Mid-Term Exam

Due Mar 6 at 5pm **Points** 50 **Questions** 42
Available Mar 6 at 2pm - Mar 6 at 5pm 3 hours **Time Limit** 75 Minutes

Instructions

- The exam on **modules 1, 2, 3, 4, 5, and 6.**
- The exam will be available on **Monday March 06, 2023 from 2:00 PM to 4:30 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	75 minutes	36 out of 50 *

* Some questions not yet graded

ⓘ Correct answers will be available on Mar 7 at 5:30pm.

Score for this quiz: **36** out of 50 *
Submitted Mar 6 at 4:45pm
This attempt took 75 minutes.

First Part: MCQs

Question 1 1 / 1 pts

Information is transformed into _____to make decisions.

- Data
- Knowledge
- Information
- File

Incorrect

Question 2

0 / 1 pts

Which is not the Phase of data Mining Process

- Prediction and interpretation
- Feature Selection
- Data preprocessing
- Data Discarding

Question 3

1 / 1 pts

Data mining activities can be subdivided into two major investigation streams , which are:

- Forecast and Prediction
- Interpretation and Sampling
- Interpretation and Prediction
- Sampling and Forecast

Question 4

1 / 1 pts

Which of the statement is not true about Data Mining?

- The data mining process is based on inductive learning methods
- The term data mining refer to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired.
- Data mining activities can be subdivided into two major investigation streams, interpretation and prediction.
- Data mining analysis is to draw a fresh conclusion without investigating the past data, observations and interpretations

Question 5

1 / 1 pts

Which of the following is related to data mining?

- All the three
- Database technology
- Statistics
- Machine learning

Question 6

1 / 1 pts

Which of the following is not among functionalities (tasks) of data mining?

- Classification
- Visualization
- Clustering
- Association

Question 7

1 / 1 pts

_____ learning analyses are not guided by a target attribute.

- Guided
- Unguided
- Unsupervised
- Supervised

Question 8

1 / 1 pts

The formula dist (x_i, x_k)= $\sqrt[q]{\Sigma|x_{ij} - x_{kj}|^q}$ shows:

- Manhattan distance

- Cosine distance
- Minkowski distance
- Euclidean distance

Question 9

1 / 1 pts

Which attributes are categorical attributes without a natural ordering, such as the province of residence.

- Numerical
- Nominal
- Ordinal
- Ratio

Question 10

1 / 1 pts

Which of the following is a dimension reduction technique?

- Principal component analysis
- Stratified Sampling
- All the three
- Box plot

Question 11

1 / 1 pts

Data may contain erroneous or anomalous values, which are usually referred to as_____.

- Noise
- Reduction
- Outliers
- Inconsistencies

Question 12

1 / 1 pts

Continuous attributes are numerical attributes that assume an uncountable _____of values.

- zero
- first
- infinity
- non-zero

Question 13

1 / 1 pts

_____represent the real problem situations.

- Models

- Information
- Data
- Tools

Question 14

1 / 1 pts

Estimated procedures can become rather complex and time-consuming for a large dataset with a high percentage of_____.

- Training data
- testing data
- resulting data
- missing data

Question 15

1 / 1 pts

The confusion matrix for a binary classifier gives

- True negatives
- True Positives, true negatives
- True Positives, true negatives, false Positives, false negatives
- False Positives, false negatives

Question 16

1 / 1 pts

In weighted F-measure of precision and recall F(β), the value of β belongs to:

- [0,1]
- [-1, 1]
- [0, ∞) (selected)
- [0, 1)

Question 17

1 / 1 pts

Which of the following is correct formula for accuracy of classifier?

- Accuracy = (FP + FN)/All
- Accuracy = (TP + TN)/P
- Accuracy = (TP + TN)/All (selected)
- Accuracy = N/(TP + TN)

Question 18

1 / 1 pts

In data mining, what is the purpose of Interpretation?

- to determine useful patterns in the data
- to identify irregular patterns in the data

- (●) to express the rules and criteria for easy understanding
- ( ) All the three statements

Question 19

1 / 1 pts

The F-Measure is equal to zero if all the predictions are _____

- ( ) Correct
- ( ) Partially incorrect
- (●) Incorrect
- ( ) Partially correct

Question 20

1 / 1 pts

Training of the models is carried out using a sample of records extracted from the_____.

- ( ) Original dataset
- (●) Training dataset
- ( ) Result dataset
- ( ) Duplicate dataset

Question 21

1 / 1 pts

On which learning methods the Data Mining method is based?

- deductive learning methods
- comprehensive learning methods
- (selected) inductive learning methods
- basic learning methods

Question 22

1 / 1 pts

The precision is the proportion of _____ positive examples.

- (selected) Correctly classified
- Actually classified
- Occasionally classified
- Misclassified

Question 23

1 / 1 pts

---------- is an example of deterministic classifier, in which a classifier produces a discrete valued label.

- None of them
- (selected) Rule-based classifier
- Naïve Bayes classifier

Logistic regression

Question 24

1 / 1 pts

Naïve Bayes formula works well for-

- Association
- (selected) Classification
- Prediction
- Clustering

Incorrect

Question 25

0 / 1 pts

Why are Bayesian networks more capable for overfitting than the naïve Bayes classifier?

- All the three
- (selected) Due to its drawback of handling the presence of correlated attributes.
- Since they can provide complex forms of relationships.
- Because they usually required large datasets to be initiated.

Question 26

1 / 1 pts

Decision Trees or Association Rules are also called as?

- machine learning
- knowledge discovery in databases
- data mining
- All the three

Question 27

1 / 1 pts

A number of techniques originated in the field of computer science, such as decision trees or association rules, and are referred to as_____

- knowledge discovery in databases or deep learning
- machine learning or knowledge discovery in databases
- deep learning or machine learning
- machine learning or knowledge recovery in databases.

Question 28

1 / 1 pts

Which of the following is a basis of Naïve Bayes method?

- Conditional Probability
- Pie Chart
- Pivot Table
- Regression

Question 29

1 / 1 pts

All of the following steps are part of Naïve Bayes method except:

- Determine what classes they all belong to and which is more prevalent
- Find all the other records where the predictor values are same
- Express the probability as the product of p(x1|y) x p(x1|y) ... p(xn|y)
- Assign that class to the old record D.

Question 30

1 / 1 pts

-------- are the strategies, in which each record is covered by at least one rule.

- Exhaustive rules
- Not mutually exclusive rules
- Not exhaustive rules
- Mutually exclusive rules

Question 31

1 / 1 pts

When using the k- nearest neighbor classifier, what is the problem of choosing very small value of k?

- All the three
- The neighborhood may include points from other classes
- Misclassification rate will be very high.
- The classifier be capable of overfitting

Question 32

1 / 1 pts

One of the important characteristics of K-Nearest Neighbor Classifier is:

- They usually make their predictions based on local information
- They usually work well in the presence of irrelevant and redundant attributes
- These classifiers can handle the missing values
- All the three

Question 33

1 / 1 pts

In Logistic regression technique, input features can be

- Quantitative
- Quantitative and Qualitative
- Qualitative

- Only numeric

Question 34

1 / 1 pts

K- Nearest Neighbor Classifier is know as:

- Lazy learner
- All the three
- Instance-based learner
- Local classifier

Question 35

1 / 1 pts

Function which is used to bound the probability of x between 0 and 1?

- Cosine
- Log function
- Sine
- Sigmoid function

Question 36

1 / 1 pts

Which of the following methods do we use to best fit the data in Logistic Regression?

- Euclidean distance
- Maximum Likelihood
- Jaccard distance
- Least Square Error

Question 37

1 / 1 pts

When using K-Nearest Neighbors, you are required to have:

- A set of labeled records
- A proximity metric to compute the distance/similarity
- A number of nearest neighbors to retrieve
- All the three

Question 38

1 / 1 pts

Why data preprocessing is high recommended when using K-Nearest Neighbor Classifier?

- To let the classifier handling missing values in both the training and test sets

To avoid any situation, in which one of the attributes can dominate our distance measure

All the three

Proximity computations normally require the presence of all attributes

Second Part: Short Questions

Question 39

Not yet graded / 3 pts

Describe ***one data mining's issue*** that, in your view, may have a strong impact on the market and on society. Briefly, discuss ***how to approach such an issue***.

Your Answer:

One issue in data mining that could have a strong impact on the market and society is the potential for algorithmic bias. Algorithms can be biased if the data used to train them is not representative of the population or if the algorithm is programmed with biased assumptions. This can result in unfair or discriminatory outcomes , particularly in areas such as hiring, and criminal justice.

To approach this issue, one potential solution is to collect data from a range of sources and taking steps to mitigate any biases in the data collection process. Another approach is to use explainable AI techniques, which can help identify and correct any biases in the algorithms.

Additionally, it is important to have policies and regulations particularly in areas with significant social impact. This may involve establishing guidelines for the development and use of algorithms, as well as requiring

regular audits and evaluations to ensure that they are working as intended and not having unintended consequences.

Question 40

Not yet graded / 3 pts

Briefly, describe the steps involved in ***data mining*** when viewed as a process of ***knowledge discovery***.

Your Answer:

Data mining when viewed as a process of knowledge discovery involves several steps which are:

1. Data Selection: The first step is to select the appropriate dataset for analysis, which may involve gathering data from multiple sources, cleaning and preprocessing the data to ensure accuracy and consistency.
2. Data Preparation: This step involves transforming the data into a format that can be analyzed like normalizing the data.
3. Data Exploration: In this step, the data is explored using various techniques such as visualization, clustering, or association rule mining, to identify patterns and relationships in the data.
4. Modeling: This step involves developing predictive models using techniques such as decision trees, neural networks, or regression analysis, to identify trends and make predictions based on the data.
5. Evaluation: The models developed in the previous step are evaluated to ensure that they are accurate and reliable, using techniques such as cross-validation.
6. Deployment: The final step involves deploying the models and insights generated through the data mining process, to inform decision-making or improve business processes.

Question 41

Not yet graded / 3 pts

Consider a training set that contains **32 positive** examples and **224 negative**

examples. For the of the following candidate rule,

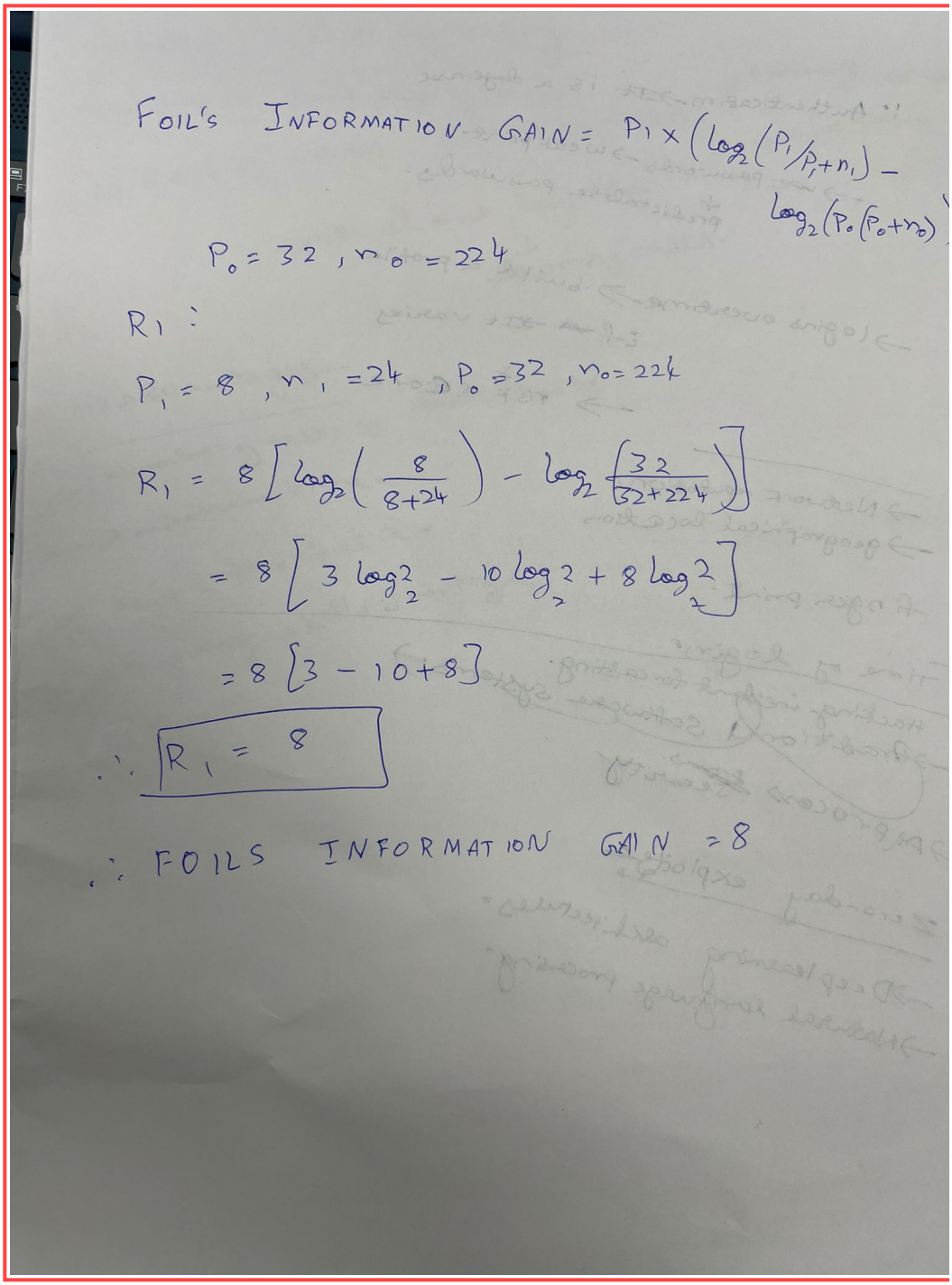
> ***R*1: *A* → + (covers 8 positive and 24 negative examples),**

Determine its FOIL's information gain.

