

# Teknik Pengukuran: Metode Klasifikasi *K-Nearest Neighbour* pada Data Pengukuran

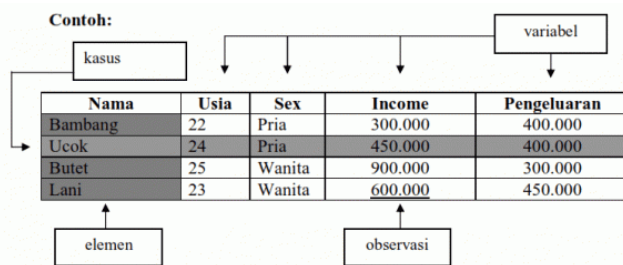
Euftrat Tsaqib Qasthari <sup>1</sup>

Departemen Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA)  
Universitas Indonesia<sup>1</sup>, Depok, 16424, Indonesia

(euftrat.tsaqib@ui.ac.id<sup>1</sup>)

## 1. Data Pengukuran

Data pengukuran merupakan fakta atau angka-angka yang didapat dari suatu pengukuran. Dalam pengukuran seringkali kita menjumpai data yang memiliki berbagai karakteristik. Karakteristik data tersebut yaitu elemen, variabel yang menjadi perhatian, kasus yaitu informasi yang menyangkut variabel pada suatu elemen dan observasi yaitu hasil pengamatan tertentu.



## 2. Klasifikasi Data

Proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui adalah klasifikasi data. Metode klasifikasi yang beragam antara lain Naive Bayes, Fuzzy dan k-Nearest Neighbor. Pada klasifikasi data dapat dibagi ke 3-fase yaitu fase pembelajaran (training phase) dimana model di bangun dari hasil pembelajaran dari data, lalu fase percobaan (testing phase) dimana model akan diberikan beberapa tes sample yang diambil secara random pada data yang sudah ada namun bukan data pada *training phase* umumnya fase ini akan dipecah menjadi tahap validasi dan *testing*. Fase terakhir pada klasifikasi data adalah tahap penggunaan (*usage phase*) dimana permodelan akan digunakan untuk memprediksikan dan mengklasifikasi data yang belum diketahui polanya.

## 3. Supervised dan Unsupervised learning

Pada proses klasifikasi data ada dua jenis cara untuk menemukan pola data yaitu supervised dan unupervised learning. Supervised learning adalah proses klasifikasi atau pembelajaran data yang menggunakan pola data yang

sudah ada, sedangkan data yang belum memiliki pola dan ingin ditemukan polanya dapat menggunakan cara Unsupervised learning.

## 4. K-NN (K- Nearest Neighbour)

K-Nearest Neighbour adalah metode klasifikasi yang mngklasifikasi objek berdasarkan voting terbanyak dari jarak pada pola data sebelumnya (data yang ada di memori model). Ide K-NN dimulai pada awal 1970 dimana metode ini digunakan untuk mengestimasi pola pada data-data statistik. K-NN menggunakan klasifikasi berdasarkan data tetangga terdekat dari sampel uji yang baru. Jarak yang digunakan adalah *Euclidean Distance*. Perumusan jarak tersebut adalah sebagai berikut:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r x_i - a_r x_j)^2}$$

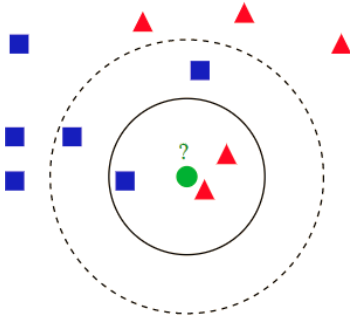
Dimana  $d(x_i, x_j)$  adalah jarak Euclidean record ke- $i$  dan record ke- $j$  dengan  $a_r$  merupakan data ke- $r$ . Selain *Euclidean distance* juga dapat digunakan *Manhattan distance* dengan perumusan sebagai berikut:

$$d(x_i, x_j) = \sum_{r=1}^n |a_r x_i - a_r x_j|$$

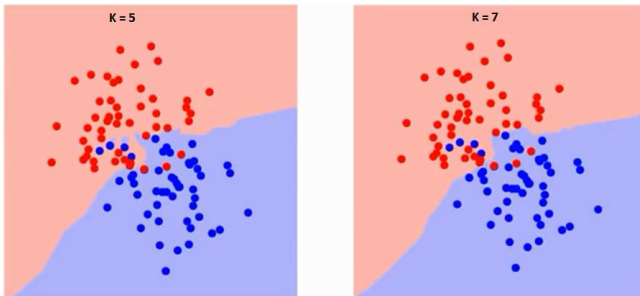
Sedangkan perumusan K-NN adalah sebagai berikut

