

## BINUS University

<b>Academic Career:</b> <i>Undergraduate / Master / Doctoral *)</i>		<b>Class Program:</b> <i>International/Regular/Smart Program/Global Class*)</i>	
<input type="checkbox"/> Mid Exam <input checked="" type="checkbox"/> Final Exam <input type="checkbox"/> Short Term Exam <input type="checkbox"/> Others Exam : _____		<b>Term : Odd/Even/Short *)</b>	
<input checked="" type="checkbox"/> Kemanggisan <input type="checkbox"/> Alam Sutera <input type="checkbox"/> Bekasi <input type="checkbox"/> Senayan <input type="checkbox"/> Bandung <input type="checkbox"/> Malang		<b>Academic Year :</b> <b><u>2024 / 2025</u></b>	
Faculty / Dept. : BGP / Magister Teknik Informatika		Deadline	Day / Date : Kamis / 20 February 2025 Time : 13.00 WIB
Code - Course : COMP8043041 – Machine Learning		Class	: LBA1, LCB1
Lecturer : Team Teaching		Exam Type	: Take Home Exam
□) <i>Strikethrough the unnecessary items</i>			
<b>The penalty for CHEATING is DROP OUT!!!</b>			

### Learning Outcomes :

- LO1: Able to explain the fundamental concept of Machine Learning and its various techniques/algorithms
- LO2: Describe the characteristics of various Machine Learning algorithms and understand how each of them works, including the mathematical principles underlying the algorithms.
- LO3: Apply relevant Machine Learning algorithms according to individual cases/problems and perform evaluation.
- LO4: Analyse the results obtained from Machine Learning experiments from several perspectives
- LO5: Able to propose suggestions to improve the system performance

#### 1. [LO1, LO2, LO3, LO4] [20%]

Untuk membantu devisi digital marketing, anda harus membuat model yang akan memprediksi apakah user yang telah mendapat program campaign akan memilih produk anda atau tidak. Anda akan bekerja dengan menggunakan data berikut:

<https://www.kaggle.com/datasets/rabieelkharoua/predict-conversion-in-digital-marketing-dataset>

Untuk membuat pemodelan maka anda perlu:

- a. (5%) Melakukan EDA dan pre pocessing data sesuai kebutuhan model
- b. (10%) Membangun model klasifikasi dengan menggunakan 3 metode Machine Learning dan anda perlu melakukan tuning hyperparameter minimal 2 hyperparameter untuk masing-masing algoritma machine learning.
- c. (5%) Jelaskan analisa anda mengenai perbandingan performance test data dari model yang dibuat, anda harus membuat summary hasil sebagai berikut:

Algoritma Machine learning	Machine learning Hyperparameter	Accuracy	Precision	Recall	F1 Score
Algoritma 1	xxx xxx	xxx xxx	xxx xxx	xxx xxx	xxx xxx
Algoritma 2	xxx xxx	xxx xxx	xxx xxx	xxx xxx	xxx xxx

Verified by,

Dr. Eng. Antoni Wibowo, S.Si., M.Kom., M.Eng (D5664) and sent to Department/Program on 24 January 2025

Algoritma 3	xxx	xxx	xxx	xxx	xxx
	xxx	xxx	xxx	xxx	xxx

## 2. [LO1, LO2, LO3, LO4] [30%]

Anda adalah seorang data scientist yang bekerja di perusahaan yang bergerak di bidang entertainment. Di akhir tahun 2024 anda ingin mengetahui gambaran lagu yang diputar di spotify berdasarkan data pada link:

<https://www.kaggle.com/datasets/nelgiriwithana/most-streamed-spotify-songs-2024>,

dan ingin melakukan analisa terhadap lagu-lagu tersebut. Untuk mempermudah analisa anda, anda bermaksud mengelompokkan terlebih dahulu. Untuk melakukan hal ini, anda perlu:

- (5%) Menjelaskan Teknik apa yang akan anda gunakan untuk melakukan pengelompokan ini, jelaskan alasan mengapa anda menggunakan teknik ini. Definisikan fitur yang anda gunakan untuk pengelompokan tersebut, berikan justifikasi mengapa fitur-fitur tersebut dipilih.
- (10%) Lakukan pre processing terhadap data sesuai kebutuhan model
- (15%) Bangunlah model pengelompokan beserta evaluasi dan analisa hasil model tersebut.