(**Hints**: 1) You can type the logarithm of base 2 as lg.

2) lg (x/y) = lg x - lg y and lg (xy) = lg x + lg y)

Your Answer:



Question 42

Not yet graded / 3 pts

For the following data set given below, give specific examples of ***classification*** and ***clustering*** tasks that can be performed on the data. For each task, state how the ***data matrix*** should

be constructed (i.e., specify the rows and columns of the matrix).

- Stock market data, which include the prices and volumes of various stocks on different trading days.

Your Answer:

rows: stock market data price, stock market data volume.
columns: Prices and volume of the data

Attributes: Number of bedrooms, house area, locality, zip code

Quiz Score: **36** out of 50

Mid-Term Exam

Due Mar 6 at 5pm **Points** 50 **Questions** 42
Available Mar 6 at 2pm - Mar 6 at 5pm 3 hours **Time Limit** 75 Minutes

Instructions

- The exam on **modules 1, 2, 3, 4, 5, and 6.**
- The exam will be available on **Monday March 06, 2023 from 2:00 PM to 4:30 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	75 minutes	35 out of 50 *

* Some questions not yet graded

ⓘ Correct answers will be available on Mar 7 at 5:30pm.

Score for this quiz: **35** out of 50 *
Submitted Mar 6 at 4:45pm
This attempt took 75 minutes.

First Part: MCQs

Question 1 1 / 1 pts

Extraction of information and knowledge from data is known as:

- Data exploration
- Data mining
- Data Optimization
- Data mart

Question 2

1 / 1 pts

Which of the statement is not true about Data Mining?

- The term data mining refer to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired.
- Data mining activities can be subdivided into two major investigation streams, interpretation and prediction.
- The data mining process is based on inductive learning methods
- Data mining analysis is to draw a fresh conclusion without investigating the past data, observations and interpretations

Incorrect

Question 3

0 / 1 pts

Which of the following is not among alternative names of data mining?

- Knowledge extraction
- Business intelligence
- Knowledge discovery in databases
- Knowledge Acquisition

Question 4

1 / 1 pts

Which is not the Phase of data Mining Process

- Prediction and interpretation
- Feature Selection
- Data Discarding
- Data preprocessing

Question 5

1 / 1 pts

What is Data Mining?

- The setting up of queries to alert management when certain criteria are met.

The automated process of discovering patterns and relationships in an organization's data.

The capability to drill down into an organization's data once a question has been raised.

The process of performing trend analysis on the financial data of an organization.

Question 6

1 / 1 pts

_____ learning analyses are not guided by a target attribute.

- Unsupervised
- Unguided
- Supervised
- Guided

Question 7

1 / 1 pts

Data mining, the extraction of hidden ..... information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

- predictive

- provocative
- preventive
- proactive

Question 8

1 / 1 pts

Which attributes are numerical attributes that assume a finite number or a countable infinity of values.

- Countable
- Numerical
- Ordinal
- Discrete

Question 9

1 / 1 pts

Which attributes are categorical attributes without a natural ordering, such as the province of residence.

- Ordinal
- Numerical
- Nominal
- Ratio

Question 10

1 / 1 pts

_____represent the real problem situations.

- (•) Data
- ( ) Tools
- ( ) Models
- ( ) Information

Question 11

1 / 1 pts

**_____are categorical attributes in relation to which a specific property
can be only true or false.**

- ( ) Ordinal
- (•) Binary
- ( ) Nominal
- ( ) Numerical

Question 12

1 / 1 pts

Continuous attributes are numerical attributes that assume an uncountable _____of values.

- first
- zero
- non-zero
- infinity

Question 13

1 / 1 pts

The purpose of feature selection, also called _____.

- feature normalization
- feature compression
- feature denormalization
- feature reduction

Question 14

1 / 1 pts

Estimated procedures can become rather complex and time-consuming for a large dataset with a high percentage of_____.

- missing data
- Training data

- testing data
- resulting data

Question 15

1 / 1 pts

Typically, classification matrix considers:

- Predicted Class
- Actual Class and Predicted Class
- Actual Class
- Target class

Question 16

1 / 1 pts

The confusion matrix for a binary classifier gives

- True negatives
- False Positives, false negatives
- True Positives, true negatives
- True Positives, true negatives, false Positives, false negatives

Question 17

1 / 1 pts

For a record i, prediction error can be calculated as:

- (●) actual value - predicted value
- ( ) predicted value × actual value
- ( ) predicted value- actual value
- ( ) actual value+ predicted value

Question 18

1 / 1 pts

If the instances belongs to more than two classes then the classification is called as _____

- ( ) Double Classification
- ( ) Binary Classification
- (●) Multiclass Classification
- ( ) High Classification

Question 19

1 / 1 pts

Training of the models is carried out using a sample of records extracted from the_____.

- ( ) Result dataset
- ( ) Duplicate dataset

- Original dataset
- Training dataset

Question 20

1 / 1 pts

The precision is the proportion of _____ positive examples.

- Misclassified
- Actually classified
- Correctly classified
- Occasionally classified

Question 21

1 / 1 pts

The F-Measure is equal to zero if all the predictions are _____

- Correct
- Partially correct
- Partially incorrect
- Incorrect

Question 22

1 / 1 pts

In weighted F-measure of precision and recall F(β), the value of β belongs to:

- (●) [0, ∞)
- ( ) [-1, 1]
- ( ) [0, 1)
- ( ) [0,1]

Question 23

1 / 1 pts

Rule-based Classification models are used to generate _____that allow the target class of future examples to be predicted.

- (●) a set of rules
- ( ) a set of targeted results
- ( ) a set of misclassified variables
- ( ) a set of predicted variables

Incorrect

Question 24

0 / 1 pts

Why are Bayesian networks more capable for overfitting than the naïve Bayes classifier?

- ( ) Because they usually required large datasets to be initiated.

- All the three
- Due to its drawback of handling the presence of correlated attributes.
- Since they can provide complex forms of relationships.

Question 25

1 / 1 pts

Which of the following is an advantage of Naïve Bayes classifier?

- Good classification performance
- Computational efficiency
- All the three
- Simplicity

Question 26

1 / 1 pts

Which of the following is necessary to establish in Naïve Bayes method?

- Beta Value
- Alpha value
- Cut-off Probability
- Gamma Value

Question 27

1 / 1 pts

Decision Trees or Association Rules are also called as?

- data mining
- knowledge discovery in databases
- All the three
- machine learning

Question 28

1 / 1 pts

-------- are the strategies, in which each record is covered by at least one rule.

- Not exhaustive rules
- Exhaustive rules
- Mutually exclusive rules
- Not mutually exclusive rules

Question 29

1 / 1 pts

Which of the following is a basis of Naïve Bayes method?

- Pie Chart
- Regression

- Pivot Table
- Conditional Probability

Question 30

1 / 1 pts

---------- is an example of deterministic classifier, in which a classifier produces a discrete valued label.

- Naïve Bayes classifier
- None of them
- Rule-based classifier
- Logistic regression

Question 31

1 / 1 pts

When using the k- nearest neighbor classifier, what is the problem of choosing very small value of k?

- All the three
- Misclassification rate will be very high.
- The classifier be capable of overfitting
- The neighborhood may include points from other classes

Incorrect

Question 32

0 / 1 pts

How to determine the class label of a test example when using the K-Nearest Neighbor?

- Choose a right method for using class labels of K nearest neighbors to determine the class label of unknown record
- All the three
- Take the majority vote of class labels among the all k nearest neighbors
- Weight the vote according to distance to reduce the impact of K neighbors

Question 33

1 / 1 pts

Function which is used to bound the probability of x between 0 and 1?

- Log function
- Sine
- Cosine
- Sigmoid function

Question 34

1 / 1 pts

In Logistic regression technique, input features can be

- Quantitative
- Quantitative and Qualitative
- Only numeric
- Qualitative

Question 35

1 / 1 pts

Logistic regression is applicable for:

- Classification
- Clustering
- Association
- Prediction

Question 36

1 / 1 pts

Why are K-Nearest neighbor classifiers known as Lazy Learners?

- Because these classifiers are required to have a data preprocessing
- Since they are modeling the training data until it is needed to classify the test instances

Due its characteristics of producing decision boundaries of arbitrary shapes

All the three

Question 37

1 / 1 pts

When using K-Nearest Neighbors, you are required to have:

A number of nearest neighbors to retrieve

All the three

A set of labeled records

A proximity metric to compute the distance/similarity

Question 38

1 / 1 pts

Logistic regression is a _____ regression technique that is used to model data having a _____outcome

Nonlinear, binary

Nonlinear, numeric

Linear, binary

Linear, numeric

Second Part: Short Questions

Question 39

Not yet graded / 3 pts

Describe ***one data mining's issue*** that, in your view, may have a strong impact on the market and on society. Briefly, discuss ***how to approach such an issue***.

Your Answer:

One data mining issue that may have a strong impact on the market and society is the potential for algorithmic bias in decision-making. Algorithmic bias occurs when the data used to train a machine learning model is biased, leading to discriminatory or unfair outcomes. For example, an employment algorithm that is trained on data that disproportionately favors male candidates may result in discrimination against female candidates. This can lead to the perpetuation of existing biases and inequalities in society.

To address this issue, it is important to ensure that the data used to train machine learning models is diverse, representative, and unbiased. This can be done by carefully selecting the data sources and preprocessing the data to remove any biases. It is also important to regularly test and evaluate the model to identify any potential biases or unfair outcomes. Additionally, incorporating ethical and legal considerations into the design of machine learning systems can help to ensure that the algorithms are fair and transparent. Finally, having a diverse team of experts in the development and deployment of machine learning models can also help to mitigate potential biases and ensure that the algorithms are designed with fairness and equity in mind.

Question 40

Not yet graded / 3 pts

What do we mean by ***pruning*** the decision tree? Given a decision tree, you have the option of (a) converting the decision tree to rules and then pruning the resulting rules, or (b) pruning the decision tree and then converting the pruned tree to rules. What advantage does (a) have over (b)?

Your Answer:

Pruning the decision tree refers to the process of removing some of the branches or nodes from the tree in order to improve its performance on new, unseen data. The goal of pruning is to simplify the tree without sacrificing too much accuracy or predictive power. The two types are known as pre-pruning and post-pruning the decision tree. In pre-pruning, the construction of the decision tree stops. when it's getting too big and complex by setting a limit for decision tree depth. In post-pruning, some branches and nodes are removed after constructing the entire decision tree.

