Artificial Intelligence Course – 2021 Assignment 2

1. Supposed you are given a coin that lands the figure of Prof. Herman Johannes with probability a and the number of 100 rupiah with probability 1 – a. Are the outcomes of successive flips of the coin independent of each other given that you know the value of a? Are the outcomes of successive flips of the coin independent of each other if you do not know the value of a? explain you answer (5 points)
2. Suppose you saw an accident involving a car in Citra Land area. All cars in Citra Land are white or black. You are certain that the car that hit the victim was white. Extensive testing shows that, under the dim lighting condition, the different between white and black is 80% reliable.
3. Is it possible to calculate the most likely colour of the car? Explain your answer with mathematical procedure. (3 points)
4. What if you know that 9 out of 10 cars in Citra Land are black? (2 points)
5. In a smart home system, there is an alarm that senses when a temperature in the kitchen is too high. The dashboard measures the temperature of the kitchen. Consider Boolean variable A (alarm sounds), Fa (alarm is faulty), Fd(dashboard is faulty), and multi valued nodes D (dashboard reading) and K (actual kitchen temperature)
6. Draw a Bayesian network for this smart home system, given that the dashboard is more likely to fail when the kitchen temperature is too high. (5 points)
7. Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the dashboard gives the correct temperature is x when it is working, but y when it is faulty. Give conditional probability table associated with A. (5 points)
8. With your own language, explain the different between supervised and unsupervised learning. (5 points)
9. Identify the type of machine learning task (clustering/classification/regression) of the following case and explain your reason (10 point – 2.5 points for each)
10. Deciding whether a customer will continue the subscription of entertainment streaming provider.
11. Finding the number or people who will take a CPNS test in a certain ministry.
12. Identifying the factors for forecasting the number of corn harvest.
13. Identifying the segmentation of customer from internet provider based on their 24 hours activities.
14. What are the advantages (give at least two) of Reinforcement Learning? (5 points)
15. Based on Fuel Consumption data on the attributes ENGINESIZE, CYLINDERS, FUELCONSUMTION\_COMB with 80% training and 20% testing
16. Calculate MSE from the model for predicting emission (10 points)
17. What is the value for each attribute? (10 points)
18. Calculate the predicted emission when ENGINESIZE = 2, CYLINDERS = 5, and FUELCONSUMTION\_COMB = 3 (10 points)

8. Create an AI program using KNN algorithm that is able to predict whether a customer will complete a purchase. In order to do so, you need to complete the implementation of load\_data, train\_model, and evaluate in shopping.py

1. Load data (10 points) This function should accept the CSV filename and return a tuple (evidence, lables) where evidence should be the list of evidence for each data points, and labels should be a list of all of the labels for each data point. - Example of evidence: [0, 0.0, 0, 0.0, 1, 0.0, 0.2, 0.2, 0.0, 0.0, 1, 1, 1, 1, 1, 1, 0] - Example of label: 0
2. Train model (10 points) This function should accept a list of evidence and list of labels, return a nearestneighbour classifier fitted on that training data.
3. Evaluate (10 points) This function should accept a list of labels and a list of predictions and return two floating-point values (sensitivity, specificity). You may assume each label will be 1 for positive result or 0 for negative result.

**Jawaban**

**1.**

A = first outcome adalah Prof Herman.

B = second outcome adalah Prof Herman.

P(A) = a

P(B) = a

Jika A dan B adalah independent, maka P(A|B) = P(A)

Dimana:

= P(A) \* P(B)

