

## Soal 2

### Proyek Akhir Kelas Mahir

#### Deskripsi:

Pada proyek akhir PSDS 7.0 Kelas Mahir anda akan melakukan:

1. Klasifikasi dan Klasterasi untuk data tabular untuk data dapat diunduh di:

<https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data>

2. Klasifikasi dengan menggunakan Transferlearning untuk data Citra (gambar) data dapat diunduh pada link berikut

<https://www.cs.toronto.edu/%7Ekriz/cifar.html>

#### A. Klasifikasi data Tabular

- Buatlah klasifikasi dengan menggunakan dataset yang sudah diberikan pada poin 1
  - Untuk splitting data anda dibebaskan memilih ratio (70:30 , 80:20 atau rasio lainnya)  
remark : rasio splitting data terkadang bisa mempengaruhi akurasi ataupun metrik evaluasi lain
  - Lakukan Klasifikasi dengan menggunakan **semua** model Machine Learning dibawah:
    1. Naive Bayes(Gaussian Naive Bayes)
    2. KNN
    3. Random Forest atau Decision Tree (pilih salah satu)
    4. Support Vector Machine(anda dapat memilih menggunakan kernel:linear, rbf ataupun kernel lainnya yang anda ketahui)
    5. Multi Layer Perceptron
  - Lalu lakukan hyperparameter tuning sehingga mendapatkan akurasi yang paling baik. clue : akurasi diatas 80% sudah termasuk bagus. Jika belum mendapatkan hasil yang terbaik dapat dicoba menggunakan feature selection atau reduksi dimensi minimal 1 yang ada pada Lampiran 1.
  - lalu pilihlah salah satu model (yang menurut anda paling baik). Terapkan Evaluasi model dengan Cari hal berikut
    - Confusion Matrix
    - Accuracy
    - Recall
    - Precision
    - F1-Score
  - Berikan penjelasan pada setiap blok kode (untuk nilai tambahan)  
clue: Anda dapat menggunakan markdown ataupun comment
- Note:Lakukan Seluruh proses dari preprocessing hingga evaluasi)

#### B. Klasterisasi data Tabular

- Gunakan data yang sama seperti pada kasus klasifikasi,lalu lakukan analisis clustering, menggunakan minimal 2 metode Kmedians,K Means,Hierarchical,Fuzzy C means,dll)

- Jumlah k kelas adalah jumlah label pada kasus klasifikasi. Karena tujuan kita adalah memilih fitur yang paling optimal dari sekumpulan fitur yang anda,  
Clue: Gunakan `code accuracy_score({data_hasil_klaster},{data_label_asli})`. score akurasi tidak berpengaruh. Namun akan lebih baik jika akurasi lebih dari 50%. Atau anda juga bisa menggunakan metrik evaluasi lain seperti ARI, Silhouette\_score, dsb.
  - Berikan penjelasan pada setiap blok kode (untuk nilai tambahan)  
clue: Anda dapat menggunakan markdown ataupun comment
- Note: Lakukan Seluruh proses dari preprocessing hingga evaluasi)

### C. Klasifikasi data Citra

- Jika sebelumnya kita terapkan algoritma machine learning pada data tabular. Maka pada bagian ini kita akan mencoba menggunakan data citra (gambar)
  - gunakan minimal 2 arsitektur pretrained model pada Lampiran 2 (anda dapat memilih 3 model dari grup yang sama, ex: *EfficientNetB1*, *EfficientNetB5*, *EfficientNetB7*).
  - Langkah pertama, gunakan data citra dengan arsitektur tersebut dengan menggunakan pre-trained model. Langkah kedua, Cobalah lakukan transfer learning dengan membangun arsitektur Neural Network (CNN, ANN, maupun RNN). Langkah ketiga, gunakan data citra dengan arsitektur tersebut menggunakan pre-trained model dengan bagian feature extraction (bukan head) Trainable = False. sehingga pengupdatean bobot hanya pada bagian head saja.
  - Lalu lakukan evaluasi seperti pada bagian A. cari hal berikut
    - Confusion Matrix
    - Accuracy
    - Recall
    - Precision
    - F1-Score
  - Berikan penjelasan pada setiap blok kode (untuk nilai tambahan)  
clue: Anda dapat menggunakan markdown ataupun comment
- Note: Lakukan Seluruh proses dari preprocessing hingga evaluasi)

3. Gabungkan apa semua yang anda kerjakan ( bagian A, B, dan C) lalu buat ke dalam 1 file Jupyter notebook (IPYNB)  
(clue : Anda dapat menggunakan Header pada markdown untuk membagi dokumen pada file IPYNB yang akan dikumpulkan). Jangan lupa bubuhkan nama kelompok dan anggota anda pada bagian awal notebook

- Sebagai nilai tambah sangat dibolehkan membuat video yang mempresentasikan apa yang anda kerjakan ke sosial media anda. link tidak boleh diprivate

### PENTING:

- Anda diperbolehkan menggunakan CHAT GPT, atau chatbot LLM lain dengan syarat anda memahami metode yang anda gunakan. Input datanya seperti apa, waktu proses apa yang terjadi, lalu proses outputnya bagaimana bisa menghasilkan hasil tersebut.
- Soal yang anda pilih harus sesuai dengan apa yang dikerjakan, jika ada indikasi perbedaan antara jawaban dan soal, nilai akhir dikurangi 30%)

## Lampiran 1

Reduksi Dimensi/Feature Selection
Recursive Feature Elimination
Variance Treshold
SelectKBest
Tree-based Method
Non-Negative Matrix Factorization
Independent Component Analysis
Principal Component Analysis
Incremental PCA
Correlation-based Feature Selection

## Lampiran 2

Kasus Klasifikasi Gambar(pretrained model) :

Group	Model
A	MobileNetV3Large
	MobileNetV3Small
B	ConvNeXtBase
	ConvNeXtLarge
	ConvNeXtSmall
	ConvNeXtTiny
	ConvNeXtXLarge
C	DenseNet121
	DenseNet169
	DenseNet201
D	EfficientNetB0
	EfficientNetB1
	EfficientNetB2
	EfficientNetB3
	EfficientNetB4
	EfficientNetB5
	EfficientNetB6

	EfficientNetB7
<b>E</b>	EfficientNetV2B0
	EfficientNetV2B1
	EfficientNetV2B2
	EfficientNetV2B3
	EfficientNetV2L
	EfficientNetV2M
	EfficientNetV2S
<b>F</b>	Inception_resnet_v2
<b>G</b>	Inception_resnet_v3
	Inception_v3
<b>H</b>	Mobilenet
	Mobilenet_v2
	Mobilenet_v3
<b>I</b>	NASNetLarge
	NASNetMobile
<b>J</b>	RegNetX002
	RegNetX004
	RegNetX006
	RegNetX008
	RegNetX016
	RegNetX032
	RegNetX064
	RegNetX080
	RegNetX120
	RegNetX320
<b>K</b>	RegNetY002
	RegNetY004
	RegNetY006
	RegNetY008
	RegNetY016
	RegNetY032
	RegNetY040
	RegNetY064
	RegNetY080
	RegNetY120
	RegNetY160

	RegNetY320
<b>L</b>	ResNet101
	ResNet152
	ResNet50
<b>M</b>	ResNetRS101
	ResNetRS152
	ResNetRS200
	ResNetRS270
	ResNetRS350
	ResNetRS420
	ResNetRS50
<b>N</b>	ResNet101V2
	ResNet152V2
	ResNet50V2
<b>O</b>	VGG16
	VGG19
	Xception