Approach a: This approach has the advantage of producing a set of rules that are easy to understand and interpret, which can be useful for explaining the decision-making process to non-experts.

Approach b: involves pruning the decision tree first and then converting the pruned tree to rules. While this approach can also lead to a simpler tree structure, it may result in a set of rules that are more difficult to understand or explain.

Overall, the advantage of approach (a) is that it produces a set of rules that are easier to understand and interpret, while still maintaining good predictive performance.

Question 41

Not yet graded / 3 pts

For each ***attribute*** given, classify its type as:

- discrete or continuous AND
- qualitative or quantitative AND

- nominal, ordinal, interval, or ratio

Indicate your reasoning if you think there may be some ambiguity in some cases.

Example: Age in years.

Answer: Discrete, quantitative, ratio.

A. Number of students enrolled in a class.
B. Daily user traffic volume at YouTube.com (i.e., number of daily visitors who visited the Web site).

Your Answer:

A. Number of students enrolled in a class :

Discrete, quantitative, ratio.

- Discrete: The number of students enrolled in a class is a whole number, and it cannot take on fractional or decimal values.
- Quantitative: It is a numerical value that represents a quantity or amount.
- Ratio: It has a true zero point, which means that it is meaningful to say that there are "zero" students in a class. Additionally, ratios between different values are meaningful (e.g., if one class has twice as many students as another class)

B. Daily user traffic volume :

Discrete, Quantitative, ratio

- Discrete: This is because the number of daily visitors can only take on certain distinct values, and cannot be measured as a continuous scale like time or temperature
- Quantitative: It is a numerical value that represents a quantity or amount.
- Ratio: It has a true zero point, which means that it is meaningful to say that there are "zero" daily visitors to the website. Additionally, ratios between different values are meaningful (e.g., if the traffic volume doubles, it means that the website has attracted twice as many users).

Question 42

Not yet graded / 3 pts

Suppose the fraction of undergraduate students who smoke is 15%, ***P*(*S*/*UG*)**, and the fraction of graduate students who smoke is 25%, ***P*(*S*/*G*)**. Assume that one-fifth of the college students are graduate, students, ***P*(*G*)**, and the rest are undergraduates, ***P(UG)***.

- Given ***P*(*G/S*) is 0.294.** Is a randomly chosen college student more likely to be a graduate or undergraduate student? (With assuming that the student is a smoker). (Hint: Using Bayesian Theorem, and Compare ***P*(*G/S*)** to ***P*(*UG/S*)**).

Your Answer:

Given,

P(s/UG) = 0.15 , p(s/G) = 0.25
P(G)= 1/5 p(UG) = 4/5

P(G/s)=P(s/G)*P(G)/P(s)
P(G/s) = (0.25 * 1/5)/ (0.25*1/5)+(4/5*0.15)
= 0.294
P(UG/s)= 1-0.2944 = 0.706

Undergraduates have the highest probability of being smokers

Quiz Score: **35** out of 50

Mid-Term Exam

Due Mar 6 at 5pm **Points** 50 **Questions** 42
Available Mar 6 at 2pm - Mar 6 at 5pm 3 hours **Time Limit** 75 Minutes

Instructions

- The exam on **modules 1, 2, 3, 4, 5, and 6.**
- The exam will be available on **Monday March 06, 2023 from 2:00 PM to 4:30 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	75 minutes	36 out of 50 *

* Some questions not yet graded

ⓘ Correct answers will be available on Mar 7 at 5:30pm.

Score for this quiz: **36** out of 50 *
Submitted Mar 6 at 4:45pm
This attempt took 75 minutes.

First Part: MCQs

Question 1 1 / 1 pts

Information is transformed into _____to make decisions.

- Data
- Knowledge
- Information
- File

Incorrect

Question 2

0 / 1 pts

Which is not the Phase of data Mining Process

- Prediction and interpretation
- Feature Selection
- Data preprocessing
- Data Discarding

Question 3

1 / 1 pts

Data mining activities can be subdivided into two major investigation streams , which are:

- Forecast and Prediction
- Interpretation and Sampling
- Interpretation and Prediction
- Sampling and Forecast

Question 4

1 / 1 pts

Which of the statement is not true about Data Mining?

- The data mining process is based on inductive learning methods
- The term data mining refer to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired.
- Data mining activities can be subdivided into two major investigation streams, interpretation and prediction.
- Data mining analysis is to draw a fresh conclusion without investigating the past data, observations and interpretations

Question 5

1 / 1 pts

Which of the following is related to data mining?

- All the three
- Database technology
- Statistics
- Machine learning

Question 6

1 / 1 pts

Which of the following is not among functionalities (tasks) of data mining?

- Classification
- Visualization
- Clustering
- Association

Question 7

1 / 1 pts

_____ learning analyses are not guided by a target attribute.

- Guided
- Unguided
- Unsupervised
- Supervised

Question 8

1 / 1 pts

The formula dist (x_i, x_k)= $\sqrt[q]{\Sigma|x_{ij} - x_{kj}|^q}$ shows:

- Manhattan distance

- Cosine distance
- Minkowski distance
- Euclidean distance

Question 9

1 / 1 pts

Which attributes are categorical attributes without a natural ordering, such as the province of residence.

- Numerical
- Nominal
- Ordinal
- Ratio

Question 10

1 / 1 pts

Which of the following is a dimension reduction technique?

- Principal component analysis
- Stratified Sampling
- All the three
- Box plot

Question 11

1 / 1 pts

Data may contain erroneous or anomalous values, which are usually referred to as_____.

- Noise
- Reduction
- Outliers
- Inconsistencies

Question 12

1 / 1 pts

Continuous attributes are numerical attributes that assume an uncountable _____of values.

- zero
- first
- infinity
- non-zero

Question 13

1 / 1 pts

_____represent the real problem situations.

- Models

- Information
- Data
- Tools

Question 14

1 / 1 pts

Estimated procedures can become rather complex and time-consuming for a large dataset with a high percentage of_____.

- Training data
- testing data
- resulting data
- missing data

Question 15

1 / 1 pts

The confusion matrix for a binary classifier gives

- True negatives
- True Positives, true negatives
- True Positives, true negatives, false Positives, false negatives
- False Positives, false negatives

Question 16

1 / 1 pts

In weighted F-measure of precision and recall F(β), the value of β belongs to:

- [0,1]
- [-1, 1]
- [0, ∞) (selected)
- [0, 1)

Question 17

1 / 1 pts

Which of the following is correct formula for accuracy of classifier?

- Accuracy = (FP + FN)/All
- Accuracy = (TP + TN)/P
- Accuracy = (TP + TN)/All (selected)
- Accuracy = N/(TP + TN)

Question 18

1 / 1 pts

In data mining, what is the purpose of Interpretation?

- to determine useful patterns in the data
- to identify irregular patterns in the data

- (●) to express the rules and criteria for easy understanding
- ( ) All the three statements

Question 19

1 / 1 pts

The F-Measure is equal to zero if all the predictions are _____

- ( ) Correct
- ( ) Partially incorrect
- (●) Incorrect
- ( ) Partially correct

Question 20

1 / 1 pts

Training of the models is carried out using a sample of records extracted from the_____.

- ( ) Original dataset
- (●) Training dataset
- ( ) Result dataset
- ( ) Duplicate dataset

Question 21

1 / 1 pts

On which learning methods the Data Mining method is based?

- deductive learning methods
- comprehensive learning methods
- (selected) inductive learning methods
- basic learning methods

Question 22

1 / 1 pts

The precision is the proportion of _____ positive examples.

- (selected) Correctly classified
- Actually classified
- Occasionally classified
- Misclassified

Question 23

1 / 1 pts

---------- is an example of deterministic classifier, in which a classifier produces a discrete valued label.

- None of them
- (selected) Rule-based classifier
- Naïve Bayes classifier

Logistic regression

Question 24

1 / 1 pts

Naïve Bayes formula works well for-

- Association
- (selected) Classification
- Prediction
- Clustering

Incorrect

Question 25

0 / 1 pts

Why are Bayesian networks more capable for overfitting than the naïve Bayes classifier?

- All the three
- (selected) Due to its drawback of handling the presence of correlated attributes.
- Since they can provide complex forms of relationships.
- Because they usually required large datasets to be initiated.

Question 26

1 / 1 pts

Decision Trees or Association Rules are also called as?

- machine learning
- knowledge discovery in databases
- data mining
- All the three

Question 27

1 / 1 pts

A number of techniques originated in the field of computer science, such as decision trees or association rules, and are referred to as_____

- knowledge discovery in databases or deep learning
- machine learning or knowledge discovery in databases
- deep learning or machine learning
- machine learning or knowledge recovery in databases.

Question 28

1 / 1 pts

Which of the following is a basis of Naïve Bayes method?

- Conditional Probability
- Pie Chart
- Pivot Table
- Regression

Question 29

1 / 1 pts

All of the following steps are part of Naïve Bayes method except:

- Determine what classes they all belong to and which is more prevalent
- Find all the other records where the predictor values are same
- Express the probability as the product of p(x1|y) x p(x1|y) ... p(xn|y)
- Assign that class to the old record D.

Question 30

1 / 1 pts

-------- are the strategies, in which each record is covered by at least one rule.

- Exhaustive rules
- Not mutually exclusive rules
- Not exhaustive rules
- Mutually exclusive rules

Question 31

1 / 1 pts

When using the k- nearest neighbor classifier, what is the problem of choosing very small value of k?

- All the three
- The neighborhood may include points from other classes
- Misclassification rate will be very high.
- The classifier be capable of overfitting

Question 32

1 / 1 pts

One of the important characteristics of K-Nearest Neighbor Classifier is:

- They usually make their predictions based on local information
- They usually work well in the presence of irrelevant and redundant attributes
- These classifiers can handle the missing values
- All the three

Question 33

1 / 1 pts

In Logistic regression technique, input features can be

- Quantitative
- Quantitative and Qualitative
- Qualitative

- Only numeric

Question 34

1 / 1 pts

K- Nearest Neighbor Classifier is know as:

- Lazy learner
- All the three
- Instance-based learner
- Local classifier

Question 35

1 / 1 pts

Function which is used to bound the probability of x between 0 and 1?

- Cosine
- Log function
- Sine
- Sigmoid function

Question 36

1 / 1 pts

Which of the following methods do we use to best fit the data in Logistic Regression?

- Euclidean distance
- Maximum Likelihood
- Jaccard distance
- Least Square Error

Question 37

1 / 1 pts

When using K-Nearest Neighbors, you are required to have:

- A set of labeled records
- A proximity metric to compute the distance/similarity
- A number of nearest neighbors to retrieve
- All the three

Question 38

1 / 1 pts

Why data preprocessing is high recommended when using K-Nearest Neighbor Classifier?

- To let the classifier handling missing values in both the training and test sets

To avoid any situation, in which one of the attributes can dominate our distance measure

All the three

Proximity computations normally require the presence of all attributes

Second Part: Short Questions

Question 39

Not yet graded / 3 pts

Describe ***one data mining's issue*** that, in your view, may have a strong impact on the market and on society. Briefly, discuss ***how to approach such an issue***.

Your Answer:

One issue in data mining that could have a strong impact on the market and society is the potential for algorithmic bias. Algorithms can be biased if the data used to train them is not representative of the population or if the algorithm is programmed with biased assumptions. This can result in unfair or discriminatory outcomes , particularly in areas such as hiring, and criminal justice.

To approach this issue, one potential solution is to collect data from a range of sources and taking steps to mitigate any biases in the data collection process. Another approach is to use explainable AI techniques, which can help identify and correct any biases in the algorithms.

Additionally, it is important to have policies and regulations particularly in areas with significant social impact. This may involve establishing guidelines for the development and use of algorithms, as well as requiring

regular audits and evaluations to ensure that they are working as intended and not having unintended consequences.

Question 40

Not yet graded / 3 pts

Briefly, describe the steps involved in ***data mining*** when viewed as a process of ***knowledge discovery***.

