

Regression Assignment
Insurance Charges Prediction
Dataset:insurance_pre.csv

1.Problem Statement: Predict the insurance charges based on the parameters provided in the dataset.

2.Basic Info about Dataset: The dataset contains 1338 rows × 6 columns. The columns are ['age', 'sex', 'bmi', 'children', 'smoker', 'charges']

3.Pre-processing Method: The columns 'sex' and 'smoker' is a string categorical data. Hence using Ordinal Encoder we are converting string to number.

4.

Model : Multiple Linear Regression

r_score/accuracy is : 0.789479034986701

Model : Support Vector Machine

C value	Kernel <i>linear</i>	Kernel <i>poly</i>	Kernel <i>rbf</i>	Kernel <i>sigmoid</i>
0.01	-0.07976206842205191	-0.08930331146175452	-0.08969568284239293	-0.08971245575310527
0.1	-0.12207668380229886	-0.08625251710262294	-0.08957624598812952	-0.08974351910465961
1	-0.11166128719608448	-0.06429258402105531	-0.08842732776913875	-0.0899412170256757
10	-0.001617632488647213	-0.09311615532848516	-0.08196910396420853	-0.09078319814614
100	0.5432818196692804	-0.09976172333666167	-0.12480367775039669	-0.11814554828411405
1000	0.6340369312632063	-0.055505937517909665	-0.11749092439183229	-1.6659081315533064
3000	0.759089036774111	0.04892896389686585	-0.09621285083097098	-12.019048105850732

The most highest r2 score and accuracy is with c=3000,kernel=linear is 0.759089036774111

Model : Decision Tree

criterion	max_features	splitter	min_sample_split	max_depth	R2 Score
squared_error	None	best	3	2	0.8569635601474321
squared_error	None	best	10	2	0.8569635601474321
squared_error	None	random	10	2	0.7830284613940716
squared_error	None	random	3	2	0.7538171837959771
squared_error	sqrt	best	3	2	0.43133147427721696
squared_error	sqrt	best	10	2	0.1203259297863446
squared_error	sqrt	random	3	2	0.7671705086777151
squared_error	sqrt	random	10	2	0.6708866467634225
squared_error	log2	best	3	2	0.6751467884822075
squared_error	log2	best	10	2	0.7239504326343124
squared_error	log2	random	3	2	0.7513136703586799
squared_error	log2	random	10	2	0.4270787261079686

criterion	max_features	splitter	min_sample_split	max_depth	R2 Score
friedman_mse	None	best	3	2	0.8569635601474321
friedman_mse	None	best	5	2	0.8569635601474321
friedman_mse	None	random	3	2	0.7469673777081562
friedman_mse	None	random	7	4	0.8724621145030536
friedman_mse	sqrt	best	3	10	0.78175532466799
friedman_mse	sqrt	best	5	10	0.7909667944419413
friedman_mse	sqrt	random	3	10	0.7369044395699538
friedman_mse	sqrt	random	5	10	0.79604328654133
friedman_mse	log2	best	3	2	0.7513136703586799
friedman_mse	log2	best	5	2	0.44543266260573744
friedman_mse	log2	random	3	10	0.7795851157345066
friedman_mse	log2	random	4	10	0.8054612864790988

criterion	max_features	splitter	min_sample_split	max_depth	R2 Score
absolute_error	None	best	5	3	0.8679279967169236
absolute_error	None	best	5	4	0.8823500165656668
absolute_error	None	random	5	4	0.8615866553087528

absolute_error	None	random	4	6	0.8716488546132559
absolute_error	sqrt	best	4	4	0.8608170102197585
absolute_error	sqrt	best	5	5	0.8712629202707325
absolute_error	sqrt	random	3	10	0.8116508518684163
absolute_error	sqrt	random	5	10	0.841964998570084
absolute_error	log2	best	7	8	0.8479826058537915
absolute_error	log2	best	3	4	0.8718019982495898
absolute_error	log2	random	3	7	0.7547237065327816
absolute_error	log2	random	4	4	0.8108096684781412

criterion	max_features	splitter	min_sample_split	max_depth	R2 Score
poisson	None	best	7	4	0.884713903308847
poisson	None	best	4	7	0.8229327876970126
poisson	None	random	4	7	0.8632836562581366
poisson	None	random	3	6	0.8739421860539673
poisson	sqrt	best	5	5	0.8653850237395369
poisson	sqrt	best	5	7	0.8061288831546138
poisson	sqrt	random	5	7	0.8512259746260167
poisson	sqrt	random	5	10	0.7728797574287636
poisson	log2	best	7	6	0.8574298342328606
poisson	log2	best	3	5	0.8542570000558944
poisson	log2	random	6	7	0.7497301270204006
poisson	log2	random	7	6	0.7955622289236295

The highest accurate model is DecisionTreeRegressor(criterion='poisson', max_depth=5, max_features=None,min_samples_split=3,splitter='best') is
0.884713903308847

Model : Random Forest

n_estimators	Max_depth	Min_samples_leaf	R2 Score
100	10	2	0.8792115089484558
300	10	5	0.8872668462833726
300	5	5	0.8921108090344467
500	5	5	0.892074804534408
500	7	10	0.8893914852816728
1000	7	10	0.8891691066052605
1000	5	5	0.8924971855669555

5.Final Model

Random Forest with n_estimators=1000,max_depth=5,min_samples_leaf=5 with
accuracy 0.8924971855669555