To predict insurance charges

Domain: Machine Learning

Goal: To predict insurance charges

Learning Selection: Requirement is to predict the insurance charges, and input and output are clear so this comes under supervised learning

Going to predict the insurance charges means going to predict numbers so it comes under ML regression

Project Name: To predict insurance charges

Stage 1: Going to process numbers so Domain is Machine Learning

Stage 2: Supervised Learning

Stage 3: ML Regression

Algorithms to try

1. Multiple Linear Regression
2. Decision Tree Regression
3. Support Vector Machine
4. Random Forest Regression

Actual Data with Nominal values

]:

	age	sex	bmi	children	smoker	charges
0	19	female	27.900	0	yes	16884.92400
1	18	male	33.770	1	no	1725.55230
2	28	male	33.000	3	no	4449.46200
3	33	male	22.705	0	no	21984.47061
4	32	male	28.880	0	no	3866.85520
...	...	...	...	...	...	...
1333	50	male	30.970	3	no	10600.54830
1334	18	female	31.920	0	no	2205.98080
1335	18	female	36.850	0	no	1629.83350
1336	21	female	25.800	0	no	2007.94500
1337	61	female	29.070	0	yes	29141.36030

1338 rows × 6 columns

Actual Data with numerical values

	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

1338 rows × 6 columns

Conclusion:

Best model is Random Forest and r_score value is 0.883475122

S. No	n_estimators	random_state	Criterion	Max features	max_depth	min_samples_split	min_samples_leaf	r score	Remarks
1	200	0	friedman_mse	sqrt	10	2	1	0.878726296	
2	200	0	friedman_mse	log2	10	2	1	0.878726296	
3	200	0	friedman_mse	sqrt	20	2	1	0.871251766	
4	200	0	friedman_mse	log2	20	2	1	0.871251766	
5	200	0	friedman_mse	sqrt	30	2	1	0.871210668	
6	200	0	friedman_mse	log2	10	5	2	0.882315637	
7	200	0	friedman_mse	sqrt	10	5	2	0.882315637	
8	200	0	friedman_mse	sqrt	20	5	2	0.882654802	
9	200	0	friedman_mse	log2	20	5	2	0.878726296	
10	200	0	friedman_mse	sqrt	30	5	2	0.882654802	
11	200	0	friedman_mse	log2	30	5	2	0.882654802	
12	200	0	friedman_mse	sqrt	10	10	4	0.881828635	
13	200	0	friedman_mse	log2	10	10	4	0.881828635	
14	200	0	friedman_mse	sqrt	30	10	4	0.881165423	
15	200	0	friedman_mse	log2	30	10	4	0.881165423	
16	200	0	friedman_mse	log2	20	10	4	0.881165423	
17	200	0	friedman_mse	sqrt	20	10	4	0.881165423	
18	200	0	friedman_mse	log2	30	5	4	0.883475122	
19	200	0	friedman_mse	sqrt	30	5	4	0.883475122	
20	200	0	friedman_mse	Log2	20	5	4	0.883475122	
21	200	0	friedman_mse	sqrt	20	5	4	0.883475122	
22	200	0	friedman_mse	log2	10	5	4	0.881792361	
23	200	0	friedman_mse	sqrt	10	5	4	0.881792361	
24	100	0	friedman_mse	sqrt	10	2	1	0.870945766	
25	100	0	friedman_mse	log2	10	2	1	0.878352695	
26	50	0	friedman_mse	sqrt	10	2	1	0.870044421	
27	50	0	friedman_mse	log2	10	2	1	0.870044421	
28			poisson						Keyerror
29			*squared_error*						Keyerror
30			*absolute_error*						Keyerror

Support Vector Machine

S. No	C	kernel	degree	gamma	shrinking	max_iter	r_score
1	100	linear	2	scale	TRUE	100	0.63526056
2	100	linear	2	scale	FALSE	100	0.63526056
3	100	rbf	2	scale	TRUE	100	0.056568869
4	100	rbf	2	scale	FALSE	100	0.056568869
5	100	poly	2	scale	TRUE	100	0.154013415
6	100	poly	2	scale	FALSE	100	0.154013415
7	100	linear	2	auto	TRUE	100	0.63526056
8	100	linear	2	auto	FALSE	100	0.63526056
9	100	rbf	2	auto	TRUE	100	0.056568869
10	100	rbf	2	auto	FALSE	100	0.056568869
11	100	poly	2	auto	TRUE	100	0.154013415
12	100	poly	2	auto	FALSE	100	0.154013415
13	100	linear	2	scale	TRUE	500	0.641045822
14	100	linear	2	scale	FALSE	500	0.641045822
15	100	rbf	2	scale	TRUE	500	0.311278356
16	100	rbf	2	scale	FALSE	500	0.311278356
17	100	poly	2	scale	TRUE	500	0.57997764
18	100	poly	2	scale	FALSE	500	0.57997764
19	100	linear	2	auto	TRUE	500	0.641045822
20	100	linear	2	auto	FALSE	500	0.641045822
21	100	rbf	2	auto	TRUE	500	0.311278356
22	100	rbf	2	auto	FALSE	500	0.311278356
23	100	poly	2	auto	TRUE	500	0.57997764
24	100	poly	2	auto	FALSE	500	0.57997764

2. Decision Tree

S. No	Criterion	Max features	splitter	r score
1	mse	auto	best	0.685020548
2	mse	auto	random	0.688248749
3	mse	sqrt	best	0.68980589
4	mse	sqrt	random	0.68980589
5	mse	Log2	best	0.682803776
6	mse	Log2	random	0.737662855
7	mae	auto	best	0.662194336
8	mae	auto	random	0.726132304
9	mae	sqrt	best	0.722891128
10	mae	sqrt	random	0.513354033
11	mae	Log2	best	0.640574025
12	mae	Log2	random	0.642391665
13	friedman_mse	auto	best	0.688859249
14	friedman_mse	auto	random	0.659013085
15	friedman_mse	sqrt	best	0.690095792
16	friedman_mse	sqrt	random	0.607810954
17	friedman_mse	Log2	best	0.764091927
18	friedman_mse	Log2	random	0.744298945

1. Multiple Linear Regression	R_score
MLR	0.789479035