Your Answer:

Data mining when viewed as a process of knowledge discovery involves several steps which are:

1. Data Selection: The first step is to select the appropriate dataset for analysis, which may involve gathering data from multiple sources, cleaning and preprocessing the data to ensure accuracy and consistency.
2. Data Preparation: This step involves transforming the data into a format that can be analyzed like normalizing the data.
3. Data Exploration: In this step, the data is explored using various techniques such as visualization, clustering, or association rule mining, to identify patterns and relationships in the data.
4. Modeling: This step involves developing predictive models using techniques such as decision trees, neural networks, or regression analysis, to identify trends and make predictions based on the data.
5. Evaluation: The models developed in the previous step are evaluated to ensure that they are accurate and reliable, using techniques such as cross-validation.
6. Deployment: The final step involves deploying the models and insights generated through the data mining process, to inform decision-making or improve business processes.

Question 41

Not yet graded / 3 pts

Consider a training set that contains **32 positive** examples and **224 negative**

examples. For the of the following candidate rule,

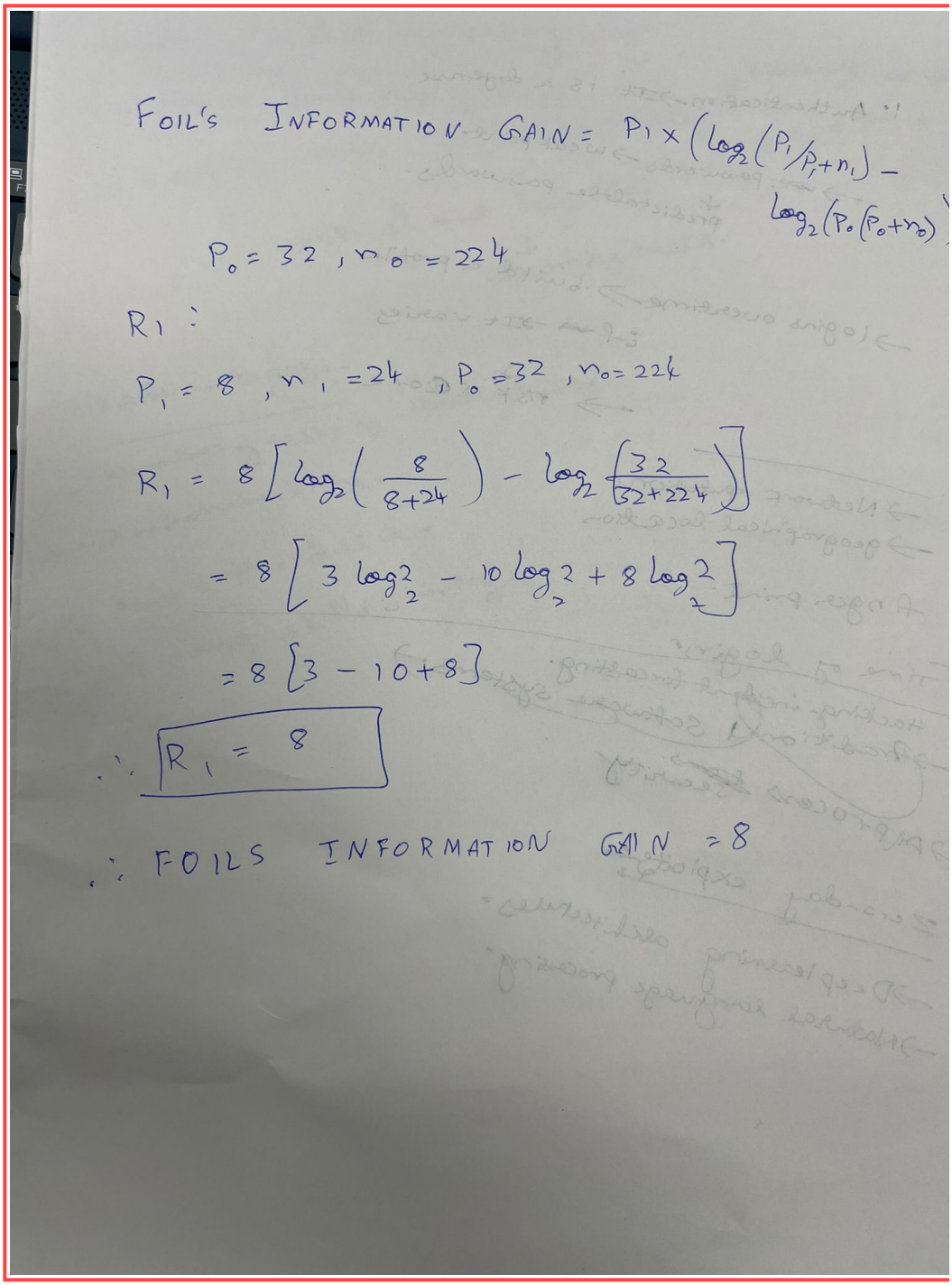
> ***R*1: *A* → + (covers 8 positive and 24 negative examples),**

Determine its FOIL's information gain.

(**Hints**: 1) You can type the logarithm of base 2 as lg.

2) lg (x/y) = lg x - lg y and lg (xy) = lg x + lg y)

Your Answer:



Question 42

Not yet graded / 3 pts

For the following data set given below, give specific examples of ***classification*** and ***clustering*** tasks that can be performed on the data. For each task, state how the ***data matrix*** should

be constructed (i.e., specify the rows and columns of the matrix).

- Stock market data, which include the prices and volumes of various stocks on different trading days.

Your Answer:

rows: stock market data price, stock market data volume.
columns: Prices and volume of the data

Attributes: Number of bedrooms, house area, locality, zip code

Quiz Score: **36** out of 50

Mid-Term Exam

Due Mar 6 at 5pm **Points** 50 **Questions** 42
Available Mar 6 at 2pm - Mar 6 at 5pm 3 hours **Time Limit** 75 Minutes

Instructions

- The exam on **modules 1, 2, 3, 4, 5, and 6.**
- The exam will be available on **Monday March 06, 2023 from 2:00 PM to 4:30 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	75 minutes	35 out of 50 *

* Some questions not yet graded

ⓘ Correct answers will be available on Mar 7 at 5:30pm.

Score for this quiz: **35** out of 50 *
Submitted Mar 6 at 4:45pm
This attempt took 75 minutes.

First Part: MCQs

Question 1 1 / 1 pts

Extraction of information and knowledge from data is known as:

- Data exploration
- Data mining
- Data Optimization
- Data mart

Question 2

1 / 1 pts

Which of the statement is not true about Data Mining?

- The term data mining refer to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired.
- Data mining activities can be subdivided into two major investigation streams, interpretation and prediction.
- The data mining process is based on inductive learning methods
- Data mining analysis is to draw a fresh conclusion without investigating the past data, observations and interpretations

Incorrect

Question 3

0 / 1 pts

Which of the following is not among alternative names of data mining?

- Knowledge extraction
- Business intelligence
- Knowledge discovery in databases
- Knowledge Acquisition

Question 4

1 / 1 pts

Which is not the Phase of data Mining Process

- Prediction and interpretation
- Feature Selection
- Data Discarding
- Data preprocessing

Question 5

1 / 1 pts

What is Data Mining?

- The setting up of queries to alert management when certain criteria are met.

The automated process of discovering patterns and relationships in an organization's data.

The capability to drill down into an organization's data once a question has been raised.

The process of performing trend analysis on the financial data of an organization.

Question 6

1 / 1 pts

_____ learning analyses are not guided by a target attribute.

- Unsupervised
- Unguided
- Supervised
- Guided

Question 7

1 / 1 pts

Data mining, the extraction of hidden ..... information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

- predictive

- provocative
- preventive
- proactive

Question 8

1 / 1 pts

Which attributes are numerical attributes that assume a finite number or a countable infinity of values.

- Countable
- Numerical
- Ordinal
- Discrete

Question 9

1 / 1 pts

Which attributes are categorical attributes without a natural ordering, such as the province of residence.

- Ordinal
- Numerical
- Nominal
- Ratio

Question 10

1 / 1 pts

_____represent the real problem situations.

- (•) Data
- ( ) Tools
- ( ) Models
- ( ) Information

Question 11

1 / 1 pts

**_____are categorical attributes in relation to which a specific property
can be only true or false.**

- ( ) Ordinal
- (•) Binary
- ( ) Nominal
- ( ) Numerical

Question 12

1 / 1 pts

Continuous attributes are numerical attributes that assume an uncountable _____of values.

- first
- zero
- non-zero
- infinity

Question 13

1 / 1 pts

The purpose of feature selection, also called _____.

- feature normalization
- feature compression
- feature denormalization
- feature reduction

Question 14

1 / 1 pts

Estimated procedures can become rather complex and time-consuming for a large dataset with a high percentage of_____.

- missing data
- Training data

- testing data
- resulting data

Question 15

1 / 1 pts

Typically, classification matrix considers:

- Predicted Class
- Actual Class and Predicted Class
- Actual Class
- Target class

Question 16

1 / 1 pts

The confusion matrix for a binary classifier gives

- True negatives
- False Positives, false negatives
- True Positives, true negatives
- True Positives, true negatives, false Positives, false negatives

Question 17

1 / 1 pts

For a record i, prediction error can be calculated as:

- (●) actual value - predicted value
- ( ) predicted value × actual value
- ( ) predicted value- actual value
- ( ) actual value+ predicted value

Question 18

1 / 1 pts

If the instances belongs to more than two classes then the classification is called as _____

- ( ) Double Classification
- ( ) Binary Classification
- (●) Multiclass Classification
- ( ) High Classification

Question 19

1 / 1 pts

Training of the models is carried out using a sample of records extracted from the_____.

- ( ) Result dataset
- ( ) Duplicate dataset

- Original dataset
- Training dataset

Question 20

1 / 1 pts

The precision is the proportion of _____ positive examples.

- Misclassified
- Actually classified
- Correctly classified
- Occasionally classified

Question 21

1 / 1 pts

The F-Measure is equal to zero if all the predictions are _____

- Correct
- Partially correct
- Partially incorrect
- Incorrect

Question 22

1 / 1 pts

In weighted F-measure of precision and recall F(β), the value of β belongs to:

- (●) [0, ∞)
- ( ) [-1, 1]
- ( ) [0, 1)
- ( ) [0,1]

Question 23

1 / 1 pts

Rule-based Classification models are used to generate _____that allow the target class of future examples to be predicted.

- (●) a set of rules
- ( ) a set of targeted results
- ( ) a set of misclassified variables
- ( ) a set of predicted variables

Incorrect

Question 24

0 / 1 pts

Why are Bayesian networks more capable for overfitting than the naïve Bayes classifier?

- ( ) Because they usually required large datasets to be initiated.

- All the three
- Due to its drawback of handling the presence of correlated attributes.
- Since they can provide complex forms of relationships.

Question 25

1 / 1 pts

Which of the following is an advantage of Naïve Bayes classifier?

- Good classification performance
- Computational efficiency
- All the three
- Simplicity

Question 26

1 / 1 pts

Which of the following is necessary to establish in Naïve Bayes method?

- Beta Value
- Alpha value
- Cut-off Probability
- Gamma Value

Question 27

1 / 1 pts

Decision Trees or Association Rules are also called as?

- data mining
- knowledge discovery in databases
- All the three
- machine learning

Question 28

1 / 1 pts

-------- are the strategies, in which each record is covered by at least one rule.

- Not exhaustive rules
- Exhaustive rules
- Mutually exclusive rules
- Not mutually exclusive rules

Question 29

1 / 1 pts

Which of the following is a basis of Naïve Bayes method?

- Pie Chart
- Regression

- Pivot Table
- Conditional Probability

Question 30

1 / 1 pts

---------- is an example of deterministic classifier, in which a classifier produces a discrete valued label.

- Naïve Bayes classifier
- None of them
- Rule-based classifier
- Logistic regression

Question 31

1 / 1 pts

When using the k- nearest neighbor classifier, what is the problem of choosing very small value of k?

- All the three
- Misclassification rate will be very high.
- The classifier be capable of overfitting
- The neighborhood may include points from other classes

Incorrect

Question 32

0 / 1 pts

How to determine the class label of a test example when using the K-Nearest Neighbor?

- Choose a right method for using class labels of K nearest neighbors to determine the class label of unknown record
- All the three
- Take the majority vote of class labels among the all k nearest neighbors
- Weight the vote according to distance to reduce the impact of K neighbors

Question 33

1 / 1 pts

Function which is used to bound the probability of x between 0 and 1?

