



FINAL EXAMINATION
NOVEMBER 2023

COURSE TITLE	PREDICTIVE ANALYTICS
COURSE CODE	RBAN3243
DATE/DAY	16 FEBRUARY 2024 / FRIDAY
TIME/DURATION	09:00 AM - 11:00 AM / 02 Hour(s) 00 Minute(s)

INSTRUCTIONS TO CANDIDATES:

1. Please read the instruction under each section carefully.
2. Candidates are reminded not to bring into examination hall/room any form of written materials or electronic gadget except for stationery that is permitted by the Invigilator.
3. Students who are caught breaching the Examination Rules and Regulation will be charged with an academic dishonesty and if found guilty of the offence, the maximum penalty is expulsion from the University.

(This Question Paper consists of 10 Printed Pages including front page)

*****DO NOT OPEN THE QUESTION PAPER UNTIL YOU ARE TOLD TO DO SO*****

This question paper consists of TWO (2) sections. Answer ALL questions in the answer booklet provided. [100 MARKS]

SECTION A

(60 Marks)

There are THIRTY (30) questions in this part. Answer all questions.

1. "_____ is a field of study that gives computers the ability to learn without being explicitly programmed". - Arthur Samuel, 1959 The blank space above best described?
 - A. Machine Learning
 - B. Predictive Analysis
 - C. Exploratory Data Analysis
 - D. Statistic Analysis

2. Which one is **NOT** a Supervised Learning Algorithm?
 - A. K-mean clustering
 - B. K nearest neighbour
 - C. Decision Tree
 - D. Logistic Regression

3. "Some photo-hosting services, such as Google Photos, are good examples of this. Once you upload all your family photos to the service, it automatically recognizes that the same person A shows up in photos 1, 5, and 11, while another person B shows up in photos 2, 5, and 7. This is the unsupervised part of the algorithm (clustering). Now all the system needs is for you to tell it who these people are. Just add one label per person 4 and it is able to name everyone in every photo, which is useful for searching photo"

The statement above best describes?

 - A. Semi supervised-learning
 - B. Supervised Learning
 - C. Unsupervised Learning
 - D. Reinforcement Learning

4. Which one below is a dimensionality reduction algorithm?
- A. Principal Component Analysis (PCA)
 - B. Apriori
 - C. Hierarchical Cluster Analysis (HCA)
 - D. One-class SVM
5. What type of Machine Learning Algorithm Category for email spam filter?
- A. Supervised learning
 - B. Semi-supervised Learning
 - C. Unsupervised Learning
 - D. Reinforcement Learning
6. "is a method of changing weights based on the loss function for each data point".This statement above best describes?
- A. Gradient Descent
 - B. Cost Function
 - C. Learning Rate
 - D. Confusion Matrix
7. What will happen to a model if the learning rate is too high?
- A. The model fails to converge and jumps from good to bad cost optimizations
 - B. The model will take too long to converge to the minimum error
 - C. The model intercept value will increase and compromise the model's accuracy
 - D. The model parameters value will increase and compromise the model performance

8. The formula below is

$$1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} ?$$

- A. R-Squared or R2 Score
 - B. Mean Squared Error (MSE)
 - C. Root Mean Squared Error (RMSE)
 - D. Absolute Mean Error
9. Which algorithm below is used to solve a binary output problem?
- A. Logistic Regression
 - B. Linear Regression
 - C. Principal Component Analysis (PCA)
 - D. Decision Tree
10. What causes the bending curve behind Logistic Regression?
- A. Sigmoid Function
 - B. Cost Function
 - C. Gradient Descent
 - D. Confusion Matrix
11. What type of evaluation matrix is used to measure classification problems?
- A. Confusion Matrix
 - B. Sigmoid Function
 - C. Gradient Descent
 - D. Cost Function

12. Which Chart is able to help us to identify outliers?

- A. Boxplot
- B. Heatmap
- C. Line Chart
- D. Bar Chart

13. This chart helps us to see the correlation between variables?

- A. Heatmap
- B. Violin Plot
- C. Boxplot
- D. Histogram

14. Which one is **NOT** the right way to handle missing values of a dataset?

- A. Slice the dataset into train and test set
- B. Drop the row of the missing value
- C. Imputation using mean or median
- D. Predictive Filling

15. Which one of the python library is used for visualization?

- A. Seaborn
- B. Numpy
- C. Pandas
- D. Colab

16. Which one of the following python functions is used to get a statistical reading of our dataset (mean, median, percentile, and count)?

- A. `df.describe()`
- B. `df.info()`
- C. `df.value_counts()`
- D. `df.head()`

17. You have a 1000 rows of house price dataset (named 'df') which have 3 columns ('house type', 'year made' and 'selling price'). Which of the python pandas code below you can choose to get the number of transactions for each house type?

