BINUS University

Academic Career: Undergraduate / Master / Doctoral *)		Class Program: International/Regular/Smart Program/Global Class*)				
☐ Mid Exam ✓ Final Exam ☐ Short Term Exam ☐ Others Exam:		Term : Odd/ Even/Short *)				
✓ Kemanggisan □ Senayan	□ Alam Sutera □ Bekasi □ Bandung □ Malang	Academic Year : 2024 / 2025				
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	
Strikethrough the upper property of the prop	nnecessary items					
The penalty for CHEATING is DROP OUT!!!						

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:

 $\frac{https://www.kaggle.com/datasets/rabieelkharoua/predict-conversion-in-digital-marketing-dataset}{}$

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 Lerning 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,

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/nelgiriyewithana/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:

- a. (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.
- b. (10%) Lakukan pre processing terhadap data sesuai kebutuhan model
- c. (15%)Bangunlah model pengelompokan beserta evaluasi dan analisa hasil model tersebut.

Notes (no 1 dan 2):

- a. Jawaban terdiri dari code dan juga penjelasan, anda diizinkan untuk menggunakan file code yang berbeda untuk nomor yang berbeda
- b. Gunakan code cell untuk menerapkan algoritma atau kalkulasi.
- c. Gunakan Markdown cell untuk menuliskan penjelasan

restecg thalch

exang oldpeal

slope

dtype: int64

55

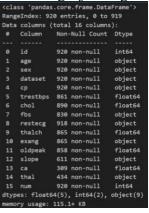
62

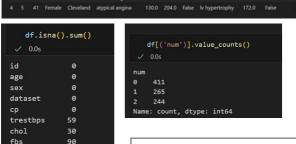
309

- d. Jika diperlukan, anda dapat menyisipkan gambar dengan menggunakan fitur 'insert image' yang tersedia di Menu Edit.
- e. Anda bebas menggunakan library yang relevan yang disediakan oleh python

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

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





Sumber data: UCI Machine Learning Data Repository-

 $\frac{https://archive.ics.uci.edu/dataset/45/heart+dis}{ease}$

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 pembangnan model prediksi menggunakan Artficial 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.