**2.**

R = Warna asli mobil

S = Warna mobil yang terlihat

B = Hitam

W = Putih

**a.**

|  |  |  |
| --- | --- | --- |
| R | P(S = W) | P(S = B) |
| B | 0.2 | 0.8 |
| W | 0.8 | 0.2 |

P(R = W|S = W) = 0.8

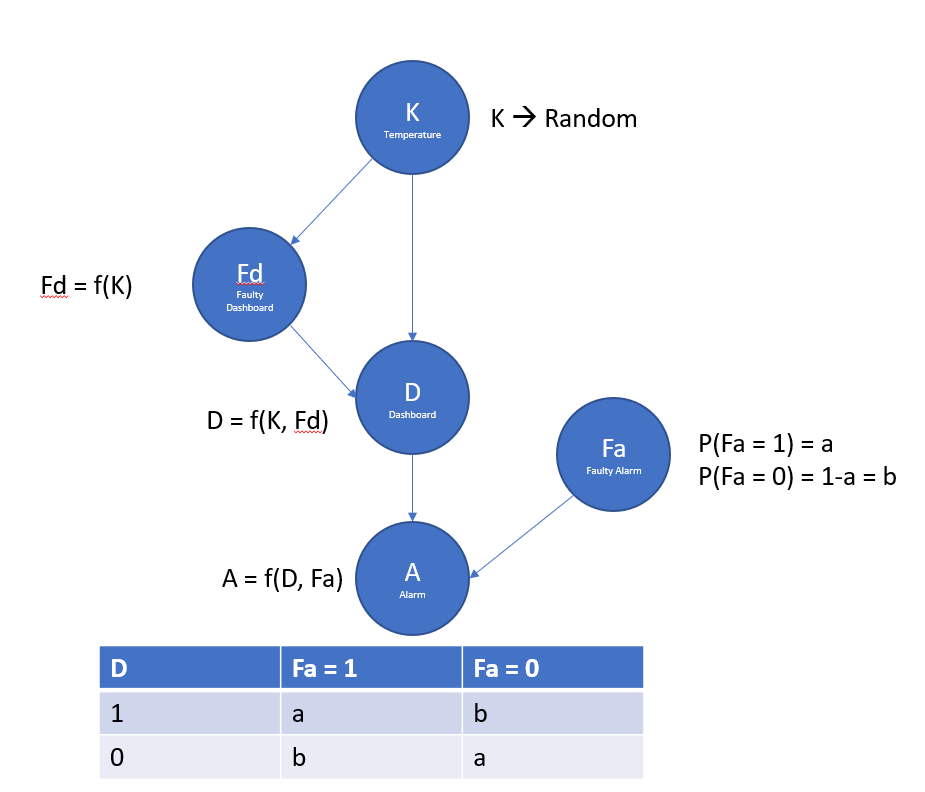
P(R = B|S = W) = 0.2

**b.**

P(R = W) = 0.1 dan P(R = B) = 0.9

P(R|S = W) = ?

**3.**

**a.** 

**b.**

x = P(Fd=1)

y = P(Fd=0)

a = P(Fa=1)

b = P(Fa=0)

|  |  |  |
| --- | --- | --- |
| K | A=1 🡪 Triggered | A=0 🡪 Not Triggered |
| 1 = High | x\*a + y\*b | x\*b + y\*a |
| 0 = Normal | x\*b + y\*a | x\*a + y\*b |

**4.**

|  |  |
| --- | --- |
| **Supervised Learning** | **Unsupervised Learning** |
| Proses Training Model dilakukan berdasarkan labeled data sebagai input karena memerlukan pengalaman sebelumnya sehingga membutuhkan supervisor. | Proses Training Model tidak memerlukan labeled data sebagai input karena akan mencari struktur tak berpola sehingga tidak membutuhkan supervisor. |
| Tujuannya untuk menemukan fungsi pemetaan untuk memetakan variabel input dengan variabel output. | Tujuannya untuk menemukan insight dari kumpulan volume data yang baru. |
| Dapat digunakan untuk kebutuhan Classification dan Regression. | Dapat digunakan terhadap permasalahan dimana hanya diketahui input datanya seperti Clustering dan Association. |
| Hasil model memiliki feedback dan dapat dibandingkan dengan label yang sudah dimiliki (source of ground truth). | Model tidak memberikan feedback apapun (hasil model tidak memiliki source of ground truth, maka sulit untuk melihat ketepatan sebuah model). |
| Computational Complexity nya sangat kompleks dengan Off-line Analysis dan jumlah Class yang diketahui. | Computational Complexity nya lebih sederhana dengan Real Time Analysis dan jumlah Class yang tidak diketahui. |
| Hasil akurasi output cenderung lebih bagus meskipun membutuhkan waktu Training yang lama. | Hasil akurasi output cenderung lebih buruk, namun hanya membutuhkan waktu Training yang lebih singkat. |

**5a.** Classification, karena jawaban keputusan pelanggan untuk melanjutkan ataupun memberhentikan berlangganan adalah berupa “ya” atau “tidak” (memiliki label).

**b.** Regression, karena memprediksi banyaknya calon peserta tes CPNS dalam bentuk angka.

**c.** Regression, karena memprediksi banyaknya panen jagung dalam bentuk angka.

**d.** Clustering, karena segmentasi termasuk ke Clustering sekaligus variasi aktivitas keseharian para pelanggan juga tidak memiliki label namun perlu dikelompokkan berdasarkan karakteristik yang serupa.

**6.**

* Proses Learning nya tidak memerlukan dataset, karena menggunakan experience serta trial and error yang diperoleh.
* Mampu meminimalisir terjadinya kesalahan (error) serupa yang berulang.
* Learning Model nya menyerupai cara pembelajaran yang dilakukan Manusia (Training yang telah diterima pada periode sebelumnya tidaklah terbuang percuma, tapi dapat menambah keakuratan dari Model). Selain itu, juga mampu mengolah permasalahan kompleks menjadi solusi yang tidak mampu dilakukan teknik konvensional (seperti Linear Regression, Decision Tree, K-Means).

7a. MSE =

b.

c. Predicted Emission = 154.10776862154998

8.