Start-Tech Academy

Classification

Method

1. Logistic Regression
2. Linear Discriminant Analysis
3. K nearest neighbor

Logistic Regression

price	crime_rate	resid_area	air_qual	room_num	age	dist1	dist2	dist3	dist4	teachers	poor_prop	airport	n_hos_beds	n_hot_rooms	waterbody	rainfall	bus_ter	parks
24	0.00632	32.31	0.538	6.575	65.2	4.35	3.81	4.18	4.01	24.7	4.98	YES	5.48	11.192	River	23	YES	0.04935
21.6	0.02731	37.07	0.469	6.421	78.9	4.99	4.7	5.12	5.06	22.2	9.14	NO	7.332	12.1728	Lake	42	YES	0.04615
34.7	0.02729	37.07	0.469	7.185	61.1	5.03	4.86	5.01	4.97	22.2	4.03	NO	7.394	101.12	None	38	YES	0.04576
33.4	0.03237	32.18	0.458	6.998	45.8	6.21	5.93	6.16	5.96	21.3	2.94	YES	9.268	11.2672	Lake	45	YES	0.04715
36.2	0.06905	32.18	0.458	7.147	54.2	6.16	5.86	6.37	5.86	21.3	5.33	NO	8.824	11.2896	Lake	55	YES	0.03947
28.7	0.02985	32.18	0.458	6.43	58.7	6.22	5.8	6.23	5.99	21.3	5.21	YES	7.174	14.2296	None	53	YES	0.04591
22.9	0.08829	37.87	0.524	6.012	66.6	5.87	5.47	5.7	5.2	24.8	12.43	YES	6.958	12.1832	River	41	YES	0.05217
22.1	0.14455	37.87	0.524	6.172	96.1	6.04	5.85	6.25	5.66	24.8	19.15	NO	5.842	12.1768	Lake	56	YES	0.05707
16.5	0.21124	37.87	0.524	5.631	100	6.18	5.85	6.3	6	24.8	29.93	YES	5.93	12.132	None	55	YES	0.0563
18.9	0.17004	37.87	0.524	6.004	85.9	6.67	6.55	6.85	6.29	24.8	17.1	YES	9.478	14.1512	River	45	YES	0.05073
15	0.22489	37.87	0.524	6.377	94.3	6.65	6.31	6.55	5.88	24.8	20.45	NO	6	11.12	Lake	29	YES	0.05778
18.9	0.11747	37.87	0.524	6.009	82.9	6.27	5.93	6.51	6.19	24.8	13.27	NO	9.278	13.1512	Lake and Riv	23	YES	0.05524
21.7	0.09378	37.87	0.524	5.889	39	5.76	5.14	5.58	5.33	24.8	15.71	YES	5.534	10.1736	Lake and Riv	57	YES	0.05742

Questions

Here are a few important questions that we might seek to address:

1. **Prediction Question**
   Will the house be sold within three months of getting listed
2. **Inferential Question**
   How accurately can we estimate the effect of each of the predictor variables on the response variable