- Log function
- Sine
- Cosine
- Sigmoid function

Question 34

1 / 1 pts

In Logistic regression technique, input features can be

- Quantitative
- Quantitative and Qualitative
- Only numeric
- Qualitative

Question 35

1 / 1 pts

Logistic regression is applicable for:

- Classification
- Clustering
- Association
- Prediction

Question 36

1 / 1 pts

Why are K-Nearest neighbor classifiers known as Lazy Learners?

- Because these classifiers are required to have a data preprocessing
- Since they are modeling the training data until it is needed to classify the test instances

Due its characteristics of producing decision boundaries of arbitrary shapes

All the three

Question 37

1 / 1 pts

When using K-Nearest Neighbors, you are required to have:

A number of nearest neighbors to retrieve

All the three

A set of labeled records

A proximity metric to compute the distance/similarity

Question 38

1 / 1 pts

Logistic regression is a _____ regression technique that is used to model data having a _____outcome

Nonlinear, binary

Nonlinear, numeric

Linear, binary

Linear, numeric

Second Part: Short Questions

Question 39

Not yet graded / 3 pts

Describe ***one data mining's issue*** that, in your view, may have a strong impact on the market and on society. Briefly, discuss ***how to approach such an issue***.

Your Answer:

One data mining issue that may have a strong impact on the market and society is the potential for algorithmic bias in decision-making. Algorithmic bias occurs when the data used to train a machine learning model is biased, leading to discriminatory or unfair outcomes. For example, an employment algorithm that is trained on data that disproportionately favors male candidates may result in discrimination against female candidates. This can lead to the perpetuation of existing biases and inequalities in society.

To address this issue, it is important to ensure that the data used to train machine learning models is diverse, representative, and unbiased. This can be done by carefully selecting the data sources and preprocessing the data to remove any biases. It is also important to regularly test and evaluate the model to identify any potential biases or unfair outcomes. Additionally, incorporating ethical and legal considerations into the design of machine learning systems can help to ensure that the algorithms are fair and transparent. Finally, having a diverse team of experts in the development and deployment of machine learning models can also help to mitigate potential biases and ensure that the algorithms are designed with fairness and equity in mind.

Question 40

Not yet graded / 3 pts

What do we mean by ***pruning*** the decision tree? Given a decision tree, you have the option of (a) converting the decision tree to rules and then pruning the resulting rules, or (b) pruning the decision tree and then converting the pruned tree to rules. What advantage does (a) have over (b)?

Your Answer:

Pruning the decision tree refers to the process of removing some of the branches or nodes from the tree in order to improve its performance on new, unseen data. The goal of pruning is to simplify the tree without sacrificing too much accuracy or predictive power. The two types are known as pre-pruning and post-pruning the decision tree. In pre-pruning, the construction of the decision tree stops. when it's getting too big and complex by setting a limit for decision tree depth. In post-pruning, some branches and nodes are removed after constructing the entire decision tree.

Approach a: This approach has the advantage of producing a set of rules that are easy to understand and interpret, which can be useful for explaining the decision-making process to non-experts.

Approach b: involves pruning the decision tree first and then converting the pruned tree to rules. While this approach can also lead to a simpler tree structure, it may result in a set of rules that are more difficult to understand or explain.

Overall, the advantage of approach (a) is that it produces a set of rules that are easier to understand and interpret, while still maintaining good predictive performance.

Question 41

Not yet graded / 3 pts

For each ***attribute*** given, classify its type as:

- discrete or continuous AND
- qualitative or quantitative AND

- nominal, ordinal, interval, or ratio

Indicate your reasoning if you think there may be some ambiguity in some cases.

Example: Age in years.

Answer: Discrete, quantitative, ratio.

A. Number of students enrolled in a class.
B. Daily user traffic volume at YouTube.com (i.e., number of daily visitors who visited the Web site).

Your Answer:

A. Number of students enrolled in a class :

Discrete, quantitative, ratio.

- Discrete: The number of students enrolled in a class is a whole number, and it cannot take on fractional or decimal values.
- Quantitative: It is a numerical value that represents a quantity or amount.
- Ratio: It has a true zero point, which means that it is meaningful to say that there are "zero" students in a class. Additionally, ratios between different values are meaningful (e.g., if one class has twice as many students as another class)

B. Daily user traffic volume :

Discrete, Quantitative, ratio

- Discrete: This is because the number of daily visitors can only take on certain distinct values, and cannot be measured as a continuous scale like time or temperature
- Quantitative: It is a numerical value that represents a quantity or amount.
- Ratio: It has a true zero point, which means that it is meaningful to say that there are "zero" daily visitors to the website. Additionally, ratios between different values are meaningful (e.g., if the traffic volume doubles, it means that the website has attracted twice as many users).

Question 42

Not yet graded / 3 pts

Suppose the fraction of undergraduate students who smoke is 15%, ***P*(*S*/*UG*)**, and the fraction of graduate students who smoke is 25%, ***P*(*S*/*G*)**. Assume that one-fifth of the college students are graduate, students, ***P*(*G*)**, and the rest are undergraduates, ***P(UG)***.

- Given ***P*(*G/S*) is 0.294.** Is a randomly chosen college student more likely to be a graduate or undergraduate student? (With assuming that the student is a smoker). (Hint: Using Bayesian Theorem, and Compare ***P*(*G/S*)** to ***P*(*UG/S*)**).

Your Answer:

Given,

P(s/UG) = 0.15 , p(s/G) = 0.25
P(G)= 1/5 p(UG) = 4/5

P(G/s)=P(s/G)*P(G)/P(s)
P(G/s) = (0.25 * 1/5)/ (0.25*1/5)+(4/5*0.15)
= 0.294
P(UG/s)= 1-0.2944 = 0.706

Undergraduates have the highest probability of being smokers

Quiz Score: **35** out of 50

Mid-Term Exam

Due Mar 6 at 5pm **Points** 50 **Questions** 42
Available Mar 6 at 2pm - Mar 6 at 5pm 3 hours **Time Limit** 75 Minutes

Instructions

- The exam on **modules 1, 2, 3, 4, 5, and 6.**
- The exam will be available on **Monday March 06, 2023 from 2:00 PM to 4:30 PM**.
- You need to answer **38 MCQs** with **1 point** for each + **4 Short questions** with **3 points** for each.
- You will have only **75 minutes** to complete your exam in **one sitting.**

Attempt History

	Attempt	Time	Score
LATEST	Attempt 1	75 minutes	35 out of 50 *

* Some questions not yet graded

ⓘ Correct answers will be available on Mar 7 at 5:30pm.

Score for this quiz: **35** out of 50 *
Submitted Mar 6 at 4:45pm
This attempt took 75 minutes.

First Part: MCQs

Question 1 1 / 1 pts

Which is not the Phase of data Mining Process

- Data preprocessing
- Feature Selection
- Data Discarding
- Prediction and interpretation

Incorrect

Question 2

0 / 1 pts

Which of the following is not among alternative names of data mining?

- Knowledge extraction
- Knowledge Acquisition
- Knowledge discovery in databases
- Business intelligence

Question 3

1 / 1 pts

Data mining, the extraction of hidden ..... information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses.

- predictive

- preventive
- provocative
- proactive

Question 4

1 / 1 pts

Which of the following is not among functionalities (tasks) of data mining?

- Clustering
- Visualization
- Classification
- Association

Question 5

1 / 1 pts

Which of the following is a new trend in data mining?

- Scalable data mining methods
- Web mining
- All the three
- Invisible data mining

Question 6

1 / 1 pts

What is Data Mining?

- The setting up of queries to alert management when certain criteria are met.
- The capability to drill down into an organization's data once a question has been raised.
- The automated process of discovering patterns and relationships in an organization's data.
- The process of performing trend analysis on the financial data of an organization.

Question 7

1 / 1 pts

Which of the following is related to data mining?

- Statistics
- Database technology
- Machine learning
- All the three

Question 8

1 / 1 pts

Continuous attributes are numerical attributes that assume an uncountable _____of values.

- non-zero
- infinity
- zero
- first

Question 9

1 / 1 pts

Which of the following is a dimension reduction technique?

- Principal component analysis
- Stratified Sampling
- Box plot
- All the three

Question 10

1 / 1 pts

The purpose of feature selection, also called _____.

- feature reduction

- feature denormalization
- feature compression
- feature normalization

Question 11

1 / 1 pts

Which attributes are categorical attributes without a natural ordering, such as the province of residence.

- Ratio
- Ordinal
- Nominal
- Numerical

Question 12

1 / 1 pts

_____represent the real problem situations.

- Information
- Models
- Tools
- Data

Question 13

1 / 1 pts

Estimated procedures can become rather complex and time-consuming for a large dataset with a high percentage of_____.

- resulting data
- missing data
- testing data
- Training data

Question 14

1 / 1 pts

Which of the following will be Euclidean Distance between the two data points A(1, 3) and B(2, 3)?

- 1
- 8
- 2
- 4

Question 15

1 / 1 pts

On which learning methods the Data Mining method is based?

- inductive learning methods

- basic learning methods
- deductive learning methods
- comprehensive learning methods

Question 16

1 / 1 pts

In weighted F-measure of precision and recall F(β), the value of β belongs to:

- [0, 1)
- [0,1]
- [0, ∞)
- [-1, 1]

Question 17

1 / 1 pts

The precision is the proportion of _____ positive examples.

- Correctly classified
- Occasionally classified
- Actually classified
- Misclassified

Question 18

1 / 1 pts

Typically, classification matrix considers:

- Target class
- Actual Class
- Predicted Class
- Actual Class and Predicted Class

Question 19

1 / 1 pts

The F-Measure is equal to zero if all the predictions are _____

- Partially correct
- Correct
- Incorrect
- Partially incorrect

Question 20

1 / 1 pts

The confusion matrix for a binary classifier gives

- False Positives, false negatives
- True negatives

- True Positives, true negatives, false Positives, false negatives
- True Positives, true negatives

Question 21

1 / 1 pts

For a record i, prediction error can be calculated as:

- actual value - predicted value
- predicted value- actual value
- predicted value × actual value
- actual value+ predicted value

Question 22

1 / 1 pts

What is classification process?

- predicts categorical class labels (discrete or nominal).
- Generalize, summarize, and contrast data characteristics.
- models continuous-valued functions.
- Predict some unknown or missing numerical values.

Question 23

1 / 1 pts

All of the following steps are part of Naïve Bayes method except:

- (●) Assign that class to the old record D.
- ( ) Express the probability as the product of p(x1|y) x p(x1|y) ... p(xn|y)
- ( ) Determine what classes they all belong to and which is more prevalent
- ( ) Find all the other records where the predictor values are same

Question 24

1 / 1 pts

Which of the following is necessary to establish in Naïve Bayes method?

- ( ) Gamma Value
- (●) Cut-off Probability
- ( ) Beta Value
- ( ) Alpha value

Question 25

1 / 1 pts

A number of techniques originated in the field of computer science, such as decision trees or association rules, and are referred to as_____

- ( ) machine learning or knowledge recovery in databases.
- ( ) knowledge discovery in databases or deep learning

- machine learning or knowledge discovery in databases
- deep learning or machine learning

Question 26

1 / 1 pts

Why are Bayesian networks more capable for overfitting than the naïve Bayes classifier?

- Because they usually required large datasets to be initiated.
- Since they can provide complex forms of relationships.
- All the three
- Due to its drawback of handling the presence of correlated attributes.

Question 27

1 / 1 pts

Which of the following is not true for Bayes model for classification?

- Naïve Bayes classifiers are highly scalable
- All the records are used instead of relying on just the matching records
- Predictors should also be categorical
- Numerical variables need not to be converted into categorical

Incorrect

Question 28

0 / 1 pts

In building a rule-based classifier, ------- use a function called One Rule function.

- C4.5rules algorithm
- Indirect method
- All the three
- FOIL

Question 29

1 / 1 pts

Rule-based Classification models are used to generate _____that allow the target class of future examples to be predicted.

- a set of targeted results
- a set of rules
- a set of misclassified variables
- a set of predicted variables

Question 30

1 / 1 pts

Which of the following is an advantage of Naïve Bayes classifier?

- Simplicity
- All the three

- Computational efficiency
- Good classification performance

Question 31

1 / 1 pts

Which of the following methods do we use to best fit the data in Logistic Regression?

