

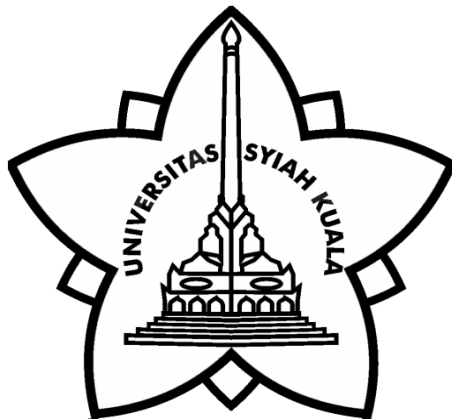
Tugas 4 Data Mining Pengukuran Jarak dalam Ruang Vector dan Metode Klasifikasi K-Nearest Neighbor (K-NN)

disusun untuk memenuhi
tugas mata kuliah Data Mining

oleh :

Khairul Umam Albi

2008107010072



**JURUSAN INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SYIAH KUALA
DARUSSALAM, BANDA ACEH**

2022

1. Dataset WDBC

Teknik mengubah data dari ekstensi lain menjadi data yang bertipe arff

- Hilangkan atribut ID sehingga tidak menjadi atribut dalam file ARFF (cek dalam dataset asli pada urutan berapa kolom ini berada).
- Letakkan class label (M atau B) pada kolom terakhir.
- Susun atribut lainnya dalam file ARFF sesuai urutan atribut pada data asli, hanya saja dalam file ARFF atribut ID dihilangkan dan atribut class label berada pada kolom terakhir.