- A. `df['house type'].value_counts()`
- B. `df.groupby('Selling Price')['house type'].count()`
- C. `df['house type'].describe()`
- D. `df['house type', 'selling price'].info()`

18. Which one of below is **NOT** the right method to identify outliers?

- A. Imputation using ML Model KNN
- B. Box plot
- C. Violin Plot
- D. Z-Score method

19. "Process of deciding whether the numerical results quantifying hypothesized relationships between variables are acceptable as descriptions of the data".

The statement above best describes?

- A. a data validation process
- B. Confusion matrix process
- C. Finding root mean square error process
- D. Finding weight for parameter of the ML model process

20. "is the similarity between observations as a function of the time lag between them".

Which is the right answer for the statement above for time series analysis?

- A. Autocorrelation
- B. Seasonality
- C. Stationarity
- D. ARIMA

21. "is the process of reducing the number of input variables when developing a predictive model". This statement best describes?
- A. Feature Selection
 - B. Reduction in Variance
 - C. Information Gain
 - D. Linear Regression
22. "The branches depend on a number of factors. It splits data into branches like these till it achieves a threshold value". This statement best represents which algorithm?
- A. Decision Tree
 - B. Confusion Matrix
 - C. Decision Branching
 - D. Reinforcement Learning
23. Which one of the below is **NOT** a splitting algorithm for a decision tree?
- A. Decision Node
 - B. Information Gain
 - C. Chi-Square
 - D. Gini Impurity
24. Which one below is **NOT** the correct use for Linear Regression?
- A. Deciding between Male and Female of Titanic Dataset
 - B. House Price Analysis
 - C. Stock Market Prediction
 - D. Predict Future Revenue

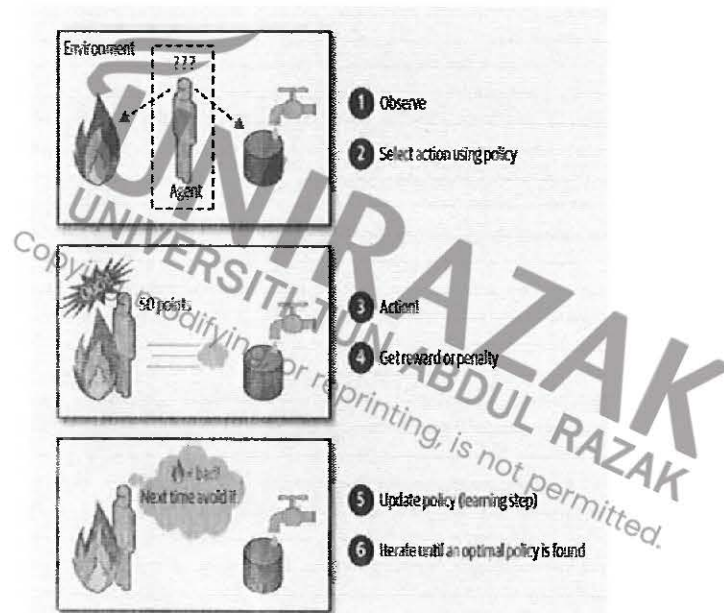
25. Based on this formula

$$y = w_0 + w_1x$$

What does w_0 stand for?

- A. Bias or intercept
- B. Coefficient
- C. Input variable
- D. Output variable

26. The image below best describes?



- A. Reinforcement Learning
- B. Supervised Learning
- C. Time Series Analysis
- D. Semi-Supervised Learning

27. 'Trend', 'Seasonality' & 'Autocorrelation'

The elements above best describe?

- A. Time Series Analysis
- B. Supervised Learning
- C. Semi-Supervised Learning
- D. Neural Network

28. Which one from below is **NOT** the purpose of Exploratory Data Analysis?

- A. Predicting the target output based on a set of inputs
- B. Identifying outliers, missing values, or human error
- C. Understanding Relation Between Variables
- D. Extracting important variables and leaving behind useless variables

29. Which one of the below is a strong case to use Time Series Analysis?

- A. Stock Market Prediction
- B. Email Spam Filter
- C. Gender Classification
- D. Image Labelling

30. Which of the python libraries below contain Machine Learning models?

- A. Sklearn
- B. Pandas
- C. Numpy
- D. Seaborn

UNIRAZAK
UNIVERSITI TUN ABDUL RAZAK
Copying, modifying, or reprinting, is not permitted.

SECTION B

(40 Marks)

There are **FOUR (4)** questions in this part. Answer All Questions

1. Explain The **SEVEN (7) Steps** in Preparing Machine Learning Models. (14 marks)
2. Name and explain the mechanism of **FIVE (5)** Machine Learning Models. (10 marks)
3. Name and explain **THREE (3)** common techniques used to deal with missing values. (6 marks)
4. Give **FIVE (5)** real-world applications for machine Learning Models. (10 marks)

UNIRAZAK
UNIVERSITI TUN ABDUL RAZAK
Copying, modifying, or reprinting, is not permitted.

***** END OF QUESTION PAPER *****