

Problem Statement:

Insurance charges Prediction

Read the dataset and Retrieve the columns present in the dataset:

	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

Data set has 1338 rows ,and 6 columns

Note: By using shape attribute we can able to get the rows and columns count

ADA BOOSTING:

S.No	n_estimators	loss	random_state	r2_score
1	50	linear	None	0.86267
2	50	square	None	0.52248
3	50	exponential	None	0.62228
4	100	linear	None	0.86667
5	100	square	None	0.47199
6	100	exponential	None	0.55467
7	150	linear	None	0.83577
8	150	square	None	0.47054
9	150	exponential	None	0.52881
10	200	linear	None	0.85546
11	200	square	None	0.45811
12	200	exponential	None	0.50162
13	250	linear	None	0.84876
14	250	square	None	0.46563
15	250	exponential	None	0.49511
16	300	linear	None	0.86261
17	300	square	None	0.45602
18	300	exponential	None	0.48252
19	350	linear	None	0.85782
20	350	square	None	0.45996
21	350	exponential	None	0.48419
22	400	linear	None	0.84321
23	400	square	None	0.46432
24	400	exponential	None	0.48145
25	450	linear	None	0.83042
26	450	square	None	0.45134
27	450	exponential	None	0.47354
28	500	linear	None	0.86645
29	500	square	None	0.44602
30	500	exponential	None	0.47867



The ADA BOOST
regressor has the
r2_score= 0.86667

XGBOOST:

S.NO	n_estimators	alpha	gamma	reg_lambda	min_child_weight	subsample	eta	r2_score
1	100	0	0	0	0	0.5	0.1	0.83851
2	100	0	0	0	0	0.5	0.3	0.77882
3	100	0	0	0	0	0.5	0.2	0.79927
4	100	100	5	100	120	1	0.5	0.87977
5	100	200	5	200	120	1	0.5	0.87333
6	100	300	5	300	120	1	0.5	0.86191
7	100	400	5	400	120	1	0.5	0.85881
8	100	500	5	500	120	1	0.5	0.85147
9	100	1000	5	1000	120	1	0.5	0.829102
10	200	100	5	100	120	1	0.5	0.87858
11	200	200	5	200	120	1	0.5	0.88541
12	200	300	5	300	120	1	0.5	0.88282
13	200	400	5	400	120	1	0.5	0.88348
14	200	500	5	500	120	1	0.5	0.8774
15	200	1000	5	1000	120	1	0.5	0.85819
16	300	100	5	100	120	1	0.5	0.87547
17	300	200	5	200	120	1	0.5	0.88099
18	300	300	5	300	120	1	0.5	0.884
19	300	400	5	400	120	1	0.5	0.88622
20	300	500	5	500	120	1	0.5	0.88478
21	300	1000	5	1000	120	1	0.5	0.87201
22	400	100	5	100	120	1	0.5	0.8718
23	400	200	5	200	120	1	0.5	0.87915
24	400	300	5	300	120	1	0.5	0.88148
25	400	400	5	400	120	1	0.5	0.88337
26	400	500	5	500	120	1	0.5	0.88495
27	400	1000	5	1000	120	1	0.5	0.88054



The XGBOOST
regressor has the
r2_score= 0.88541

LGBM:

S.No	boosting_type	n_estimators	max_depth	num_leaves	learning_rate	random_state	r2_score
1	gbdt	100	15	10	0.1	None	0.86603
2	gbdt	100	30	20	0.1	None	0.87198
3	gbdt	100	45	30	0.1	None	0.8649
4	gbdt	100	60	40	0.1	None	0.86515
5	gbdt	200	15	10	0.1	None	0.87979
6	gbdt	200	30	20	0.1	None	0.86096
7	gbdt	200	45	30	0.1	None	0.85254
8	gbdt	200	60	40	0.1	None	0.84945
9	gbdt	300	15	10	0.1	None	0.87385
10	gbdt	300	30	20	0.1	None	0.85132
11	gbdt	300	45	30	0.1	None	0.84069
12	gbdt	300	60	40	0.1	None	0.83827
13	dart	100	15	10	0.1	None	0.87767
14	dart	100	30	20	0.1	None	0.87012
15	dart	100	45	30	0.1	None	0.86814
16	dart	100	60	40	0.1	None	0.86759
17	dart	200	15	10	0.1	None	0.88806
18	dart	200	30	20	0.1	None	0.88129
19	dart	200	45	30	0.1	None	0.87488
20	dart	200	60	40	0.1	None	0.87422
21	dart	300	15	10	0.1	None	0.88725
22	dart	300	30	20	0.1	None	0.8773
23	dart	300	45	30	0.1	None	0.86981
24	dart	300	60	40	0.1	None	0.86852

The LGBM regressor
has the r2_score=
0.88806

Conclusion:

- The Ada Boost , XGBoost, LGBM r2_score values is < than 0.90
- But comparing these three algorithm the XGBoost and LGBM algorithm performed good, The r2_score value is near 0.88. So that we can choose either anyone of them.