

Metode Mencari Nilai K

Untuk mencari nilai K yang pas, diperlukan untuk menguji hasil dari prediksi KNN dengan nilai seharusnya. Karena file `DataTest_Tugas3.csv` nilai Y belum diketahui, maka akan digunakan sebagian dari data di file `DataTrain_Tugas3.csv`.

Persentase Akurasi

Untuk mengetahui persentase akurasi dari prediksi KNN. Hasil dari prediksi dibandingkan dengan hasil seharusnya. Lalu akurasinya dapat dilihat dengan menggunakan rumus $x/y * 100$

Dimana

X : Jumlah data yang benar

Y : Jumlah data yang harus di prediksi

Cross Validation

Cross Validation adalah sebuah teknik untuk validasi dan menilai dari hasil prediksi.

Cross validation akan memisahkan *Data Train* menjadi bagian-bagian kecil yang dinamakan *fold*. Jumlah dari *fold* yang dibuat dinamakan *k_fold*.

Selanjutnya akan dilakukan perulangan sebanyak *k_fold* yang akan menjalankan KNN dengan *k* tertentu. KNN dijalankan dengan *Data Test* berupa *fold* ke *i* dimana *i* adalah nilai perulangan saat ini. *Data Train* yang digunakan adalah gabungan dari semua *fold* tanpa *fold* ke *i*.

Hasil dari KNN tadi kemudian akan dihitung akurasinya.

Hasil dari Cross Validation adalah rata-rata dari akurasi yang didapat.

Pencarian Nilai K

Disini pencarian nilai K akan dilakukan dengan nilai K antara 1 sampai 50.

Metode Cross Validation akan digunakan untuk melihat tingkat akurasi dari K tersebut.

Nilai *k_fold* yang dipakai adalah 10.

Hasil Pencarian Nilai K

Berikut hasil pencarian nilai K dengan metode Cross Validation

K = 1 Accuracy = 82.5
K = 2 Accuracy = 82.5
K = 3 Accuracy = 83.875
K = 4 Accuracy = 84.375
K = 5 Accuracy = 84.375
K = 6 Accuracy = 85.25
K = 7 Accuracy = 85.875
K = 8 Accuracy = 85.75
K = 9 Accuracy = 86.5
K = 10 Accuracy = 86.25
K = 11 Accuracy = 85.875
K = 12 Accuracy = 85.5
K = 13 Accuracy = 85.375
K = 14 Accuracy = 85.625
K = 15 Accuracy = 85.875
K = 16 Accuracy = 85.0
K = 17 Accuracy = 85.25
K = 18 Accuracy = 85.125

K = 19 Accuracy = 85.125
K = 20 Accuracy = 84.875
K = 21 Accuracy = 84.375
K = 22 Accuracy = 84.0
K = 23 Accuracy = 84.5
K = 24 Accuracy = 83.75
K = 25 Accuracy = 84.375
K = 26 Accuracy = 83.875
K = 27 Accuracy = 83.25
K = 28 Accuracy = 83.75
K = 29 Accuracy = 83.25
K = 30 Accuracy = 83.375
K = 31 Accuracy = 83.0
K = 32 Accuracy = 82.875
K = 33 Accuracy = 82.625
K = 34 Accuracy = 82.25
K = 35 Accuracy = 82.5
K = 36 Accuracy = 82.5
K = 37 Accuracy = 81.75
K = 38 Accuracy = 81.5
K = 39 Accuracy = 81.625
K = 40 Accuracy = 81.5
K = 41 Accuracy = 81.875
K = 42 Accuracy = 81.625
K = 43 Accuracy = 81.875
K = 44 Accuracy = 81.875
K = 45 Accuracy = 81.5
K = 46 Accuracy = 81.25
K = 47 Accuracy = 81.375
K = 48 Accuracy = 81.5
K = 49 Accuracy = 81.375
K = 50 Accuracy = 81.125

Dari hasil diatas diketahui K terbaik adalah
9 dengan akurasi 86.5%

KNN Pada Test Data Dari Soal

Setelah K terbaik didapatkan prediksi pada
Test Data dapat dilakukan.

Parameter yang digunakan adalah:

K = 9

Hasil

Berikut hasil dari prediksi pada Test Data.

1	1	51	1	101	2	151	2
2	1	52	1	102	2	152	3
3	1	53	0	103	3	153	3
4	1	54	1	104	3	154	3
5	1	55	1	105	2	155	3
6	1	56	3	106	2	156	2
7	1	57	3	107	3	157	3
8	1	58	1	108	3	158	2
9	1	59	1	109	0	159	2
10	1	60	1	110	3	160	3
11	1	61	1	111	3	161	2
12	0	62	1	112	0	162	3
13	1	63	1	113	3	163	3
14	1	64	1	114	3	164	2
15	0	65	0	115	2	165	3
16	1	66	1	116	2	166	2
17	0	67	0	117	2	167	3
18	0	68	1	118	2	168	2
19	0	69	3	119	2	169	2
20	0	70	1	120	2	170	2
21	1	71	0	121	2	171	2
22	0	72	0	122	3	172	3
23	0	73	1	123	2	173	0
24	1	74	1	124	2	174	2
25	1	75	1	125	3	175	2
26	0	76	1	126	3	176	2
27	0	77	0	127	2	177	2
28	0	78	0	128	2	178	2
29	1	79	1	129	3	179	3
30	0	80	1	130	2	180	3
31	0	81	1	131	2	181	3
32	1	82	0	132	3	182	3
33	0	83	0	133	2	183	2
34	1	84	1	134	3	184	2
35	1	85	1	135	3	185	2
36	1	86	1	136	2	186	2
37	1	87	0	137	2	187	3
38	1	88	0	138	2	188	2
39	0	89	0	139	2	189	3
40	1	90	0	140	2	190	3
41	0	91	0	141	2	191	2
42	0	92	0	142	0	192	3
43	2	93	1	143	3	193	2
44	0	94	3	144	2	194	3
45	0	95	0	145	2	195	3
46	0	96	0	146	2	196	3
47	0	97	0	147	3	197	3
48	1	98	1	148	3	198	3
49	0	99	0	149	3	199	3
50	1	100	0	150	2	200	3