- (•) Maximum Likelihood
- Jaccard distance
- Euclidean distance
- Least Square Error

Incorrect

Question 32

0 / 1 pts

One of the important characteristics of Logistic Regression is:

- (•) Understands the relationships between attributes and class labels
- Can handle interacting attributes
- All the three
- Works more robustly even in high dimensional settings

Question 33

1 / 1 pts

When using K-Nearest Neighbors, you are required to have:

- A proximity metric to compute the distance/similarity
- A set of labeled records
- A number of nearest neighbors to retrieve
- (selected) All the three

Question 34

1 / 1 pts

Why data preprocessing is high recommended when using K-Nearest Neighbor Classifier?

- All the three
- Proximity computations normally require the presence of all attributes
- To let the classifier handling missing values in both the training and test sets
- (selected) To avoid any situation, in which one of the attributes can dominate our distance measure

Question 35

1 / 1 pts

One of the important characteristics of K-Nearest Neighbor Classifier is:

These classifiers can handle the missing values

They usually work well in the presence of irrelevant and redundant attributes

They usually make their predictions based on local information

All the three

Question 36

1 / 1 pts

K- Nearest Neighbor Classifier is know as:

Local classifier

All the three

Instance-based learner

Lazy learner

Question 37

1 / 1 pts

When using the k- nearest neighbor classifier, what is the problem of choosing very small value of k?

The classifier be capable of overfitting

All the three

Misclassification rate will be very high.

- The neighborhood may include points from other classes

Question 38

1 / 1 pts

Logistic regression is a _____ regression technique that is used to model data having a _____outcome

- Linear, binary
- Linear, numeric
- Nonlinear, binary
- Nonlinear, numeric

Second Part: Short Questions

Question 39

Not yet graded / 3 pts

In data mining, what are the similarities and the differences between ***classification*** *and* ***clustering*** tasks?

Your Answer:

Similarities:

Both classification and clustering are used to discover patterns or structures in data and require a set of features or attributes to represent the data

Also, both these tasks can be applied to a wide range of domains and applications, including marketing, healthcare, finance, etc.

Differences:

Classification is a supervised learning task, where the goal is to predict the class label of a new instance based on its features. In contrast, clustering is an unsupervised learning task, where the goal is to group similar instances together based on their features.

Also, in classification, the class labels of the training data are known and used to train a model that can generalize to new data. In clustering, there are no predefined classes, and the algorithm groups the data into clusters based on the similarity between the instances.

Question 40

Not yet graded / 3 pts

Briefly, describe the steps involved in ***data mining*** when viewed as a process of ***knowledge discovery***.

Your Answer:

Data mining can be viewed as a process of knowledge discovery, which involves several steps or phases. Here is a brief overview of the common steps involved in the data mining process:

1. Problem Definition: In this step, the problem is identified, and the goals and objectives of the data mining project are determined.
2. Data Collection: in this step we collect and the data from various sources, here we also determine the quality and quantity of the data needed to achieve the goals of the project
3. Data Preparation: Data preparation involves cleaning, transforming, and pre-processing the data to ensure it is in a format suitable for analysis. This step also includes dealing with missing data.
4. Data Exploration: Data exploration involves visualizing and summarizing the data to identify patterns, trends, and relationships

that may be of interest.

5. Data Modeling: This step includes selecting an appropriate algorithm, training the model on the data, and validating the model to ensure it generalizes well to new data.
6. Evaluation: This step includes measuring the accuracy, precision, recall, and other performance metrics of the model.
7. Deployment: Deployment involves integrating the model into the business or research process, and putting the results of the data mining project into action.

Question 41

Not yet graded / 3 pts

For the following data set given below, give specific examples of ***classification*** and ***clustering*** tasks that can be performed on the data. For each task, state how the ***data matrix*** should be constructed (i.e., specify the rows and columns of the matrix).

- Stock market data, which include the prices and volumes of various stocks on different trading days.

Your Answer:

Classification tasks:

1. Predicting stock price movement: A classification task can be performed on this data to predict whether the stock price of a particular company will rise or fall on a particular trading day. The data matrix for this task should consist of rows representing different trading days and columns representing different stocks. Industry classification: Another classification task that can be performed on this data is to classify stocks into different industries based on their prices and trading volumes. The data matrix for this task should consist of rows representing different stocks and columns representing different attributes such as industry, price, and volume.

Clustering tasks:

1. Stock similarity clustering: Clustering can be performed on this data to group stocks that have similar price and volume patterns over a period of time. The data matrix for this task should consist of rows representing different trading days and columns representing different stocks.
2. Time-series clustering: Clustering can also be performed on this data to group trading days that have similar price and volume patterns across all stocks. The data matrix for this task should consist of rows representing different trading days and columns representing different attributes such as stock price and volume.

Question 42

Not yet graded / 3 pts

Consider a training set that contains **32 positive** examples and **224 negative**

examples. For the of the following candidate rule,

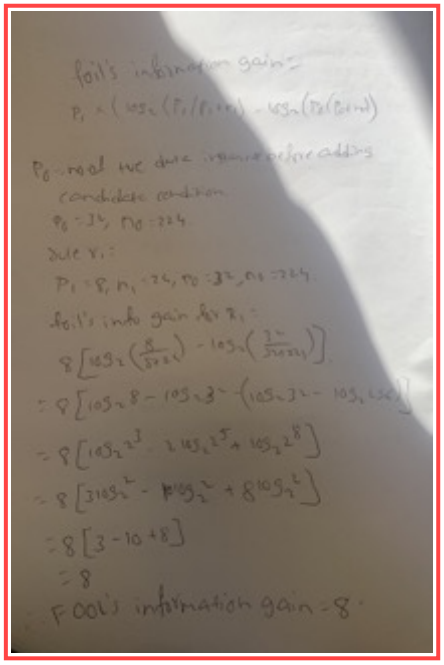
> ***R*1: *A* → + (covers 8 positive and 24 negative examples),**

Determine its FOIL's information gain.

(**Hints**: 1) You can type the logarithm of base 2 as lg.

2) lg (x/y) = lg x - lg y and lg (xy) = lg x + lg y)

Your Answer:



Quiz Score: **35** out of 50

B.N.N College,Bhiwandi
Department Of Information Technology
Subject: Business Intelligence
Questions Bank

UNIT-01	
	1. Business intelligence system provides tools and methodologies to knowledge workers to help them to take_____. A. Effective decision. B. Timely decision C. Both 1 and 2. D. Efficient Decision. Answer: C
	2. _____ is the outcome of extraction and processing activities carried out on data. A. Knowledge B. Information C. Data D. Wisdom Answer: B
	3. The objective of B.I is A. To support decision-making and complex problem solving. B. To support information gathering. C. To support data collection. D. To support data analysis. Answer: A
	4. Which of the following is not a component of business intelligence analysis cycle? A. Analysis B. Insight C. Decision D. Design Answer: D

	5. In BI Architecture, It is used to gather and integrate the data stored in various primary and secondary sources. A. Data Warehouse B. Data mart C. Data Sources D. Data explosion. Answer: C
	6. Extraction of information and knowledge from data is known as A. Data mining B. Optimisation C. Data exploration D. Data mart Answer: A
	7. Which phases is used for planning of Development of a business intelligence system. A. Analysis and Design B. Planning C. Implementation and Control D. Maintenance Answer: B
	8. Decision making process is of _____ phases. A. Three B. Five C. Two D. Six Answer: B
	9. Well defined and recurring decision making procedure is called A. Structured B. Semi-structured C. Operational

	D. Unstructured Answer: A
	10. In _____ approach, a decision maker considers economic, tactical legal, ethical, procedural and political factors. A. Absolute rationality approach B. Bounded rationality approach C. Rational approach D. Un-Bounded rationality approach Answer: C
	11. a) Data mart is a subset of data warehouse. b) Data marts contain repositories of summarized data collected for analysis on a specific section or unit within an organization. A. Only a is correct B. Only b is correct C. Both are correct D. Both are wrong. Answer: C
	12. In ETL 'E' stands for A. External B. Extraction C. Extreme D. Extra-ordinal Answer: B
	13. DSS stands for: A. Decision Support System. B. Definition support System.

	C. Day–to-day System Support. D. Data storage system. Answer: A
	14. _____represent the real problem situations. A. Data B. Models C. Tools D. Information Answer: A
	15. During the _____ phase, additional data conversion occurs to performed to obtain the summaries that will reduce the response time. A. Loading. B. Extraction. C. Transformation D. Performance Evaluation. Answer: C
	16. Optimization is: A. Determine the best solution.Successful marketing approaches to achieve the optimum outcome. B. Un-Successful marketing approaches to achieve the optimum outcome. C. Getting the greatest return for the least expenditure of time, effort, manpower. D. Determine the Worst solution. Answer: A
	17 . (a)A decision support system helps in decision making but does not necessarily give a decision itself. (b) decision support system is a computer-based application that

	collects organizes and analyses business data to facilitate quality business decision making for management, operations, and planning. A. only a is correct. B. only b is correct C. both are correct. D. both are wrong. Answer: C
	18. _____measurements express the level of conformity of a given system to the objectives for which it was designed. A. Effectiveness B. Efficiency C. Evaluation D. Feedback Answer: A
	19. _____is the first stage in developing in decision support system. A. Analysis B. Design C. Knowledge Acquisition D. Planning Answer: C
	20. Data by itself is not useful unless A. It is massive B. It is processed to obtain information C. It is collected from diverse sources D. It is properly stated Answer: B

	22. Decision support systems are used for A. Management decision making B. Providing tactical information to management C. Providing strategic information to management D. Better operation of an organization Answer: C
	23. Business intelligence (BI) is a broad category of application programs which includes: A. Decision support,Datamining,OLTP B. Decision support,Datacollection,OLAP C. Decision Expert,Datamining,OLAP D. Decision support,Datamining,OLAP Answer: D
	24. Decision support systems are used by A. Line managers. B. Top-level managers. C. Middle level managers. D. System users ANSWER: B
	25. Which of following is not phase of decision making process A. Design B. Analysis C. Intelligence D. Choice ANSWER: B
	26. _____is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. A. best practice

	B. data mart C. business information warehouse D. business intelligence ANSWER: D
	27. Organizations have hierarchical structures because A. it is convenient to do so B. it is done by every organization C. specific responsibilities can be assigned for each level D. it provides opportunities for promotions ANSWER: C
	28. Strategic information is needed for A. Day to day operations B. Meet government requirements C. Long range planning D. Short range planning ANSWER: C
	29. Decision support systems are essential for A. Day–to-day operation of an organization. B. Providing statutory information. C. Top level strategic decision making. D. Ensuring that organizations are profitable. ANSWER: C
	30. Computer information system are most successful in providing information for A. Control decisions B. Planning decision C. Strategic decision D. Design decision

	ANSWER: A
	31. Close System Cycle defined as A. Which is System cycle does not need any output. B. Able to modify their own output flows based on feedback. C. Able to modify their own output flows based on process. D. Able to modify their own output without any process. ANSWER: B
	32. Information is transformed into _____to make decisions. A. Data B. Information C. Knowledge D. File ANSWER: C
	33. Mathematical models are developed by A. doing mathematical calculation B. exploring the relationships with data. C. developing mathematical logic D. exploiting the relationships among system control variables ,parameters and evaluation metrics. ANSWER: D
	34. BI supports _____ for data. A. Files B. Data Marts & Data Warehouse. C. Data Mining D. System database ANSWER: B
	35. The fourth phase of the BI cycle involves performances measurement_____ A. Design B. Evaluation C. Intelligence D. Graph ANSWER:B

	36. _____technologies enabling factors that have facilitate development of BI with complex organizations A. Application B. System C. business information D. Hardware and Software ANSWER: D
	37. Phases are in Decision making process A. Intelligence, Design, Choice, Implementation, Control. B. Design, Develop, Apply C. Data, Process, Develop, Implement, Control D. Search, Sort, Design, Develop, Implement ANSWER: A
	38. Decisions can be classified into their _____ A. concept and thoughts B. Supply and Demand C. Nature and Scope D. Techniques ANSWER: C
	39. Knowledge Management increase in the _____ A. no. of options considered. B. no. of difficulties considered C. no. of risks considered D. no. of models considered ANSWER: A
	40. Business intelligence allows making _____ and timely decisions. A. Effective B. Correct C. Right D. Beneficiary ANSWER: A
	41. It is necessary to recognize and accurately spell out the problem in _____ phase in the Cycle of a business intelligence analysis A. Analysis