$$D_{nn}(C_1, C_2) = \min_{1 \leq i \leq r, 1 \leq j \leq s} d(y_i, z_j)$$

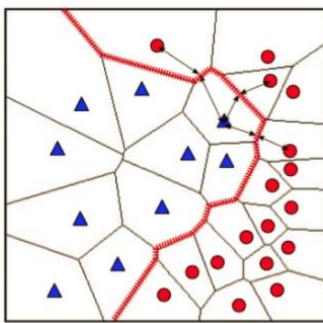
Metode k-NN akan menentukan nilai berdasarkan voting dari jumlah  $k$  data *training* terdekat yang ditentukan berdasarkan jaraknya (*Euclidean Distance* atau *Manhattan Distance*). Nilai  $k$  adalah integer positif ( $k=1,2,3,4$  dst.) yang menentukan jumlah tetangga yang ada pada titik *test* tertentu.



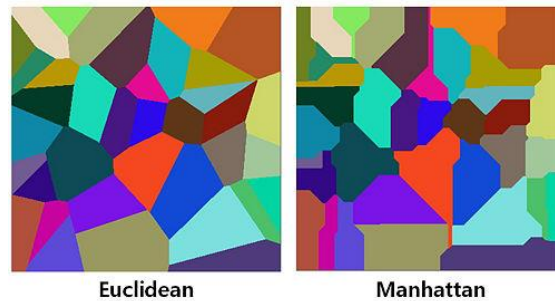
Pada gambar diatas diperlihatkan dua lingkaran, yaitu lingkaran solid dan lingkaran titik-titik pada lingkaran solid nilai  $k$  adalah 3, karena ada tiga buah data *training* sebelumnya untuk menentukan klasifikasi sampel *test* yang digambarkan dengan bulat hijau, pada kejadian ini sampel *test* diklasifikasikan menjadi tipe data segitiga merah karena pada  $k=3$  segitiga merah mempunyai voting terbanyak (dua). Sedangkan apabila kita mengganti nilai  $k$  menjadi lima maka hasil voting terbanyak ada pada kotak biru sehingga sampel *test* akan diklasifikasikan menjadi kotak biru. Pada data yang banyak elemennya pada umumnya nilai  $k$  yang besar akan mengurangi efek data *noise* pada klasifikasi, namun membuat garis pembagi menjadi tidak mengklasifikasi secara baik contoh:



## 5. Voronoi Tesselation pada k-NN



Voronoi Tesselation adalah metode partisi untuk membagi bagian-bagian pada suatu ruang berdasarkan letak titik yang tersebar pada ruang tersebut. Setiap partisi pada Voronoi Tesselation disebut dengan Voronoi Cells. Teknik pembagian partisi dengan Voronoi Tesselation biasanya digunakan untuk menggambarkan klasifikasi pada titik-titik yang terklasifikasi dengan menggunakan k-NN agar dapat terlihat daerah klasifikasi antara satu titik dengan titik lainnya.



Perbedaan pengambilan jarak (*Euclidean Distance* atau *Manhattan Distance*) dapat mempengaruhi batas partisi Voronoi Tesselation karena Euclidean mengambil jarak dari kuadrat akar jumlah-jumlah titiknya dan Manhattan mengambil jarak dari mutlak jumlah-jumlah titiknya.

## Referensi

1. <http://scikit-learn.org/stable/modules/neighbors.html> diakses 24 Desember 2017, 12.30 WIB
2. <http://hyperphysics.phy-astr.gsu.edu/hbase/magnetic/motdc.html> diakses 16 Desember 2017, 12.32 WIB
3. Krisandi, Helmi, Prihandono (2013). ALGORITMA k-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA HASIL PRODUKSI KELAPA SAWIT PADA PT. MINAMAS KECAMATAN PARINDU, Universitas Tanjungpura.
4. <https://math.stackexchange.com> diakses 24 Desember 2017, 12.36 WIB
5. <https://www.ethz.ch> diakses 25 Desember 2017, 12.40 WIB
6. <http://chem-eng.utoronto.ca/~datamining/Presentations/K-NN.pdf> diakses 25 Desember 2017, 12.45 WIB