

BINUS University

Academic Career: <i>Undergraduate / Master / Doctoral / International / BASE / BINUS Online*)</i>	Class Program: <i>Regular / Global Class*)</i>
<input checked="" type="checkbox"/> Mid Exam <input type="checkbox"/> Compact Term Exam <input type="checkbox"/> Final Exam <input type="checkbox"/> Others Exam : _____	Term : Odd / Even / Compact *) Period (Only for <i>BINUS Online</i>): 1 / 2 *)
<input checked="" type="checkbox"/> Kemanggisan <input type="checkbox"/> Senayan <input type="checkbox"/> Semarang <input type="checkbox"/> Alam Sutera <input type="checkbox"/> Bandung <input type="checkbox"/> Bekasi <input type="checkbox"/> Malang	Academic Year : 2024 / 2025
Exam Type* : Onsite / Online	Faculty / Dept. : School of Computer Science
Day / Date** : Friday, 24 th January 2025	Code - Course : COMP6577001— Machine Learning
Time** : 08.30 – 10.30 (120 Minutes)	Code - Lecturer : D6423 — Hidayaturrahman, S.Kom., M.T. D6673 — Bakti Amirul Jabar, S.Si., M.Kom. D6668 — Ade Putera Kemala, S.Kom., M.Kom.
Exam Specification*** : <input type="checkbox"/> Open Book <input type="checkbox"/> Open Notes <input type="checkbox"/> Close Book <input type="checkbox"/> Submit Project <input type="checkbox"/> Open E-Book <input type="checkbox"/> Oral Test	BULC (Only for : BINUS Online) Class : 11LA, 11LB, & 11LC (PPTI 17, 18, & 19)
Equipment*** : <input type="checkbox"/> Exam <input checked="" type="checkbox"/> Laptop <input type="checkbox"/> Drawing Paper – A3 Booklet <input type="checkbox"/> Tablet <input type="checkbox"/> Drawing Paper – A2 <input type="checkbox"/> Calculator <input type="checkbox"/> Smartphone <input type="checkbox"/> Notes <input type="checkbox"/> Dictionary	Student ID *** : Name *** : Signature *** :
*) <i>Strikethrough the unnecessary items</i> **) <i>For Online Exam, this is the due date</i> ***) <i>Only for Onsite Exam</i>	
<p><i>Please insert the test paper into the exam booklet and submit both papers after the test.</i></p> <p><i>The penalty for CHEATING is DROP OUT!</i></p>	

Learning Outcomes:

- LO 1: Explain the fundamental of machine learning concept
- LO 2: Interpret the distribution of dataset using regression method
- LO 3: Experiment classification and clustering algorithm from given dataset

INSTRUKSI UMUM

- Pastikan Anda memiliki Jupyter Notebook terinstal di perangkat Anda atau dapat diakses melalui platform online seperti Google Colab.
- Jawaban untuk soal essay harus ditulis dalam format markdown. Gunakan heading, poin-poin, atau tabel jika diperlukan.
- Jawaban berbentuk kode harus ditulis dalam cell kode yang telah disediakan atau ditambahkan jika diperlukan.
- Semua jawaban, baik essay maupun kode, harus dikerjakan dan disimpan dalam satu file Jupyter Notebook (.ipynb) tunggal.
- Dataset dapat diakses di link berikut : <https://shorturl.at/KEB83>
- Submit jawaban anda melalui form berikut : <https://forms.office.com/r/euFSfTS0nS>

1. **[LO 1, 2, & 3 – 30 Poin]** Dataset Auto-MPG adalah data yang berkaitan dengan konsumsi bahan bakar pada kendaraan bermotor, di mana variabel utama (target) adalah miles per gallon (MPG). Dataset ini terdiri atas 9 kolom, yaitu MPG (miles per gallon), Cylinders, Displacement, Horsepower, Weight, Acceleration, Model Year, Origin, dan Car Name. Dengan dataset ini, Anda ditantang untuk menganalisis dan memprediksi jumlah konsumsi bahan bakar sebuah kendaraan. Adapun contoh dataset sebagai berikut.

