

BINUS University

Academic Career: <i>Undergraduate / Master / Doctoral *)</i>		Class Program: <i>International/Regular/Smart Program/Global Class*)</i>	
<input checked="" type="checkbox"/> Mid Exam <input type="checkbox"/> Final Exam <input type="checkbox"/> Short Term Exam <input type="checkbox"/> Others Exam : _____		Term : Odd /Even/ Short *)	
<input checked="" type="checkbox"/> Kemanggisan <input checked="" type="checkbox"/> Alam Sutera <input checked="" type="checkbox"/> Bekasi <input type="checkbox"/> Senayan <input type="checkbox"/> Bandung <input type="checkbox"/> Malang		Academic Year : 2021 / 2022	
Faculty / Dept. : School of Computer Science		Deadline	Day / Date : Rabu / 27 April 2022 Time : 17.00
Code - Course : COMP6745001 – Machine Learning		Class : All Classes	
Lecturer : Team		Exam Type : Online	
*) <i>Strikethrough the unnecessary items</i>			
<i>The penalty for CHEATING is DROP OUT!!!</i>			

Learning Outcomes:

LO1 : Explain the fundamental of machine learning concept

LO2 : Interpret the distribution of dataset using regression method

Seluruh pertanyaan pada ujian ini memiliki beberapa cara atau pendekatan dalam menjawabnya. Silahkan gunakan cara atau pendekatan yang anda ketahui untuk menyelesaikan pertanyaan di bawah ini. Anda dapat menggunakan *library* python apapun yang tersedia untuk membantu jawaban anda. Bagi mahasiswa/i yang terbukti mencontek, maka akan dikenakan sanksi yang berlaku.

I. Esai (20%)

1. Ketika Ira bermain *game* catur melawan program komputer, probabilitas dia menang sebesar 0.6, probabilitas kalah sebesar 0.1, dan sebanyak 30% dari hasil permainan adalah seri. Ira bermain sebanyak 8 kali. Asumsikan hasil permainan *independent* satu sama lain.
 - a. **[LO1, 2 poin]** Apakah contoh kasus di atas termasuk *discrete random variable* atau *continuous random variable*?
 - b. **[LO1, 3 poin]** Dari kedelapan kali permainan tersebut, berapakah probabilitas Ira menang 4 kali, kalah 1 kali, dan seri 3 kali?
 - c. **[LO1, 3 poin]** Masih dari 8 kali permainan, berapakah probabilitas Ira menang setidaknya 5 kali?
 - d. **[LO1, 2 poin]** Hitung *expected value (mean)* dari kasus tersebut!
2. **[LO1, 10 poin]** Kota San Fransokyo merupakan sebuah kota teknologi tingkat tinggi terutama pada bidang robotik. Ada tiga perusahaan besar yang menguasai kota tersebut: Granville, Axcel, dan Hamada. Perusahaan Granville menguasai sekitar 50% dari total pembuatan robot di kota San Fransokyo; Perusahaan Axcel menguasai 35%; sedangkan sisanya dikuasai oleh perusahaan Hamada. Berdasarkan informasi yang kalian peroleh, ternyata 3% robot yang dibuat oleh Granville memiliki kendala pada kemampuan *sensing* robotnya. Sebanyak 5%

Verified by,

Hidayaturrahman (D6423) and sent to Program on April 1st, 2022

robot yang dibuat oleh Axcel dan 1% robot yang dibuat oleh Hamada juga memiliki kendala yang sama. Kalian ingin membeli robot untuk membantu usaha kalian dan secara acak mengambil tepat satu robot dari ketiga perusahaan tersebut. Setelah barang tersebut sampai, kalian melakukan *test*, dan ternyata robot tersebut mengalami malfungsi di komponen *sensing*-nya. Dengan menggunakan pendekatan **Bayes Rule**, perusahaan manakah yang diprediksi membuat robot yang Anda beli tersebut? Justifikasi jawaban Anda dengan perhitungan yang memadai!

II. Kasus (80%)

1. **[LO1, LO2, 30 poin]** Buka *file* data “**insurance-cost.csv**” yang diberikan pada *link* berikut ini: <https://bit.ly/3ugXXf2>. *Dataset* tersebut berisi profil para pemegang asuransi beserta nominal premi asuransinya. Anda diminta membangun model untuk memprediksi nominal biaya premi asuransi tersebut. Lakukan **Data Cleaning** dan **Visual Exploratory Data Analysis**, dengan detail yang dimaksud sebagai berikut:
 - a. Gunakan variable-variabel yang ditentukan berikut ini:
 - *age*,
 - *sex*,
 - *bmi*,
 - *children*,
 - *smoker*,
 - *region*,
 - *charges* (nominal premi asuransi).
 - b. Periksa *missing values* dan jika terdapat missing value, maka isilah *missing values* tersebut dengan ketentuan sebagai berikut:
 - Variabel yang bersifat *continuous* diisi dengan nilai rata-rata dari variabel.
 - Variabel yang bersifat *categorical* diisi dengan nilai modus dari variabel.
 - Pengisian *missing values* juga mempertimbangkan jenis kelamin. Artinya, jika nilai variabel yang kosong adalah milik pemegang asuransi berjenis kelamin pria, maka nilai modus atau *mean* hanya dihitung terhadap *samples* berjenis kelamin pria.
 - c. Buat visualisasi terhadap setiap *independent variable* dan jelaskan hubungan antar setiap *independent variables*.
 - d. Buat grafik *correlation* antara setiap *independent variable* dengan *dependent variable*. Lalu jelaskan masing-masing *correlation* tersebut!
2. **[LO1, LO2, 10 poin]** Buatlah **Training Set** dan **Test Set** dengan proporsi 4:1, lalu tampilkan **Training Set**.
3. **[LO1, LO2, 10 poin]** Buatlah permodelan **Linear Regression**.
 - a. Lakukan prediksi menggunakan *test set*.
 - b. Tampilkan hasil prediksi tersebut dan nilai sebenarnya dari *test set*.

Verified by,

Hidayaturrahman (D6423) and sent to Program on April 1st, 2022

4. **[LO1, LO2, 20 poin]** Lakukan evaluasi terhadap permodelan *Linear Regression*. Berikut adalah ketentuan yang harus dipenuhi:
 - a. Tampilkan persamaan regresi.
 - b. Buatlah evaluasi terhadap *coefficient* untuk setiap variabel.
 - c. Buatlah evaluasi dari hasil nilai prediksi menggunakan beberapa metrics evaluasi, seperti MAE, MSE, RMSE, dan R^2 .
 - d. Jelaskan hasil dari metrics evaluasi yang dilakukan pada tahapan sebelumnya.
5. **[LO1, LO2, 10 poin]** Buatlah plot yang sesuai untuk menggambarkan hasil dari ***Predicted Value vs Measure/Actual Value***

-- Selamat Mengerjakan --

Verified by,

Hidayaturrahman (D6423) and sent to Program on April 1st, 2022