a. File arff

```
File Edit Selection View Go Run Terminal Help
wdbc.arff
1 @RELATION wdbc
2
3 @attribute mean_radius real
4 @attribute mean_texture real
5 @attribute mean_perimeter real
6 @attribute mean_area real
7 @attribute mean_smoothness real
8 @attribute mean_compactness real
9 @attribute mean_concavity real
10 @attribute mean_concave_points real
11 @attribute mean_symmetry real
12 @attribute mean_fractal_dimension real
13 @attribute standart_error_radius real
14 @attribute standart_error_texture real
15 @attribute standart_error_perimeter real
16 @attribute standart_error_area real
17 @attribute standart_error_smoothness real
18 @attribute standart_error_compactness real
19 @attribute standart_error_concavity real
20 @attribute standart_error_concave_points real
21 @attribute standart_error_symmetry real
22 @attribute standart_error_fractal_dimension real
23 @attribute worst_radius real
24 @attribute worst_perimeter real
25 @attribute worst_area real
26 @attribute worst_smoothness real
27 @attribute worst_compactness real
28 @attribute worst_concavity real
29 @attribute worst_concave_points real
30 @attribute worst_symmetry real
31 @attribute worst_fractal_dimension real
32 @attribute Diagnosis (M, B)
33
34
35
36 17,99,10,30,122,0,1001,0,1184,0,2770,0,3001,0,1471,0,2419,0,07871,1,005,0,0053,0,509,113,4,0,000399,0,04004,0,05177,0,01507,0,03003,0,000103,25,30,17,33,104,6,2019,0,1622,0,4656,0,7119,0,2054,0,4601,0,1109,M
37 20,57,17,77,132,9,1326,0,08474,0,07864,0,0808,0,0781,0,1812,0,05667,0,5435,0,7339,1,302,74,00,0,00525,0,01308,0,0106,0,01309,0,00312,24,99,21,41,158,0,1956,0,1230,0,1806,0,2416,0,186,0,275,0,00002,M
38 19,49,21,70,136,103,0,1002,0,1596,0,1074,0,1279,0,2009,0,05099,0,7456,0,7009,0,583,94,03,0,00015,0,00000,0,03032,0,00004,0,0255,0,000513,23,57,25,53,153,5,1709,0,1444,0,4345,0,4504,0,243,0,303,0,00709,M
39 11,42,20,30,77,58,106,1,0,1425,0,1039,0,2414,0,1052,0,2597,0,07044,0,4056,1,156,3,445,27,23,0,00011,0,07458,0,05061,0,01007,0,05063,0,000200,14,91,26,5,90,0,07567,0,0,2000,0,0663,0,6809,0,2575,0,6630,0,173,M
40 20,29,14,14,115,1,1297,0,1003,0,1120,0,100,0,1063,0,1009,0,05081,0,7977,0,7033,1,434,94,44,0,01149,0,02463,0,05040,0,01009,0,01704,0,00115,22,54,14,47,152,2,1575,0,1170,0,200,0,4,0,1625,0,2304,0,07670,M
41 12,45,17,7,83,57,47,1,0,1078,0,117,0,1570,0,00009,0,2087,0,07013,0,1345,0,0002,2,227,27,35,0,00793,0,01345,0,03077,0,0117,0,02345,0,005002,15,47,23,75,103,4,741,6,0,1793,0,5249,0,5355,0,1343,0,3005,0,1244,M
42 10,25,19,50,110,6,1040,0,00063,0,100,0,1127,0,074,0,1794,0,05742,0,4467,0,7732,3,10,53,01,0,000314,0,01302,0,02254,0,01039,0,01309,0,002179,22,00,27,66,153,2,1006,0,1442,0,2576,0,3704,0,1932,0,3003,0,00360,M
43 13,71,20,83,90,1257,9,0,1109,0,1045,0,09366,0,05005,0,2100,0,07031,0,5035,1,377,1,056,50,96,0,000005,0,03029,0,02400,0,05040,0,01040,0,005012,17,00,20,14,110,6,007,0,1054,0,3002,0,0070,0,1550,0,1151,M