### Notes (no 1 dan 2):

- Jawaban terdiri dari code dan juga penjelasan, anda diizinkan untuk menggunakan file code yang berbeda untuk nomor yang berbeda
- Gunakan code cell untuk menerapkan algoritma atau kalkulasi.
- Gunakan Markdown cell untuk menuliskan penjelasan
- Jika diperlukan, anda dapat menyisipkan gambar dengan menggunakan fitur 'insert image' yang tersedia di Menu Edit.
- Anda bebas menggunakan library yang relevan yang disediakan oleh python

## 3. [LO1, LO3, LO4, LO5] [25%]

Diberikan data rekam medis pasien yang digunakan untuk deteksi awal abonormalitas fungsi jantung. Informasi terkait data diberikan dalam beberapa gambar sebagai berikut:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 920 entries, 0 to 919
Data columns (total 16 columns):
 #   Column        Non-Null Count  Dtype  
---  --
 0   id            920 non-null    int64  
 1   age           920 non-null    object  
 2   sex           920 non-null    object  
 3   dataset       920 non-null    object  
 4   cp            920 non-null    object  
 5   trestbps      861 non-null    float64 
 6   chol          890 non-null    float64 
 7   fbs           830 non-null    object  
 8   restecg       918 non-null    object  
 9   thalach       865 non-null    float64 
10   exang         865 non-null    object  
11   oldpeak       858 non-null    float64 
12   slope         611 non-null    object  
13   ca            309 non-null    float64 
14   thal          434 non-null    object  
15   num           920 non-null    int64  
dtypes: float64(5), int64(2), object(9)
memory usage: 115.1+ KB
```

	id	age	sex	dataset	cp	trestbps	chol	fbs	restecg	thalch	exang	oldpeak	slope	ca	thal	num
0	1	63	Male	Cleveland	typical angina	145.0	233.0	True	lv hypertrophy	150.0	False	2.3	downsloping	0.0	fixed defect	0
1	2	67	Male	Cleveland	asymptomatic	160.0	286.0	False	lv hypertrophy	108.0	True	1.5	flat	3.0	normal	2
2	3	67	Male	Cleveland	asymptomatic	120.0	229.0	False	lv hypertrophy	129.0	True	2.6	flat	2.0	reversible defect	1
3	4	37	Male	Cleveland	non-anginal	130.0	250.0	False	normal	187.0	False	3.5	downsloping	0.0	normal	0
4	5	41	Female	Cleveland	atypical angina	130.0	204.0	False	lv hypertrophy	172.0	False	1.4	upsloping	0.0	normal	0

```
df.isna().sum()
✓ 0.0s

id      0
age      0
sex      0
dataset  0
cp       0
trestbps 59
chol     30
fbs      90
restecg  2
thalch   55
exang    55
oldpeak  62
slope    309
ca       611
thal     486
num      0
dtype: int64
```

```
df[('num')].value_counts()
✓ 0.0s

num
0    411
1    265
2    244
Name: count, dtype: int64
```

Sumber data: UCI Machine Learning Data Repository-  
<https://archive.ics.uci.edu/dataset/45/heart+dis+ease>

Verified by,

Dr. Eng. Antoni Wibowo, S.Si., M.Kom., M.Eng (D5664) and sent to Department/Program on 24 January 2025

Jika hendak dilakukan pembangunan model prediksi, maka berdasarkan informasi data diatas

- A. **(15%)** Implementasikan proses *Pre-processing data* yang sesuai dengan kondisi data diatas. Berikan penjelasan pada setiap langkah *pre-processing data* yang dilakukan.  
(10 points)
- B. **(10%)** Jika dilakukan pembangunan model prediksi menggunakan Artificial Neural Network, lakukan analisa pada beberapa aspek berikut: (15 points)
  - a. Arsitektur ANN dengan memodifikasi jumlah hidden layer dan jumlah hidden nodes
  - b. Fungsi Aktivasi
  - c. Algoritma yang digunakan untuk optimasi
  - d. Learning rate

Berikan analisa anda terkait beberapa aspek diatas terhadap performa model prediksi yang dihasilkan

#### 4. [LO1, LO2, LO3, LO4] [25%]

Buka link dataset pada <https://archive.ics.uci.edu/dataset/145/statlog+heart>. Download dataset file 'heart.dat' dan dokumentasinya 'heart.doc'. Bacalah dokumentasinya terlebih dahulu terutama bagian attribute information, attribute types, dan variable to be predicted. Bagilah dataset tersebut menjadi 2 bagian menggunakan `train_test_split`, dengan random state 5 dan `test_size=0.2`.

- a. **(5%)** Jelaskan konsep feature engineering maupun selection yang menurut kalian perlu
- b. **(15%)** Tunjukkan hasil evaluasi menggunakan Decision Tree dan Naive Bayes Classifier (silahkan melakukan optimasi). Gunakan metric yang kalian anggap sesuai
- c. **(5%)** Berikan analisis dari hasil evaluasi performa dari kedua model klasifikasi tsb.

Gunakan Python untuk implementasi metode klasifikasi pada dataset tersebut. Berikan script kode Python, screenshot hasilnya.

-----oOo-----

Verified by,

Dr. Eng. Antoni Wibowo, S.Si., M.Kom., M.Eng (D5664) and sent to Department/Program on  
24 January 2025