	B. Decision C. Insight D. Evaluation ANSWER: A
	42. ETL process consists of Extract, _____ and Load A. Translate B. Transmit C. Track D. Transform ANSWER: D
	43. Analysis, Design, Planning and _____ are the Phases in the development of a business intelligence system A. Implementation B. Control C. Implementation and control D. Coding ANSWER: C
	44. The evaluation metrics can be categorized into two main classes - effectiveness and _____. A. Perfection B. Performance C. Concreteness D. Correctness ANSWER: B

	45. Depending on scope, decisions can be classified as strategic, tactical and _____ A. Accurate B. Sequential C. Operational D. Procedural ANSWER: C
	46. A decision support system has been defined as an interactive computer system helping decision makers to combine data and _____ to solve semi-structured and unstructured problems. A. Information B. Model C. Knowledge D. Experience ANSWER: B
	47. A DSS must be _____ and adaptable A. Flexible B. Adjustable C. Changeable D. Knowledgeable ANSWER : A
	48. Usage of data by _____organizations that is improper and does not respect the individuals' right to privacy should not be tolerated A. Public B. Private C. Public and Private D. Government ANSWER : C
	49. _____ decisions affect only parts of an enterprise and are usually restricted to a single department. A. Strategic B. Operational C. Ethical

	D. Tactical ANSWER : D
	50. _____ decisions refer to specific activities carried out within an organization and have a modest impact on the future. A. Strategic B. Operational C. Ethical D. Tactical ANSWER : B
	51.What is MDX: A. (MDX) is a query language for online transection processing (OLTP) using a database management system. B. (MDX) is a query language for online analytical processing (OLAP) using a database management system. C. (MDX) is a query language for online analytical processing (OLAP) using a power BI. D. (MDX) is a structure language for online analytical processing (OLAP) using a database management system. Answer: B
	52. Fullform of MDX: A. Multidimensional Expressions B. Multidata Expressions C. Multidirect Expressions D. Multidimensional Exponential Answer:A

	53. Data warehouse is also known as _____ A. Enterprise data warehouse B. Enter data warehouse C. Extra data warehouse D. Easy data warehouse Answer: A
	54. Full form of OLAP: A. Online analytical process B. Online analog process C. Online anal process D. Online analogy process Answer: A
	55. What is explain by given figure:  A. Features of business intelligence B. Features of power BI C. Features of Data warehouse

	D. Features of data explosion Answer:B
Unit -02	
	Q. 1 What is a model? A. a selective abstraction of real world B. a selective imagination of 1st world C. a selective proposal of real world D. a selective example of second word Correct answer: A
	Q. 2 A material representation of a real system, whose behaviour is imitated for the purpose of the analysis is called as ? A. Analogical Model B. Iconic Model C. Symbolic Model D. Static Model Correct answer: B
	Q. 3 In which model some input information represents random events characterized by a probability distribution? A. Stochastic B. Iconic C. Symbolic D. Static Correct answer: A
	Q.4 What is allowed by Sensitivity and Scenario analyses be assessed ? A. the robustness of optimal decisions from variations in the input parameters. B. the robustness of optimal decisions to variations into input parameters. C. the robustness of optimal decisions to variations in the input data. D. the robustness of optimal decisions to variations in the input parameters. Correct answer: D

	Q. 5 Which model observes the status of a system only at the beginning or at the end of discrete intervals ? A. Discrete-time dynamic models B. Static-time dynamic models C. Symbolic -time dynamic models D. Iconic-time dynamic models Correct answer: A
	Q. 6 The purpose of explanatory models is to functionally identify a possible relationshipbetween a dependent variable and _____ ? A. a set of positive attributes B. a set of negative attributes C. a set of neutral attributes D. a set of independent attributes Correct answer: D
	Q. 7 Which is one of the primary objective of mathematical models ? A. to identify regular patterns in the data B. to identify irregular patterns in the data C. to identify negative patterns in the data D. to identify neutral patterns in the data Correct answer: A
	Q.8 The conceptual paradigm outlined determines a wide and popular class of mathematical models for decision making, represented by _____models A. optimization models B. stochastic models C. supervised models D. iconic models Correct answer: A
	Q.9 What is the aim of Data Mining? A. extracting information and knowledge B. useful for knowledge workers in decision making C. extracting raw data D. Both A & B Correct answer: D

	Q.10 On which learning methods the Data Mining method is based? A. inductive learning methods B. deductive learning methods C. basic learning methods D. comprehensive learning methods Correct answer: A
	Q.11 What is the purpose of Interpretation? A. to identify regular patterns in the data B. to express the rules and criteria for easy understanding C. to identify irregular patterns in the data D. Both A & B Correct answer: D
	Q.12 Classification Trees or Association Rules are also called as? A. machine learning B. knowledge discovery in databases C. deep learning D. A & B Correct answer: D
	Q.13 Which is the last Phases of mathematical models for decision making (a) Problem Identification (b) Implementation and Testing (c) Model Formation (d) Development of Algorithm Answer b
	Q.14 Which mathematical model aims at understand the mechanisms that regulate the development of intelligence, ability to extract knowledge from past experience in order to apply it in the future. a. Risk analysis models b. Optimization models c. Pattern recognition Models d. Waiting line models Answer C
	Q.15 In which Mathematical mode the decision maker is required to choose among a number of available alternatives. a. Risk analysis models b. Optimization models

	c. Pattern recognition Models d. Waiting line models Answer A
	Q.16 Which of the statement is not true about Data Mining? a)The term data mining refer to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge acquired. b) Data mining analysis is to draw a fresh conclusion without investigating the past data, observations and interpretations c) Data mining activities can be subdivided into two major investigation streams, interpretation and prediction. d) The data mining process is based on inductive learning methods Answer b
	Q.17 Which is not the Phase of data Mining Process A. Data Gathering B. Selection of Attributes C. Prediction and interpretation D. Data Discarding Answer D
	Q.18 Data Inception Means (a) inspection of each missing value (b) identify missing values (c) replacement of missing Data (d) discard all records Answer a
	Q.19 Data Elimination Means (a) inspection of each missing value (b) identify missing values (c) replacement of missing Data (d) discard all records Answer d

	Q.20 Data mining activities can be subdivided into two major investigation streams , which are a. Interpretation and Sampling b. Interpretation and Prediction. c. Forecast and Prediction d. Forecast and Interpretation Answer b.
	Q.21 Which is the Application of Data Mining . a. Fraud Detection b. Risk Analysis c. Both a & b d. Only b Answer C
	Q.22 The conceptual paradigm outlined determines a wide and popular class of mathematical models for decision making, represented by optimization models. a. Project management model b. learning model c. Predictive model D.optimization model Answer d
	Q.23 A number of techniques originated in the field of computer science, such as classification trees or association rules, and are referred to as_____ a. machine learning or knowledge discovery in databases b. Deep learning c. machine learning or knowledge recovery in databases. d. Deep learning or knowledge discovery in databases. Answer a
	Q.24 _____methods are based on the contributions of various disciplines, such as business organization, behavioral psychology and operations research. a. Project management model b. learning model c. Predictive model D.optimization model Answer a

	Q.25 A significant proportion of the models used in business intelligence systems, such _____models, require input data concerned with future events. a. Project management model b. learning model c. Predictive model D.optimization model Answer c
	Q.26 _____ learning analyses are not guided by a target attribute. a. Supervised b. Guided c. Unguided d. Unsupervised Answer d
	Q.27 Once a mathematical model has been defined, one will naturally wish to proceed with its solution to assess decisions and to select the best alternative. a. Development of algorithms b. Mathematical relationships c. Numerical parameters d. Evaluation criteria Answer a
	Q.28 _____are categorical attributes in relation to which a specific property can be true or false. a.Counts b.Nominal c.Numerical d.Ordinal Answer: a
	Q.29 Which attributes are categorical attributes without a natural order-ing, such as the province of residence. a.Counts b.Nominal c.Numerical d.Ordinal Answer: b
	Q.30 which attributes are numerical attributes that assume a finite number or a countable infinity of values. A.counts B. Discrete c.Numerical d.Ordinal

	Answer: b
	Q.31 The purpose of a data mining process is some-times to provide a simple and concise representation of the information stored in a _____. A.small dataset b.large dataset. C. numeric dataset D.stored dataset Answer: b
	Q.32 It is possible to discard ____ records for which the values of one or more attributes are missing. A.All b.some C. combined D. half Answer: a
	Q.33 Most learning models benefit from a preventive standardization of the data, also called_____. A.normalization b.de-normalize C. data segregation D. data prevention Answer: A
	Q.34 Transformations of this kind are usually referred to as feature _____. a. Transformations b. Extraction c. Scaling d. Standardization Answer:b
	Q.35 The use of such methods will be explained within the classification methods called _____. A K-means B K-nearest C. support vector machines D. Normalization Answer: c

	Q.36 The purpose of feature selection, also called _____. a. feature reduction b. feature compression c. feature normalization d. feature denormalization Answer: a
	Q.37 An _____model is a material representation of a real system, whose behavior is imitated for the purpose of the analysis. A.Iconic B.b.Analogical C..Symbolic D.Stochastic Answer:a
	Q.38 A model is called _____when all input data are supposed to be known a priori and with certainty. A.iconic B.analogical c.deterministic D.static Answer: c
	Q.39 _____ models consider a given system through several temporal stages, corresponding to a sequence of decisions. A.iconic B.analogical c.deterministic d.Dynamic Answer: d
	Q.40 The term _____theory is reserved for the variety of mathematical models and methods that can be found at the core of each datamining analysis and that are used to generate new knowledge. A. mathematical learning B. machine learning C. deep learning D. static learning Answer: a
	Q.41 Continuous attributes are numerical attributes that assume an uncountable _____of values. A.infinity B.zero C.first D. non-zero Answer: a
	Q.42 Training of the models is carried out using a sample of records extracted from the_____.

	a. Duplicate dataset b. original dataset c. Training dataset d. Result dataset Answer:b
	Q.43 An _____in the company information systems, expected to supervise the access to the information sources. a.expert B.trainer C.developer D.tester Answer: a
	Q.44 Trough which the effort of representation is justified by the remarkable conciseness of the information achieved . A.through a well-designed documentation B.through a class diagram C.through a well-designed chart D. through a well-designed journal Answer: c
	Q.45 Data may contain erroneous or anomalous values, which are usually referred to as_____. A.Noise B.outliers C.Inconsistencies d.Reduction Answer:b
	Q.46 estimate procedures can become rather complex and time-consuming for a large dataset with a high percentage of_____. A Training data B. missing data C.result data D. expert data Answer: b
	Q.47 whichtechnique uses this equation:

	$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$ A.Min-max standardization B.z- score C.standardisation d.zero scaling Answer: B
	Q.48 _____methods select the relevant attributes before moving on to the subsequent learning phase, and are therefore independent of the specific algorithm being used. A.Filter B.embedded C.wrapper D. Scaling Answer: A
	Q.49 how many distinct myopic search schemes are followed? A.two B.four C.three D.one Answer: c
	Q.50 Data mining projects differ in many respects from both classical statistics and _____analyses. A.OLAP B.OLTP C. OLAM D.HOLAP Answer: a
	Q.51 Which of the following is a dimension reduction technique? a) Sampling b) Box plot c) Principal component analysis d) Z – Test Answer: C

	Q.52 Which of the following is not an objective of PCA? a) To reduce attribute space from a larger number of variables to a smaller number of variables b) To identify new meaningful underlying variables c) To convert a set of observations of possibly uncorrelated variables into a set of values of linearly correlated variables d) To reduce number of dimensions Answer: C
	Q.53 The first principal component in PCA is obtained by means of _____ associated with the maximum eigenvalues. a) Eigenvalues b) Eigen vectors c) Eigenvalues and Eigenvectors d) Vector addition Answer: B
	Q.54 The full form for PCA is a) Principal Component Analysis b) Principle Component Analysis c) Partition Component Analysis d) Partial Component Analysis Answer: A
	Q.55 Why is dimensionality reduction useful? a) To achieve parsimony b) To avoid overfitting c) It reduces processing time d) To achieve parsimony, To avoid overfitting, It reduces processing time Answer: D
UNIT-03	
Unit-03	1. Classifications models are used to generate _____that allow the target class of future examples to be predicted. A. a set of rules B. a set of results C. a set of predicted variables D. a set of misclassified variables Answer: A

