

## Tugas 3

### CTI-2G3 Sistem Cerdas

### Genap 2020/2021

Diberikan DataSetTB3.xls yang berisikan sheet Data dan Submit untuk kasus hand-writing digit recognizer (bagian dari MNIST data yang diunduh pada [www.kaggle.com](http://www.kaggle.com)). Data itu merepresentasikan gambar grayscale dari tulisan tangan dengan dimensi 28 x 28 pixel (total 784 pixel). Setiap pixel dapat berisi nilai 0-255 yang merepresentasikan derajat keabuan dari pixel, semakin tinggi semakin gelap.

Sheet Data terdiri atas 1001 baris dan 786 kolom, dimana baris pertama merupakan label attribute yaitu idData, label-kelas dan idPixel. Sheet Submit terdiri atas 786 kolom dengan kolom pertama adalah idData, kolom kedua adalah label kelas dari data 0-9 dan kolom lainnya merepresentasikan nilai pixel dengan rentang 0-255 (dalam 1 vektor)

Sheet Submit terdiri atas 501 baris dan 785 kolom, baris pertama merupakan label attribute yaitu idData dan idPixel. Sheet Submit tidak memiliki label kelas hanya mengandung nilai pixel yang disimpan dalam 785 kolom.

**Sheet Data**

A	B	C	D	E	F
idData	label	pixel0	pixel1	pixel2	pixel3
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0

**Sheet Submit**

A	B	C	D	E	F
idData	pixel0	pixel1	pixel2	pixel3	pixel4
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0

Catatan:

- ✓ **Sheet Submit hanya digunakan untuk pengujian akhir (“pengetahuan terbaik” dari kNN yang telah diujikan secara lokal)**
- ✓ **Silahkan menggunakan sheet Data untuk melakukan pelatihan dan uji peformansi lokal**

#### **Hal yang dapat diobservasi/dipaparkan:**

- Teknik/strategi pemrosesan data pada Sheet Data dengan melakukan hold-out 70/30 untuk menghasilkan data latih (70%) dan data uji (30%), dan/atau normalisasi data jika dibutuhkan
- Strategi penggunaan algoritma kNN yang meliputi pemilihan ukuran jarak, nilai k dan cara mengambil keputusan

#### **Proses yang harus dibangun (bisa berbentuk fungsi/prosedur):**

- Membaca file inputan Sheet Data
- Prapemrosesan data yang meliputi: pembagian data latih dan data uji, serta normalisasi jika dibutuhkan

- Menentukan minimal dua ukuran jarak (Eucliden, Manhattan, Minkowski, Supremum atau metode yang lain)
- Menentukan minimal 2 nilai k dalam kNN
- Klasifikasi dengan “**pengetahuan**” kNN yang dibangun dengan data latih (70%) pada data uji (30%), kemudian menghasilkan file **OutputLatih.xls** dengan format 2 kolom awal berisikan idData dan klasifikasi kelas. Kemudian menghasilkan ukuran performansi berupa akurasi yang diletakkan pada kolom 3 baris 2
- Klasifikasi dengan “**pengetahuan terbaik**” kNN yang dibangun dengan data latih pada Sheet Submit kemudian menghasilkan file **OutputSubmit.xls** dengan format 2 kolom, kolom pertama adalah idData dan kolom kedua adalah klasifikasi kelas

**Output** dari sistem terdiri atas 2, yaitu:

1. **File OutputLatih.xls** untuk hasil pengujian lokal dengan menggunakan Sheet Data, Data Latih (70%) dan Data Uji (30%). Akurasi mungkin dihitung karena Data Uji adalah bagian dari Sheet Data yang telah memiliki Label sehingga dapat dibandingkan dengan hasil dari Knn.
2. **File OutputSubmit.xls** adalah sebuah file luaran. yang berisi satu vektor kolom berisi 10 baris angka bernilai integer (bilangan bulat) yang menyatakan nomor baris/record id restoran (1-100) pada file masukan.xls.

