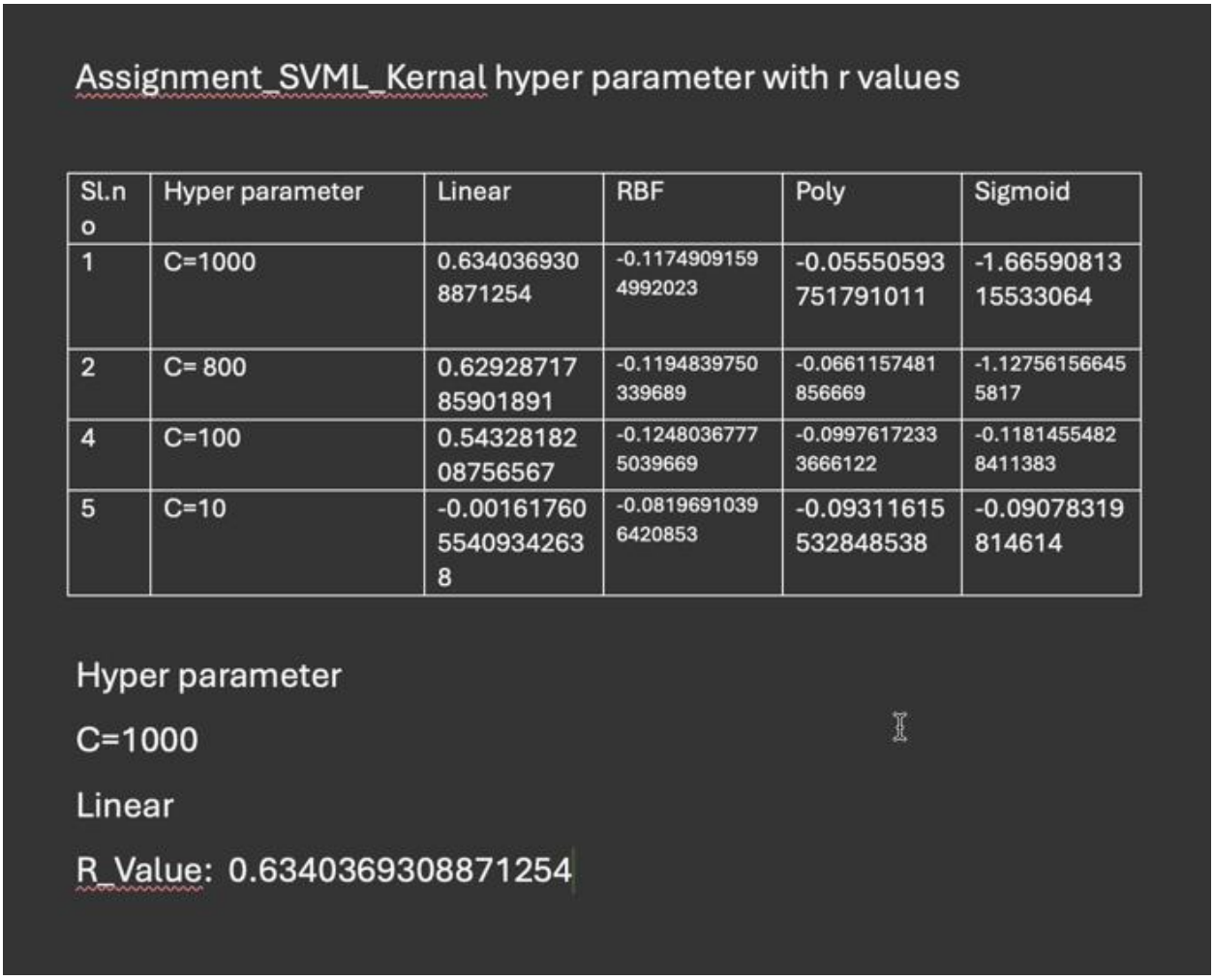
3. Yes, We need preprocessing method since we have nominal datas (Sex, Smoker) to convert String data into numeric.

4. R2_score values:

   1. Multiple Linear Regression:

      R2_value : 0.7894790349867009

2. Support Vector ML Regression



Assignment_SVML_Kernal hyper parameter with r values

Sl.no	Hyper parameter	Linear	RBF	Poly	Sigmoid
1	C=1000	0.6340369308871254	-0.11749091594992023	-0.05550593751791011	-1.6659081315533064
2	C= 800	0.6292871785901891	-0.1194839750339689	-0.0661157481856669	-1.127561566455817
4	C=100	0.5432818208756567	-0.12480367775039669	-0.09976172333666122	-0.11814554828411383
5	C=10	-0.001617605540934263 8	-0.08196910396420853	-0.09311615532848538	-0.09078319814614

Hyper parameter

C=1000

Linear

R_Value: 0.6340369308871254

kernel='linear'

C=1000

R2_Score value ： 0.6340369308871254

3. Decision Tree

Assignment: Decision tree Regression- Hyper parameters with R values

Sl.N0	Criteriaon	Max_features	Splitter	R_Value
1	Squared_error	Sqrt	best	0.6795533641359489
2	Squared_error	Sqrt	Random	0.6973832909326495
3	Squared_error	auto	best	0.6623132903803661
4	Squared_error	auto	Random	0.6980933984406023
5	Squared_error	log2	best	0.6398772872609122
6	Squared_error	log2	Random	0.6738577749631401
7	friedman_mse	Sqrt	best	0.7149923349432595
8	friedman_mse	Sqrt	Random	0.7076350768998017
9	friedman_mse	auto	best	0.658603507799159
10	friedman_mse	auto	Random	0.7193753715117888
11	friedman_mse	log2	best	0.6795796589117227
12	friedman_mse	log2	Random	0.7003518431318947

13	poisson	Sqrt	best	0.6370151046264717
14	poisson	Sqrt	Random	0.5985342184884453
15	poisson	log2	best	0.6597147233352301
16	poisson	log2	Random	0.6302975320496209
17	poisson	auto	best	0.6568278169253381
18	poisso	auto	random	0.7189475453842333
19	absolute_error	sqrt	best	0.7245625680549648
20	absolute_error	sqrt	Random	0.655196143279222

21	absolute_error	log2	best	0.6968008978205906
22	absolute_error	log2	random	0.5737609430307047
23	absolute_error	auto	best	0.668027408988 5903
24	absolute_error	auto	random	0.7241032787882982

criterion='absolute_error'

max_features='sqrt'

splitter='best'

R2_Value: 0.7245625680549648

4. Random forest:

Assignment:

Random Forest Regressor_hyper paramenter with R2_score values

Sl.No	n_estimators	criterion	max_features	C_Value	
1	100	-	-	0.8250491220 17278	
2	100	squared_error	sqrt	0.8388598173 145339	
3	50	squared_error	squrt	0.8369776106 713841	
4	100	squared_error	log2	0.8388598173 145339	
5	50	squared_error	log2	0.8369776106 713841	
6	100	friedman_mse	squrt	0.8390780475 542116	
7	100	friedman_mse	log2	0.8390780475 542116	

7	100	friedman_mse	log2	0.8390780475542116	
8	50	absolute_error	sqrt	0.843193121438361	Best Model
9	50	absolute_error	log2	0.843193121438361	Best Model
10	50	poisson	sqrt	0.8409579434651613	
11	50	poisson	log2	0.8409579434651613	
12	100	poisson	log2	0.8417099421816833	

n_estimators=50

random_state=0

criterion='absolute_error'

max_features='sqrt'

R2 value: 0.843193121438361

Conclusion:

In Random Forest supervised learning method, we are getting the best model out of all other methods. We can deploy in Production.