

INSURANCE CHARGE PREDICTION

Baby steps-3

1. Stages

Machine learning
Supervised learning
Regression

2. Dataset details

Rows-1338
Columns- 6
Sex and smoker columns are categorical.
5 input variables.

3. Pre-processing and table creation

Converting them into nominal data using get_dummies from pandas.
After, Indep , dep is separated, Training and test set is created.

4. MODEL CREATIONS

- Multiple Linear Regression

R2 score=0.7894

- SVM

s.no	Hyper parameter C value	Linear	Poly	rbf	sigmoid	Precomputed
1	C10	-0.0016	-0.0931	-0.081	-0.0907	-
2	C100	0.5432	-0.0997	-0.124	-0.1181	-
3	C1000	0.6340	-0.0555	-0.117	-1.6659	-
4	C2500	0.7135	0.0234	-0.102	-8.4899	-

SVM R2 score=0.7135

- Decision tree

<input type="checkbox"/> criterion	splitter	max_features	R_score
squared_error	best	auto	0.693936183
squared_error	random	auto	0.68867211
squared_error	best	sqrt	0.754973374
squared_error	random	sqrt	0.609715771
squared_error	best	log2	0.735579151
squared_error	random	log2	0.678165532
friedman_mse	best	auto	0.694712031
friedman_mse	random	auto	0.67475693
friedman_mse	best	sqrt	0.664551385
friedman_mse	random	sqrt	0.702069718
friedman_mse	best	log2	0.658567844
friedman_mse	random	log2	0.680453145
absolute_error	best	auto	0.672856346
absolute_error	random	auto	0.745585477
absolute_error	best	sqrt	0.718793574
absolute_error	random	sqrt	0.666803184
absolute_error	best	log2	0.718946467
absolute_error	random	log2	0.676553716
poisson	best	auto	0.696393038
poisson	random	auto	0.690120097
poisson	best	sqrt	0.595131885
poisson	random	sqrt	0.633963344
poisson	best	log2	0.647422306
poisson	random	log2	0.644108831

Decision tree R2 score=0.754973374

- Random forest

s.no	N_estimators	R2-score
1	10	0.8338
2	50	0.8509
3	75	0.8537
4	100	0.8550

Random forest R2_score= 0.8550

- Comparing all the Evaluation metrics from all models, Random forest is chosen to be the best model due to its highest of 85.5%.

So, random forest model is saved and used in deployment.