DETERMINING THE BEST ALGORITHM WITH R2 METRICS FOR GIVEN DATASET

Name: E. Mariappan

The following shows the determination of best algorithm (best fit) for the given problem statement using the evaluation metrics r-square value.

The dataset is subjected to various algorithms such as linear regression, SVM and decision tree with varying parameters (for tuning the model).

1. Linear Regression

The following is evaluation metrics value obtained using Linear Regression Algorithm.

The evaluation metrics R^2 = 0.935868

2. SVM

The evaluation metrics value is observed using Support Vector Machine Algorithm.

S.No	kernel	citem	r_scpre
1	poly	1	0.050890
2	poly	100	0.465663
3	poly	200	0.578570
4	poly	300	0.588686
5	poly	400	0.616834
6	poly	500	0.620774
7	poly	600	0.624580
8	poly	700	0.630473
9	poly	800	0.634029
10	poly	900	0.637312
11	poly	1000	0.640324
12	poly	1100	0.643065
13	poly	1200	0.645534
14	poly	1300	0.647732
15	poly	1400	0.646930
16	poly	1500	0.648883
17	poly	1600	0.655675
18	poly	1700	0.661766
19	poly	1800	0.665255
20	poly	1900	0.668583
21	poly	2000	0.671748
22	poly	2100	0.674751

23	poly	2200	0.677591
24	poly	2300	0.680270
25	poly	2400	0.683152
26	poly	2500	0.699169
27	poly	2600	0.695718
28	poly	2700	0.691634
29	poly	2800	0.689933
30	poly	2900	0.690466
31	sigmoid	1	0.057499
32	sigmoid	100	0.058780
33	sigmoid	200	0.060080
34	sigmoid	300	0.061386
35	sigmoid	400	0.062698
36	sigmoid	500	0.064017
37	sigmoid	600	0.065341
38	sigmoid	700	0.066672
39	sigmoid	800	0.068009
40	sigmoid	900	0.069352
41	sigmoid	1000	0.070701
42	sigmoid	1100	0.072057
43	sigmoid	1200	0.073418
44	sigmoid	1300	0.074786
45	sigmoid	1400	0.076160
46	sigmoid	1500	0.077540
47	sigmoid	1600	0.078926
48	sigmoid	1700	0.080319
49	sigmoid	1800	0.081718
50	sigmoid	1900	0.083122
51	sigmoid	2000	0.084533
52	sigmoid	2100	0.085950
53	sigmoid	2200	0.087374
54	sigmoid	2300	0.088803
55	sigmoid	2400	0.090239
56	sigmoid	2500	0.091681
57	sigmoid	2600	0.093129
58	sigmoid	2700	0.094583
59	sigmoid	2800	0.096043
60	sigmoid	2900	0.097510
61	rbf	1	0.057317
62	rbf	100	0.030236

63	rbf	200	0.001162
64	rbf	300	0.023153
65	rbf	400	0.036504
66	rbf	500	0.050018
67	rbf	600	0.073677
68	rbf	700	0.096564
69	rbf	800	0.118681
70	rbf	900	0.140026
71	rbf	1000	0.160600
72	rbf	1100	0.180404
73	rbf	1200	0.199436
74	rbf	1300	0.217697
75	rbf	1400	0.231320
76	rbf	1500	0.236118
77	rbf	1600	0.246089
78	rbf	1700	0.257437
79	rbf	1800	0.268271
80	rbf	1900	0.278590
81	rbf	2000	0.288395
82	rbf	2100	0.293675
83	rbf	2200	0.298817
84	rbf	2300	0.304842
85	rbf	2400	0.320190
86	rbf	2500	0.334722
87	rbf	2600	0.348438
88	rbf	2700	0.361338
89	rbf	2800	0.373422
90	rbf	2900	0.384689
91	linear	1	0.895077

The evaluation metrics using SVM , R^2 = 0. 895077

3. Decision Tree

The same run is done for Decision Tree Algorithm.

S.No	criterion	splitter	max_features	r_sqr
0	poisson	random	log2	-0.331469
1	poisson	random	sqrt	0.562794
2	poisson	best	log2	-0.385106
3	poisson	best	sqrt	0.631192

4	friedman_mse	random	log2	-0.162306
5	friedman_mse	random	sqrt	0.848997
6	friedman_mse	best	log2	0.459646
7	friedman_mse	best	sqrt	0.561413
8	squared_error	random	log2	0.460638
9	squared_error	random	sqrt	0.450317
10	squared_error	best	log2	0.362213
11	squared_error	best	sqrt	-0.656995
12	absolute_error	random	log2	0.357039
13	absolute_error	random	sqrt	0.0146508
14	absolute_error	best	log2	0.861044
15	absolute_error	best	sqrt	0.656919

Best fit at [0.89805567] in the run- model combination is criterion:['squared_error'], splitter:['random'], max_features:['sqrt']

The evaluation metrics R^2 = 0. 861044

Linear Regression is one of the top models for the given dataset.