44 13,01,02,07,5,115,0,0,0,1771,0,1037,0,1009,0,00353,0,720,0,07309,0,3001,1,002,2,400,14,12,0,00711,0,03302,0,03353,0,01235,0,02143,0,003700,15,49,30,15,100,2,770,3,0,1700,0,5401,0,530,0,206,0,4370,0,1027,M
45 12,46,24,04,03,97,075,9,0,1100,0,1096,0,2273,0,00543,0,201,0,00243,0,2976,1,599,2,009,21,94,0,00710,0,07217,0,07743,0,01432,0,01709,0,01000,15,09,40,48,97,65,711,4,0,1033,1,003,1,100,0,221,0,4366,0,2075,M
46 16,02,23,20,102,7,797,0,0,00200,0,00000,0,03020,0,03023,0,1500,0,05007,0,3951,1,107,2,000,00,53,0,000420,0,000000,0,01010,0,007951,0,0300,0,000042,19,15,13,00,123,0,1150,0,1101,0,1551,0,1059,0,00452,M
47 15,70,17,09,103,6,713,0,00710,0,1202,0,00004,0,00000,0,1002,0,00002,0,3000,0,0001,1,004,16,16,0,00771,0,00001,0,02701,0,01230,0,03000,0,000104,20,42,27,20,136,1,1209,0,1300,0,5000,0,3003,0,101,0,1792,0,1000,M
48 19,17,24,0,132,4,1123,0,00704,0,2450,0,0003,0,1110,0,2397,0,070,0,0555,1,508,11,07,116,1,0,00339,0,00297,0,0009,0,0000,0,00404,0,01204,20,90,29,94,151,7,1332,0,1037,0,3003,0,3039,0,1767,0,3376,0,1003,M
49 15,25,23,36,102,7,702,2,0,00001,0,1002,0,00000,0,05300,0,0107,0,00330,0,4003,1,070,2,003,30,50,0,00070,0,03100,0,00010,0,01002,0,00001,16,94,27,66,112,070,5,0,1117,0,1004,0,2302,0,1119,0,2000,0,00307,M
50 13,73,22,03,93,6,570,3,0,1112,0,2203,0,1100,0,00025,0,2009,0,07002,0,1221,1,109,2,001,53,23,0,000429,0,05036,0,05001,0,01020,0,01001,0,000003,10,03,32,01,100,6,097,7,0,1051,0,7720,0,0943,0,2200,0,3506,0,1431,M
51 14,54,27,54,96,73,050,0,0,1110,0,1505,0,1039,0,07064,0,2303,0,07007,0,17,1,033,2,070,12,55,0,005007,0,0426,0,04741,0,0100,0,01037,0,005466,17,40,17,13,124,1,041,2,0,1670,0,0577,0,7026,0,1712,0,4210,0,1341,M
52 14,00,20,13,94,7,050,0,0,00002,0,0770,0,07390,0,02350,0,1504,0,00022,0,07271,20,1,195,45,4,0,000710,0,01042,0,01000,0,01009,0,01041,0,000201,19,07,00,00,122,6,1110,0,1464,0,1077,0,2014,0,1000,0,3020,0,00216,M
53 15,10,20,60,100,1,700,0,0,117,0,2022,0,1722,0,1020,0,2164,0,07350,0,5002,1,073,8,54,54,18,0,007026,0,02501,0,01000,0,01007,0,01000,0,000142,20,96,31,40,136,8,1515,0,1700,0,4233,0,4704,0,2073,0,3706,0,1442,M
54 19,01,22,15,130,1200,0,00031,0,1027,0,1470,0,00040,0,1502,0,05395,0,7002,1,017,1,003,112,4,0,000404,0,01003,0,03301,0,01221,0,01150,0,001997,27,32,30,00,106,0,2300,0,1512,0,315,0,5372,0,2300,0,2700,0,07015,M
55 13,54,10,30,07,40,566,1,0,0070,0,00120,0,00006,0,04033,0,1300,0,07306,0,2000,0,7000,2,005,23,16,0,000402,0,01040,0,01007,0,01035,0,01000,0,0003,15,13,10,20,90,7,711,2,0,244,0,1771,0,200,0,1200,0,2017,0,07200,M
56 13,08,15,71,05,03,120,0,1075,0,127,0,04500,0,0111,0,1067,0,00011,0,1052,0,7477,1,303,14,47,0,000407,0,01000,0,01000,0,00040,0,01070,0,000425,14,5,20,40,96,09,450,5,0,1312,0,2770,0,100,0,07203,0,3104,0,00103,M
```