**OutputLatih.xls**

A	B	C
idData	Klasifikasi	Akurasi
1	0	90%
2	1	
3	3	
4	4	
5	2	
6	9	
7	8	
8	7	

**OutputSubmit.xls**

A	B
idData	Klasifikasi
1	1
2	9
3	8
4	6
5	5
6	9
7	8
8	7

## Aturan dan Penilaian

### 1. Pengumpulan

- Tugas dikumpulkan maksimal hari **Sabtu** tanggal **3 Juli 2021** pukul **20.00 WIB** melalui **LMS**.
- Pengumpulan berupa:
  1. *source code* program (file sama antara anggota kelompok)
  2. Laporan Observasi (file sama antara anggota kelompok)
  3. Video Presetasi Program (video berbeda antara mahasiswa)
- Penamaan file zip/rar **HARUS** menggunakan format **KELOMPOK\_KELAS\_NIM:**
  - Misal: **01\_IT4301\_1303194106.zip** atau **01\_IT4304\_1303194106.rar**
  - Tidak ada tanda hubung (-) atau garis bawah (\_) di penamaan kelas
- Penamaan file dan format source code yang tidak sesuai akan menyebabkan pengumpulan **TIDAK DINILAI**

### 2. Source Code Program

- Source Code harus bisa dijalankan dan terkumpul dalam **SATU FILE** Source Code

- Sistem boleh dibangun menggunakan bahasa pemrograman apapun
- **Tidak diperbolehkan menggunakan library** yang secara langsung melakukan proses kNN atau proses yang harus dibangun/observasi
- Penamaan file dan format source code yang tidak sesuai akan menyebabkan pengumpulan **TIDAK DINILAI**
- **JANGAN** mengumpulkan file .exe

### 3. Laporan Observasi

- Laporan minimum 5 halaman, tidak termasuk cover laporan
- Font: Times New Roman, font size: 12, spacing: 1, dalam format **.pdf**
- Berisi analisis dan penjelasan strategi penyelesaian masalah (teknik menentukan nilai-nilai parameter atau pemilihan strategi) terkait **Hal yang harus diobservasi**. Serta buktikan strategi yang digunakan telah **sesuai dengan code program yang dibuat**.
- Tambahkan **screenshot** terkait hal yang diobservasi agar terlihat **kesesuaian antara strategi yang digunakan dengan code program**.
- Ketidak sesuaian strategi yang dijelaskan dengan code program yang dibuat akan mempengaruhi penilaian.
- Tuliskan nilai k, metode ukuran jarak yang mendapatkan hasil terbaik untuk pengujian lokal.
- **Berikan kesimpulan dan hasil terkait program kNN** yang dibuat serta berikan **screenshot** hasil running berdasarkan nilai-nilai parameter optimum (tampilan hasil akurasi terbaik dari pengujian lokal dan hasil tebakan untuk 10 data awal Sheet Submit).

### 4. Video Presentasi

- Video presentasi program maximum berdurasi selama 10 menit.
- Presentasikan strategi yang dibangun dalam merancang aplikasi kNN tersebut. Jelaskan fungsi atau prosedur utama yang dibuat.
- Upload video tersebut kedalam youtube atau google drive. Tuliskan link video yang dapat diakses kedalam laporan observasi.

### 5. Detil Penilaian

- **Nilai = E** bagi siapapun yang terbukti melakukan **KECURANGAN**.
- Kesamaan kode dan/atau laporan **hingga 80% (kecuali dalam satu kelompok)**
- Mengumpulkan kode program dan/atau laporan **milik orang lain (kecuali dalam satu kelompok)**.

Bandung, 18 Juni 2021



(Tim Dosen CTI2G3 Sistem Cerdas)



**Daftar Asisten Dosen CTI2G3 SISTEM CERDAS**  
**Genap 2020/2021**

Kelas	Nama
IT-42-01	Angel Metanosa Afinda - 082129916992
IT-42-02	Dimitri Irfan Dzidny - 081283083410
IT-42-03	Angel Metanosa Afinda - 082129916992
IT-43-01	Muhammad Farrel - 08114551997
IT-43-02	Muhammad Aziz Pratama - 085273309914
IT-43-03	Angel Metanosa Afinda - 082129916992