ID	mpg	cylinders	Displacement	Horsepower	weight	Acceleration	model year	origin	car name
0	18.0	8	307.0	130	3504	12.0	70	1	chevrolet chevelle malibu
1	15.0	8	350.0	165	3693	11.5	70	1	buick skylark 320
2	18.0	8	318.0	150	3436	11.0	70	1	plymouth satellite
3	16.0	8	304.0	150	3433	12.0	70	1	amc rebel sst
4	17.0	8	302.0	140	3449	10.5	70	1	ford torino

- a. **[LO 1, 2, & 3 – 10 Poin]** Gunakan data pada kolom Weight dan MPG untuk melakukan analisis hubungan antara berat kendaraan dan konsumsi bahan bakar. Jelaskan apakah hubungan tersebut positif, negatif, atau tidak ada hubungan sama sekali, serta deskripsikan kekuatan hubungannya.
- b. **[LO 1, 2, & 3 – 20 Poin]** Lakukan analisis prediksi nilai konsumsi bahan bakar (MPG) menggunakan dataset Auto-MPG dengan memilih tiga variabel yang paling relevan/paling penting, dengan alasan Anda memilih variabel tersebut. Split dataset menjadi training dan testing set dengan rasio 70:30, lalu kembangkan model Linear Regression dengan regularisasi L1 (Lasso) atau L2 (Ridge) untuk menghindari overfitting. Evaluasi model menggunakan metrik seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R-squared. Kemudian analisa dampak dari proses regularisasi.
2. **[LO 1, 2, & 3 – 50 Poin]** Anda adalah seorang karyawan baru yang mengisi posisi ML Engineer di perusahaan XYZ. Anda diminta untuk melanjutkan project yang sudah dikerjakan sebagian oleh senior anda sebelumnya. Silakan selesaikan tugas berikut menggunakan notebook (EDA_Insurance_Claim.ipynb) dan dataset

(*Klaim_asuransi_mobil.csv*) yang sudah diberikan. Anda akan bekerja untuk memprediksi apakah pengemudi telah mengajukan klaim asuransi mereka atau tidak (kolom *OUTCOME*).

- a. **[LO 1, 2, & 3 – 10 Poin]** Evaluasi EDA yang sudah dilakukan oleh pekerja sebelumnya. Berikan pendapat anda untuk masing-masing EDA yang sudah diberikan, apakah sudah tepat atau belum. Apabila ada yang kurang tepat berikan saran atas apa yang sudah mereka lakukan.
 - b. **[LO 1, 2, & 3 – 10 Poin]** Cek nilai yang hilang dalam DataFrame dan bersihkan data tersebut. Berikan penjelasan tentang metode pembersihan data yang Anda pilih!
 - c. **[LO 1, 2, & 3 – 10 Poin]** Periksa korelasi antar fitur dan pilih fitur yang akan Anda gunakan untuk membuat model klasifikasi. Jelaskan alasan Anda!
 - d. **[LO 1, 2, & 3 – 10 Poin]** Pilih algoritma klasifikasi (*Logistic Regression* atau *Naive Bayes*) dan buat model klasifikasi menggunakan 80% dataset sebagai data pelatihan.
 - e. **[LO 1, 2, & 3 – 10 Poin]** Evaluasi model menggunakan metrik *F1 Score*. Jelaskan hasil yang Anda peroleh!
3. **[LO 1 & 2 – 20 Poin]** Anda adalah seorang data scientist di MusicMatch, sebuah platform streaming musik yang ingin membangun sistem rekomendasi untuk meningkatkan pengalaman pengguna. Data yang dimiliki mencakup:
- Data Pengguna: User ID, usia, lokasi, dan genre favorit.
 - Data Lagu: Song ID, genre, popularitas artis (1-100), dan rating rata-rata lagu (1-5).
 - Data Interaksi: Rating pengguna terhadap lagu, jumlah pemutaran, dan riwayat pencarian.

Perusahaan sedang mempertimbangkan tiga pendekatan untuk sistem rekomendasi: content-based, collaborative filtering, dan hybrid model. Sistem rekomendasi yang dibangun harus mampu meningkatkan keterlibatan pengguna, mengatasi masalah cold-start untuk pengguna dan lagu baru, serta menangani masalah matriks utilitas yang jarang (sparse utility matrix). Jawab pertanyaan berikut ini:

- a) **[LO 1 & 2 – 5 Poin]** Jelaskan prinsip utama dari content-based, collaborative filtering, dan hybrid recommendation systems.
- b) **[LO 1 & 2 – 5 Poin]** Evaluasi kelebihan dan kekurangan dari masing-masing pendekatan di konteks platform streaming musik.
- c) **[LO 1 & 2 – 5 Poin]** Diskusikan bagaimana masing-masing metode dapat mengatasi tantangan seperti cold-start problem dan sparse utility matrix.
- d) **[LO 1 & 2 – 5 Poin]** Pilih metode yang paling sesuai untuk diterapkan di MusicMatch dan berikan alasan yang kuat berdasarkan kebutuhan platform dan jenis data yang tersedia.

--Selamat Mengerjakan--