



Materi Kuliah – [7]: Data Mining

K–Nearest Neighbour

lizda.iswari@uui.ac.id

April 2012

Pengantar

- ▶ Konsep dasar dari **K-NN** adalah mencari **jarak terdekat** antara data yang akan dievaluasi dengan K tetangga terdekatnya dalam data pelatihan.
- ▶ Penghitungan jarak dilakukan dengan konsep Euclidean.
- ▶ Jumlah kelas yang paling banyak dengan jarak terdekat tersebut akan menjadi kelas dimana data evaluasi tersebut berada.

Algoritma

- ▶ Algoritma
 - Tentukan parameter K = jumlah tetangga terdekat.
 - Hitung jarak antara data yang akan dievaluasi dengan semua data pelatihan.
 - Urutkan jarak yang terbentuk (urut naik) dan tentukan jarak terdekat sampai urutan ke- K .
 - Pasangkan kelas (C) yang bersesuaian.
 - Cari jumlah kelas terbanyak dari tetangga terdekat tersebut, dan tetapkan kelas tersebut sebagai kelas data yang dievaluasi.

Contoh

- ▶ Kasus perusahaan makanan ringan.
- ▶ Variabel penentu:
 - Kenaikan derajat keasaman (v1)
 - Penyusutan volume (v2)
- ▶ Kategori kelas: Baik dan Buruk
- ▶ Lihat: Kasun_KNN.pdf



Latihan: Data Hasil Tes Asisten Laboratorium

Data	Pemrog. Java	Pemrog. php	Wawancara	Laboratorium
1	8	7	7	Lab. Sirkel
2	8	7	9	Lab. KSC
3	9	8	8	Lab. PIT
4	9	9	8	Lab. Sirkel
5	7	8	8	Lab. KSC
6	8	9	7	Lab. PIT
7	7	7	8	Lab. Sirkel
8	9	7	8	Lab. KSC
9	9	7	7	Lab. PIT
10	7	8	7	Lab. PIT

Latihan: Test Data

- ▶ Jika ada calon asisten dengan kriteria hasil tes:
 - Pemrograman Java: 7;
 - Pemrograman PHP: 9;
 - Wawancara: 7
- ▶ Tentukan posisi lab apakah yang tepat dengan mempertimbangkan 5 data terdekat!

