

Hendro Susanto (1301160284)

## Analisis Masalah

Terdapat 800 data train yang berisi 5 atribut input dan 1 output yang memiliki 4 label, lalu terdapat 200 data test yang berisi 5 atribut dan 1 output yang belum ditentukan labelnya. Akan ditentukan label dari setiap data test dengan membaca dari data train menggunakan metode k-Nearest Neighbors. 200 label dari data test akan dioutputkan berupa file TebakanTugas3.csv.

## Strategi Penyelesaian Masalah

Untuk menentukan label menggunakan metode KNN terdapat 4 step yaitu :

Step 1: Menentukan nilai K

Nilai k yang dipilih adalah 7. Karena nilai 7 paling optimal setelah dilakukan validasi dengan cara menguji data train sebanyak 240 dari 800(30%) dengan persentase akurasi cukup tinggi.

Step 2: Menghitung jarak setiap data

Nilai setiap data pada data test dihitung nilai euclidannya dengan seluruh data dari

data train dengan rumus : 
$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$
 . Maka akan terdapat 800 nilai yang berisi jarak antara 1 data test dengan 800 data train.

```
static double euclidan(String[] test, String[] train){
    return Math.sqrt(
        Math.pow( (Double.parseDouble(train[1])-Double.parseDouble(test[1])),2)+
        Math.pow( (Double.parseDouble(train[2])-Double.parseDouble(test[2])),2)+
        Math.pow( (Double.parseDouble(train[3])-Double.parseDouble(test[3])),2)+
        Math.pow( (Double.parseDouble(train[4])-Double.parseDouble(test[4])),2)+
        Math.pow( (Double.parseDouble(train[5])-Double.parseDouble(test[5])),2) );
}
```

Step 3: Memilih data terkecil sebanyak-k

Setiap mendapatkan nilai euclidan maka akan dicek apakah nilai baru lebih kecil dari nilai yang lama. Nilai yang lebih kecil akan disimpan beserta label dari data train sebanyak 7 data yang terpilih.

```
static double[][] check(double[][] kterdekat, double a, double b){
    int idmax=0;
    for(int i=0; i<kterdekat.length; i++){
        if(kterdekat[idmax][0]<kterdekat[i][0]) idmax=i;
    }
    if(kterdekat[idmax][0]>a){
        kterdekat[idmax][0]=a;
        kterdekat[idmax][1]=b;
    }
    return kterdekat;
}
```

Step 4 : Memilih label terbanyak diantara data sebanyak-k yang ada

Setelah didapatkan 7 data terkecil beserta labelnya maka akan dilakukan voting untuk menentukan label dari data test yang dicari. Label terbanyak diantara 7 data label akan dipilih sebagai label dari data test.

```
static double voting(double[][] kterdekat){
    double nol=0; double satu=0; double dua=0; double tiga=0;
    for (int i = 0; i < kterdekat.length; i++) {
        if (kterdekat[i][1]==0.0) nol++;
        else if (kterdekat[i][1]==1.0) satu++;
        else if (kterdekat[i][1]==2.0) dua++;
        else tiga++;
    }
    if (nol>=satu && nol>=dua && nol>=tiga) return 0;
    else if (satu>=nol && satu>=dua && satu>=tiga) return 1;
    else if (dua>=satu && dua>=nol && dua>=tiga) return 2;
    else return 3;
}
```

Setelah dilakukan 4 step diatas maka akan diperoleh 200 baris angka yang menyatakan label yang bersesuaian dari data test yang disimpan pada file TebakanTugas3.csv.

Screenshoot :

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