b. data input oleh Perangkat Lunak Weka

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **None** Apply Stop

Current relation: Relation: wdbc Instances: 569 Attributes: 31 Sum of weights: 569

Attributes:

No.	Name
1	mean_radius
2	mean_texture
3	mean_perimeter
4	mean_area
5	mean_smoothness
6	mean_compactness
7	mean_concavity
8	mean_concave_points
9	mean_symmetry
10	mean_fractal_dimension
11	standart_error_radius
12	standart_error_texture
13	standart_error_perimeter
14	standart_error_area
15	standart error smoothness

Remove

Status OK Log

Selected attribute: Name: mean_radius Distinct: 456 Type: Numeric Missing: 0 (0%) Unique: 359 (63%)

Statistic	Value
Minimum	6.981
Maximum	28.11
Mean	14.127
StdDev	3.524

Class: Diagnosis (Nom) Visualize All

6.98 17.55 28.11

2. Dataset Abalone

Teknik dalam mengubah data ke dalam bentuk arff

- Letakkan atribut nominal (jenis kelamin) pada kolom terakhir.
- Selanjutnya, susun atribut dalam file ARFF dalam urutan sebagai berikut: Diameter, Rings, Height, Length, Shell weight, Shucked weight, Whole weight, Viscera weight.

a. File arff

```

1 @RELATION abalone
2
3 @ATTRIBUTE Length real
4 @ATTRIBUTE Diameter real
5 @ATTRIBUTE Height real
6 @ATTRIBUTE Whole_weight real
7 @ATTRIBUTE Shucked_weight real
8 @ATTRIBUTE Viscera_weight real
9 @ATTRIBUTE Shell_weight real
10 @ATTRIBUTE Rings real
11 @ATTRIBUTE Sex {M,F,I}
12
13 @data
14 0.455,0.365,0.095,0.514,0.2245,0.301,0.15,15,M
15 0.35,0.265,0.09,0.2225,0.0995,0.0445,0.07,7,M
16 0.53,0.42,0.135,0.677,0.2545,0.1435,0.21,9,F
17 0.46,0.305,0.125,0.525,0.2355,0.116,0.355,10,M
18 0.33,0.255,0.08,0.205,0.0895,0.0395,0.055,7,I
19 0.425,0.3,0.095,0.5515,0.141,0.0755,0.12,8,I
20 0.53,0.435,0.15,0.7775,0.237,0.1435,0.33,20,F
21 0.545,0.425,0.125,0.708,0.294,0.1495,0.26,16,F
22 0.475,0.37,0.125,0.5095,0.2255,0.1125,0.165,9,M
23 0.35,0.44,0.15,0.8945,0.3495,0.151,0.32,19,I
24 0.525,0.38,0.14,0.0805,0.194,0.1475,0.21,14,F
25 0.43,0.375,0.11,0.606,0.1675,0.081,0.135,10,M
26 0.49,0.38,0.135,0.5415,0.2375,0.095,0.19,11,M
27 0.535,0.405,0.145,0.6845,0.2775,0.171,0.205,10,F
28 0.47,0.355,0.1,0.4355,0.1075,0.0895,0.105,14,I
29 0.5,0.4,0.13,0.6445,0.258,0.133,0.24,12,M
30 0.355,0.28,0.085,0.2905,0.095,0.0395,0.115,7,I
31 0.46,0.34,0.1,0.431,0.108,0.067,0.13,10,F
32 0.365,0.285,0.08,0.2555,0.097,0.041,0.1,7,M
33 0.45,0.32,0.1,0.381,0.1705,0.075,0.115,9,M
34 0.555,0.28,0.095,0.2455,0.0955,0.062,0.075,11,M
35 0.38,0.275,0.1,0.2255,0.08,0.049,0.085,10,I
36 0.565,0.46,0.155,0.9395,0.4275,0.214,0.27,12,F
37 0.35,0.435,0.135,0.7035,0.318,0.21,0.2,9,F
38 0.615,0.48,0.165,1.1,0.515,0.513,0.301,0.305,10,F
39 0.56,0.46,0.16,0.9205,0.3825,0.198,0.2,11,F
40 0.58,0.45,0.185,0.9955,0.3945,0.272,0.205,11,F
41 0.59,0.445,0.14,0.931,0.356,0.234,0.28,12,M
42 0.605,0.475,0.18,0.9305,0.395,0.229,0.205,11,M
43 0.575,0.425,0.14,0.8655,0.393,0.227,0.2,11,M
44 0.58,0.47,0.165,0.9975,0.3935,0.242,0.33,10,M
45 0.68,0.56,0.165,1.039,0.6095,0.289,0.46,15,F
46 0.665,0.525,0.165,1.138,0.555,0.3575,0.35,18,M
47 0.68,0.55,0.175,1.798,0.815,0.3925,0.455,19,F
48 0.705,0.55,0.2,1.7,0.895,0.433,0.435,0.49,13,F
49 0.465,0.355,0.105,0.4795,0.227,0.124,0.125,8,M
50 0.54,0.475,0.155,1.227,0.5305,0.3075,0.34,16,F
51 0.45,0.355,0.105,0.5225,0.227,0.1155,0.165,8,F
52 0.575,0.445,0.135,0.883,0.383,0.2035,0.26,11,F
53 0.355,0.28,0.09,0.2275,0.154,0.080,0.09,9,M
54 0.45,0.355,0.105,0.425,0.1805,0.091,0.115,9,F
55 0.55,0.425,0.135,0.8315,0.362,0.196,0.27,14,F
56 0.24,0.175,0.045,0.07,0.07015,0.0255,0.02,5,I

```

b. data input oleh Perangkat Lunak Weka

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** | Apply | Stop

Current relation: Relation: abalone, Instances: 4177, Attributes: 9, Sum of weights: 4177

Attributes: All | None | Invert | Pattern

No.	Name
1	Length
2	Diameter
3	Height
4	Whole_weight
5	Shucked_weight
6	Viscera_weight
7	Shell_weight
8	Rings
9	Sex

Remove

Status: OK

Selected attribute: Name: Length, Missing: 0 (0%), Distinct: 134, Type: Numeric, Unique: 6 (0%)

Statistic	Value
Minimum	0.075
Maximum	0.815
Mean	0.524
StdDev	0.12

Class: Sex (Nom) | Visualize All

Histogram showing the distribution of Length values for different Sex categories (M, F, I).

3. Dataset Bank Marketing

Teknik mengubah data ke dalam bentuk arff

- Susun atribut dalam file ARFF dalam urutan sebagaimana aslinya.
- Ubah kedua file csv (bank.csv dan bank-full.csv) ke dalam format ARFF.

a. File arff

Note : kiri merupakan data arff dari nama file bank.arff dan sebelah kanan dengan nama bank-full.arff

b. data input oleh Perangkat Lunak Weka dari file bank.arff

- c. data input oleh Perangkat Lunak Weka dari file bank-full.arff

