

Regression Assignment

Usually a proper way to approach a dataset is by identifying the three stages of AI Prediction

1. Identify the Problem statement

There are three stages of AI Prediction

Stage 1:

Domain Selection: Machine Learning

Stage 2:

Learning: Supervised

Stage 3:

Classification: Regression

2. Information about the dataset

Insurance_pre.csv contains 1338 rows × 6 columns

3. Preprocessing Method

The Insurance_pre.csv has Nominal Data

4. r2 values for each algorithm and their respective max_features, C parameter

1. Multiple Linear Regression **r2 value () is 0.7894790349867009**

2. Support Vector Machine

serial.No	Hyper parameter	Linear R value	RBF(Non Linear R value)	Poly (r value)	Sigmoid (r value)
1	C10	0.46246841 4233968	-0.0322732 9390671052	0.0387162 227602313 45	0.0393071 437827434 7
2	C100	0.62887928 57320359	0.32003178 32050831	0.6179569 624059795	0.5276103 546510407
3	C500	0.76310579 7597537	0.66429846 45143137	0.8263683 54126896	0.4446061 033869479 5
4	C1000	0.76493117 38608445	0.81020648 51758545	0.8566487 675946551	0.2874706 948697669 5

5	C2000	0.74404183 08107487	0.85477664 2539298	0.8605579 27577388	-0.593950 973128350 8
6	C3000	0.74142365 99249815	0.86633939 53081686	0.8598930 084494408	-2.124419 478668986 3

Support Vector Machine best **r2 value** (**RBF(Non Linear)**, Hyperparameter (**C3000**)) is **0.8663393953081686**

3. Descision Tree

Serial.No	<i>criterion</i>	Max_features	Splitter	R_score Value
1	Mse	auto	best	0.6871945632 68898
2	Mse	auto	random	0.7058331807 027536
3	Mse	sqrt	best	0.5780449056 811394
4	Mse	Sqrt	random	0.6717749850 77185
5	Mse	Log2	best	0.6799833114 025472
6	Mse	Log2	random	0.6801072400 316461
7	Mae	auto	best	0.6648432062 010856
8	Mae	auto	random	0.7500551403 583591
9	Mae	sqrt	best	0.7339690095 918607
10	Mae	sqrt	random	0.6941169283 379627
11	Mae	Log2	best	0.5658811496 163172
12	Mae	Log2	random	0.7028754036 814628
13	<i>friedman_mse</i>	auto	best	0.7188171695 197177

14	<i>friedman_mse</i>	auto	random	0.6783110980 709615
15	<i>friedman_mse</i>	sqrt	best	0.7055531104 846458
16	<i>friedman_mse</i>	sqrt	random	0.5952600931 249437
17	<i>friedman_mse</i>	Log2	best	0.7963033250 752416
18	<i>friedman_mse</i>	Log2	random	0.6226609825 934093
19	<i>poisson</i>	auto	best	0.6792193646 949658
20	<i>poisson</i>	auto	random	0.6982855340 536165
21	<i>poisson</i>	sqrt	best	0.6106388856 607337
22	<i>poisson</i>	sqrt	random	0.6542909381 256115
23	<i>poisson</i>	Log2	best	0.5227100972 953048
24	<i>poisson</i>	Log2	random	0.6467486767 283614

Decision Tree best **r2 value** (criterion ***friedman_mse***), Max_features (Log2), splitter (Best))

is **0.7963033250752416**

4.Random Forest

Serial.No	<i>criterion</i>	Max_features	n_estimators	R_score Value
1	Mse	auto	10	0.8446897393 177992
2	Mse	auto	100	0.8582064953 815349
3	Mse	sqrt	10	0.8535877981 749798

4	Mse	sqrt	100	0.8722476146 127074
5	Mse	Log2	10	0.85841787167 08083
6	Mse	Log2	100	0.8706412205 393324
7	Mae	auto	10	0.8205612752 559868
8	Mae	auto	100	0.8501514295 69389
9	Mae	sqrt	10	0.8500750883 670041
10	Mae	sqrt	100	0.87174566434 95153
11	Mae	Log2	10	0.8543698071 729846
12	Mae	Log2	100	0.8736539550 478307

Random Forest best **r2 value** (*criterion (mae), Max_features (Log2), N_estimator is 100*)
is **0.8736539550478307**

The final Machine Learning Best Model of Regression for insurance dataset :

From implementing algorithm in order to find the best model .I found that both **Random Forest** and **Support Vector Machine** has best **R_score values**

1 . Random Forest best **r2 value** (*criterion (mae), Max_features (Log2), N_estimator is 100*)
is **0.8736539550478307**

or

2 . Support Vector Machine best **r2 value** (**RBF(Non Linear)** and HyperParameter(**C3000**)) is **0.8663393953081686**