



Departemen Teknik Elektro, Program Studi S2 Teknik Elektro

Ujian Tengah Semester
Semester Genap 2020/2021

Mata Kuliah (Kode)	: Rekayasa Data Terapan (ENEE802606)		
Dosen	: Dr. Prima Dewi Purnamasari, Dr. Anak Agung Putri Ratna		
Hari, Tanggal	: Jumat, 16 April 2021	Waktu	: 1 hari. Dikumpulkan Sabtu 17 April 2021, jam 10.00 WIB
Kompetensi (CPL)	: 5	Sifat Ujian	: take home

Petunjuk Ujian

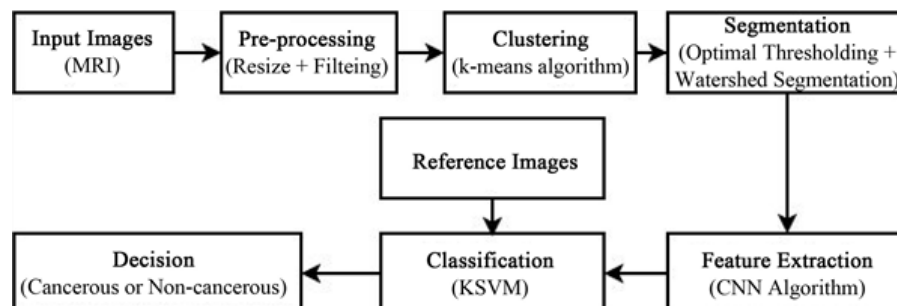
Kerjakan ujian ini dalam bentuk Notebook Python, dengan bagian penjelasan dituliskan dalam bentuk "markdown text" dan bagian program dituliskan dalam bentuk kode Python. Kumpulkan melalui area yang tersedia di EMAS. **File yang dikumpulkan berjumlah 3 buah:**

- 1) sourcecode dalam bentuk .ipynb diberi nama file: nama_anda_code.ipynb,
- 2) file data training dalam bentuk .csv dengan nama file: nama_anda_train.csv, dan
- 3) file data testing dalam bentuk .csv dengan nama file: nama_anda_test.csv.

1. Rancanglah sebuah solusi untuk permasalahan klasifikasi teks dari data yang diberikan dengan metode machine learning. Jenis classifier yang digunakan adalah **Non-ANN classifier**, misal Naïve Bayes seperti yang dipelajari di kelas, atau classifier lain yang telah dipelajari di kelas lain.

Rancangan dituliskan dengan cara menggambarkan langkah-langkah desain dalam bentuk *block diagram* (image dapat di-insert dalam area *markdown text*). Langkah-langkah yang dituliskan harus lengkap sesuai dengan apa yang akan diimplimentasikan dalam program. Bagian *Pre-processing* dan *Feature Extraction* sebagian besar (atau seluruhnya) mengikuti apa yang dipelajari di kelas dan harus di-break down sesuai dengan langkah yang akan dilakukan (bukan seperti gambar di bawah).

Ket: block diagram bukanlah flowchart, melainkan hanya kumpulan blok yang menunjukkan urutan kerja. Contoh block diagram adalah sebagai berikut:





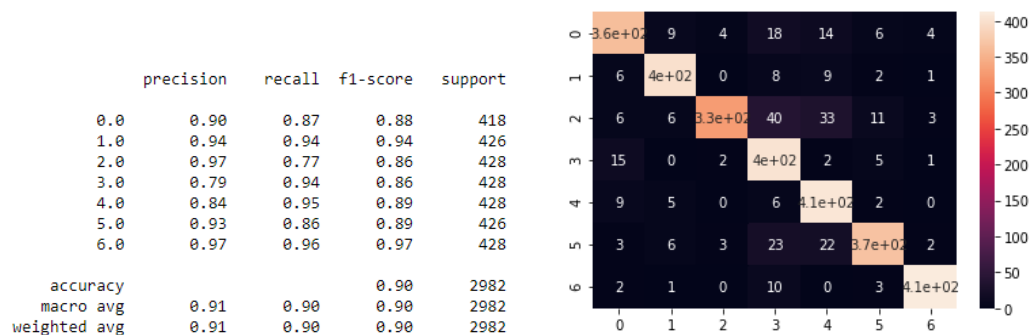
2. Bagilah data yang diberikan ke dalam 2 file CSV terpisah, yaitu untuk data training, dan untuk data testing. Data training dan data testing dibagi dengan proporsi 80% untuk data training, dan 20% untuk data testing. Proses pemisahan data ini dapat dilakukan dengan software apapun, tidak harus Python, misal menggunakan Matlab, Ms Excel dsb. Tuliskan proses yang Anda lakukan dalam bentuk markdown text. Jelaskan atribut dan kelas data. File data training, dan data testing dilampirkan Ketika pengumpulan ujian.
3. Implementasikan rancangan pada nomer 1 dalam kode Python dengan penjelasan yang memadai di tiap blok prosesnya, dalam bentuk markdown text. Jadi, kode program dipisah per blok proses. Data pada no 2 di-*insert* ke dalam program dalam bentuk upload file, agar langsung dapat dijalankan di Google Colab, contoh:

```
# It will prompt you to select a file. Click on "Choose Files" then select and upload the file.
# wait for the file to be 100% uploaded.

from google.colab import files
uploaded = files.upload()

# To store dataset in a Pandas Dataframe
import io
df2 = pd.read_csv(io.BytesIO(uploaded['Filename.csv']))
```

4. Tampilkan hasil klasifikasinya dalam bentuk tabel dan confusion matrix, contoh:



5. Buatlah analisis dan kesimpulan dari hasil pada nomer 4.
6. Jika Anda menyadur sebagian atau seluruh langkah yang digunakan untuk mengerjakan ujian ini dari Internet, maka tuliskan referensinya dengan lengkap. Misal untuk pengolahan data dari referensi A, untuk klasifikasi dari referensi B dsb.