UNIVERSITI
TEKNOLOGI
PETRONAS

FINAL EXAMINATION
MAY 2024 SEMESTER

COURSE : **TEB3123/TFB3123 - MACHINE LEARNING**

DATE : **8 AUGUST 2024 (THURSDAY)**

TIME : **9:00 AM - 12:00 NOON (3 HOURS)**

INSTRUCTIONS TO CANDIDATES

1. Answer **ALL** questions in the Answer Booklet.

2. Begin **EACH** answer on a new page in the Answer Booklet.

3. Indicate clearly answers that are cancelled, if any.

4. Where applicable, show clearly steps taken in arriving at the solutions and indicate **ALL** assumptions, if any.

5. **DO NOT** open this Question Booklet until instructed.

Note :

    i. There are **EIGHT (8)** pages in this Question Booklet including the cover page

    ii. **DOUBLE-SIDED** Question Booklet.

1. a. The bias-variance tradeoff is a fundamental concept in supervised machine learning that refers to the tradeoff between the error introduced by bias and the error introduced by variance when training a machine learning model.

    i. Describe how the bias and variance change as the complexity of the model increases.

    [2 marks]

    ii. Illustrate the fundamental concept of bias-variance tradeoff using a graphical representation (bias-variance decomposition plot).

    [6 marks]

    b. For each of the listed descriptions below, state with justification whether the experimental set up is good or problematic.

    i. A project team proposed the use of dockers to create containers that encapsulate the machine learning model prior to deployment.

    [3 marks]

    ii. A project team dealing with image data proposed the use of data augmentation with the aim of avoiding model underfitting.

    [3 marks]

    iii. A project team proposed the use of early stopping with the aim of avoiding model overfitting.

    [3 marks]

    iv. A project team applied min-max scaler in a case study involving a highly chaotic dataset.

    [3 marks]

2. Autonomous driving requires complex decision-making in dynamic environments. Traditional rule-based systems struggle with the unpredictability of real-world driving scenarios. Reinforcement learning provides a solution by allowing systems to learn optimal driving strategies through trial and error, improving performance over time.

   a. Formulate the problem by identifying the agent, environment, state, action, penalty and reward for the above case study.

   [10 marks]

   b. Develop a theoretical Q- learning model approach and explain all processes involved.

   [6 marks]

   c. List **TWO (2)** advantages and disadvantages of reinforcement leaning in development of autonomous driving solutions.

   [4 marks]

3.   Consider a fictional dataset that describes the weather conditions for playing a game of golf. Given the weather conditions, each tuple classifies the conditions as fit("Yes") or unfit("No") for playing golf as shown in **FIGURE Q3**.

	Outlook	Temperature	Humidity	Windy	Play Golf
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

FIGURE Q3: Data Sample

    i.   Convert the outlook condition into a frequency table.

[3 marks]

    ii.   Create a Likelihood table for the weather conditions.

[3 marks]

    iii.   Use Naive Bayesian equation to calculate the posterior probability for each outlook condition.

[12 marks]

    iv.   Identify the **ONE (1)** advantage and **ONE (1)** disadvantage of developing such a model.

[2 marks]

4. a. Educational institutions face numerous challenges in administering online exams, such as ensuring a consistent testing environment, preventing cheating, and managing resources efficiently. Discuss how Docker containers can provide solutions in machine learning model deployment by creating isolated, reproducible environments for each student, ensuring consistency and security.

[5 marks]

b. Association rule mining is a rule-based machine learning method that is based on the concept that customer purchase behaviour has a pattern which can be exploited for selling more items to the customer in the future. **TABLE Q4** represents a sample of a transaction dataset.

TABLE Q4: Transaction dataset

TransactionId	Items
1	{A,C,D}
2	{B,C,D}
3	{A,B,C,D}
4	{B,D}
5	{A,B,C,D}

i. Each row in **TABLE Q4** consists of a transaction, determine the number of item sets, and calculate their corresponding support values.

[6 marks]

ii. Determine the confidence for each itemset.

[6 marks]

iii. With a user defined support threshold of 20%, generate all frequent item sets for the transaction dataset in **TABLE Q4.**

[3 marks]

6

5.   Image classification involves analysing pixel data in images to identify patterns and features that correspond to different categories. Traditional methods relied on handcrafted features and simple classifiers, but the advent of Convolutional Neural Networks (CNNs), has significantly improved accuracy and efficiency as shown in **FIGURE Q5**.
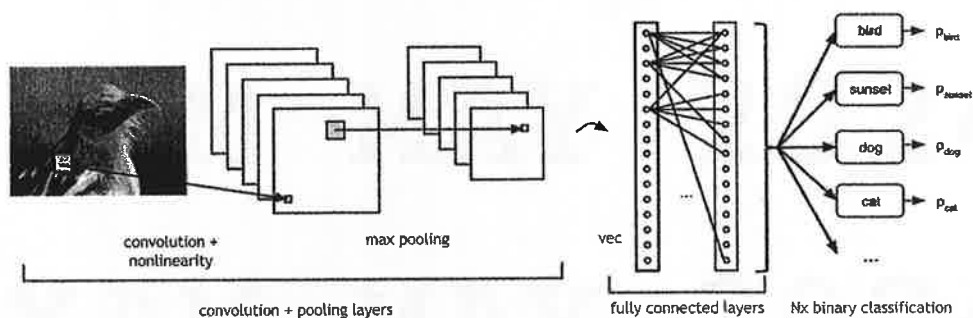

FIGURE Q5: Convolutional Neural Network

a.   Elaborate how data augmentation can be applied as a feature engineering option in image classification problems.

[2 marks]

b.   Discuss the functions of the following layers:

    i.   Convolution

[3 marks]

    ii.   Pooling with max and average polling

[3 marks]

    iii.   Fully connected

[3 marks]

c. Provided with the option of two optimizers (adam, rmsprop), dropout rate range (0 to 1), kernel size (3,5,7,9), epochs and batch size (8,16,32,64).

    i. Elaborate how grid search can be applied in hyper parameter tuning.

[3 marks]

    ii. Elaborate how random search can be applied in hyper parameter tuning.

[3 marks]

d. Explain how k-fold cross validation can be applied in evaluating the performance of machine learning models in general.

[3 marks]

- END OF PAPER -