	2. The target attribute is also called as _____ A. Class or Label B. Goal C. Leaf D. Variable Answer: A
	3. If the instances belong to two classes only, the classification is called as _____ A. Binary Classification B. Multiclass Classification C. Multicategory Classification D. Unary Classification Answer: A
	4.If the instances belongs to more than two classes than the classification is called as _____ A. Binary Classification B. Unary Classification C. Binary or Unary Classification D. Multicategory or Multiclass Classification Answer: D
	5.A prediction is obtained by applying the _____ during the training phase to the explanatory variables that describe the new instance. A. Rules generated B. Variables generated C. Values generated D. Labels generated Answer: A
	6.In the development of classification model Test phase comes after _____ A. Training Phase B. Prediction Phase C. Development phase D. Identification Phase Answer: A
	7.Which model is not the category of the classification models? A. Heuristic Models B. Separation Models C. Probabilistic Models D. Prediction Models Answer: C

	8. In $f(x) = sgn(g(x))$, where sgn(.) is a function indicating the sign of its argument, it takes the values of the two classes _____ A. {-1, 1} B. {0, 1} C. {1, 0} D. {-1, -2} Answer: A
	9. Among the following which method guarantees that each observation of the dataset appears the same number of times in the training set and exactly once in the test set. A. Cross Validation B. Holdout method C. Repeated Random Sampling D. Holdout method and Repeated Random Sampling Answer: A
	10. Which of the following is not true about Classification Matrix? A. Most accuracy measures are derived from the classification matrix B. Classification matrix is also called as confusion matrix C. It summarizes only incorrect classifications D. Rows and columns of the classification matrix correspond to the true and predicted classes Answer: C
	11. Which of the following is correct formula for accuracy? A. Accuracy= (p + v)/m B. Accuracy= p/(p + q) C. Accuracy= q/(p + q) D. Accuracy= (p + u)/m Answer: A
	12. True negative rate is calculated by the formula: A. tn = p/ (p + q) B. tn = q/ (p + q) C. tn = u/ (u + v) D. tn = p/ (p + p) Answer: A
	13. The precision is the proportion of _____ positive examples. A. Correctly classified B. Misclassified C. Actually classified D. Occasionally classified Answer: A

	14. The F-Measure is equal to zero if all the predictions are _____ A. Incorrect B. Correct C. Partially correct D. Partially incorrect Answer: A
	15. In the formula for F-Measure, $F = \frac{(\beta^2-1)\text{tp} \times \text{prc}}{\beta^2\ \text{prc} + \text{tp}}$ the value of β belongs to _____ A. [0, ∞) B. [0,1] C. [-1, 1] D. [0, 1) Answer: A
	16. The correct formula for the Geometric mean is _____ A. $gm = \sqrt{tp \times prc}$ B. $gm = \sqrt{tp \times tn}$ C. $gm = \sqrt{t \times tp}$ D. $gm = \sqrt{tp \times tp}$ Answer: A
	17. ROC in performance metrics stands for? A. Reverse operating characteristic B. Receiver operating characteristic C. Remote operating characteristic D. Revise operating characteristic Answer: B
	18. For two class classifiers, if the probability of a record being a class 1 member is greater than 0.5, then that record is classified as A. 1 B. 0 C. 1 and 0 D. -1 Answer: A
	19. Typically, classification matrix considers- A. Actual Class B. Predicted Class C. Actual Class and Predicted Class D. Target class Answer: C
	20. For a record i, prediction error can be calculated as- A. predicted value- actual value B. actual value - predicted value C. actual value+ predicted value

	D. predicted value × actual value Answer: B
	21. Which of the following is a graphical way to assess predictive performance? A. Pie Chart B. Box plot C. Histogram D. Lift curve 1. Answer: D
	22. ROC chart is a _____ plot. A. Two-dimensional B. One-dimensional C. Multi-dimensional D. Three-dimensional 2. Answer: A
	23. In ROC chart the proportion of false positive fp is on _____ and the proportion of true positive tp is on _____ A. The horizontal axis, the vertical axis B. The vertical axis, the horizontal axis C. The horizontal axis, the x-axis D. The vertical axis, the y-axis 3. Answer: A
	24. All of the following steps are part of Naïve Bayes method except- A. Find all the other records where the predictor values are same B. Determine what classes they all belong to and which is more prevalent C. Assign that class to the old record D. Express the probability as $p(x\|y) = \prod_{j=1}^{n} p(x_j \|y)$ 4. Answer: C
-	25. Which of the following is a basis of Naïve Bayes method? A. Pie Chart B. Regression C. Conditional Probability D. Pivot Table Answer: C
	26. Which of the following is necessary to establish in Naïve Bayes method? A. Beta Value B. Cut-off Probability C. Alpha value D. Gamma Value Answer: B

	27. Which of the following is an advantage of Naïve Bayes classifier? A. Simplicity B. Computational efficiency C. Good classification performance D. Simplicity, Computational efficiency, Good classification performance Answer: D
	28. Which of the following is not true for Bayes model for classification? A. Predictors should also be categorical B. Numerical variables need not to be converted into categorical C. Naïve Bayes classifiers are highly scalable D. All the records are used instead of relying on just the matching records Answer: B
	29. Naïve Bayes formula works well for- A. Prediction B. Classification C. Association D. Clustering Answer: B
	30. Naïve Bayes formula works well for- A. Prediction B. Classification C. Association D. Clustering Answer: B
	31. In Logistic regression technique input features can be A. Quantitative B. Qualitative C. Quantitative and Qualitative D. Only numeric Answer: C
	32. Function which is used to bound the probability of x between 0 and 1? A. Log function B. Sigmoid function C. Sine D. Cosine Answer: B

	33. The confusion matrix for a binary classifier gives A. True Positives, true negatives B. False Positives, false negatives C. True Positives, true negatives, false Positives, false negatives D. True negatives Answer: C
	34. In confusion matrix, which one of the following is called as power of the test? A. True negative B. True positive C. False positive D. False negative Answer: A
	35. Logistic regression is applicable for: A. Classification B. Profiling C. Clustering D. Classification, profiling Answer: D
	36. Logistic regression is a _____ regression technique that is used to model data having a _____outcome A. Linear, numeric B. Linear, binary C. Nonlinear, numeric D. Nonlinear, binary Answer: D
	37. Which of the following methods do we use to best fit the data in Logistic Regression? A. Least Square Error B. Maximum Likelihood C. Jaccard distance D. Euclidean distance Answer: B
	38. In logistic regression the logit is: A. The natural logarithm of the odds B. An instruction to record the data C. A logarithm of a digit D. The cube root of the sample size Answer: A
	39. Maximum likelihood estimation methods work for: A. Logit B. Probit C. Logit, Probit D. Least Square Error

	Answer: C 40. If x1, x2 are independent variables and y the dependent variable, which of the following represents a linear regression model? A. $y = a_0 + a_1/x_1 + a_2/x_2$ B. $y = a_0 + a_1x_1 + a_2x_2$ C. $y = a_0 + a_1x_1 + a_2x_2^2$ D. $y = a_0 + a_1x_1^2 + a_2x_2$ Answer: B
	41. Layers between the input and output layers are known as: A. Multilayer B. Resultant layer C. Hidden layer D. Output layer Answer: C
	42. In feed- forward networks, the connections between layers are _____ from input to output. A. Bidirectional B. Unidirectional C. Multidirectional D. Directional Answer: A
	43. Which of the following statement is NOT true about clustering? A. It is a supervised learning technique B. It is an unsupervised learning technique C. It uses clusters for data analysis. D. It groups the data Answer: A
	44. Which of the following is true about cluster analysis? A. Clustering is referred to as an unsupervised learning method. B. It can't uncover previously undetected relationships in a complex dataset. C. Cluster analysis is the process of ungrouping objects into subsets that have meaning in the context of a particular problem D. Clustering is referred to as a supervised learning method. Answer: D
	45. _____ is a clustering procedure characterized by the development of a tree-like structure. A. Non-hierarchical clustering B. Hierarchical clustering C. K-Means clustering D. K-Medoids clustering Answer: B

	46. HAC stands for- A. Hierarchical aggregative clustering B. Hierarchical agglomerative clustering C. Heightened agglomerative clustering D. Hierarchical absolute clustering Answer: B
	47. Which statement is not true about cluster analysis? A. Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters. B. Cluster analysis is also called classification analysis or numerical taxonomy. C. Groups or clusters are suggested by the data, not defined a priori. D. Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature. Answer: D
	48. A _____ is a tree diagram for displaying clustering results. Vertical lines represent clusters that are joined together. A. Dendrogram B. Scatter plot C. Scree plot D. Histogram Answer: A
	49. Which of the following will be Euclidean Distance between the two data point A(1, 3) and B(2, 3)? A. 1 B. 2 C. 4 D. 8 Answer: A
	50. Which of the following is an advantage of choosing k>1? A. Maximizes misclassification rate B. Provides smoothing that reduces the risk of over fitting C. Minimizes classification rate D. Doesn't maximize classification rate Answer: B
	51. The aim of clustering models is to subdivide the records of a dataset into_____ A. Homogeneous groups B. Heterogeneous groups C. Learning groups. D. Problem-Solving groups Answer: A

	52. From the following which method is not the clustering method? A. Partition B. Hierarchical C. Density based D. Divide-and-conquer based Answer: D
	53. Which clustering method develops a subdivision of the given dataset into a predetermined number K of non empty subset? A. Partition method B. Hierarchical method C. Density-based method D. Grid method Answer: A
	54. _____ Method derives clusters from the number of observations locally falling in a neighborhood of each observation. A. Partition method B. Hierarchical method C. Density-based method D. Grid method Answer: C
	55. Grid methods first derive a _____ of the space of the observations, obtaining a grid structure consisting of cells. A. Discretization B. Digitization C. Characterization D. Standardization Answer: A
	56. clustering problems actually belong to the class of_____ A. NP-hard B. NP-complete C. Complicated problem D. Complex problem Answer: A
	57. Euclidean distance between the vector associated with the pair of observations x_i and x_k is calculated by the formula_____. A. dist(x_i, x_k)= $\sqrt{\sum_{j=1}^{n}(x_{ij} - x_{kj})^2}$ B. dist(x_i, x_k)= $\sqrt{\sum_{j=1}^{n}(x_{ij} + x_{kj})^2}$ C. dist(x_i, x_k)= $\sqrt{\sum_{j=1}^{n}(x_{ij} \times x_{kj})^2}$

	D. dist(x_i, x_k)= $\sqrt{\sum_{j=1}^{n}(x_{ij} - x_{kj})^3}$ Answer: A
	58. Formula for Manhattan distance using two dimensional example is: A. dist(x_i, x_k)= $\sum_{j=1}^{n} \mid x_{ij} - x_{kj} \mid$ B. dist(x_i, x_k)= $\sum_{j=1}^{n} \mid x_{ij} + x_{kj} \mid$ C. dist(x_i, x_k)= $\sum_{j=1}^{n} \mid x_{ij} - x_{kj} \mid^2$ D. dist(x_i, x_k)= $\sum_{j=1}^{n} \mid x_{ij} - x_{kj} \mid^{-2}$ Answer: A
	59. The formula dist(x_i, x_k)= $\sqrt[q]{\sum_{j=1}^{n} \mid x_{ij} - x_{kj} \mid^q}$ shows: A. Euclidean distance B. Manhattan distance C. Minkowski distance D. Cosine distance Answer: C
	60. _____ distance depends on the angle formed by the vectors associated with the observation. A. Minkowski B. Manhattan C. Euclidean D. Arccosine Answer: D
	61. Partition algorithms usually stop when A. During the same iteration no reallocation occurs, subdivision appears stable with respect to the evaluation criterion chosen. B. Subdivisions appear unstable C. Reallocation occurs and subdivision appears unstable D. All the observation are assigned to the cluster Answer: A
	62. The subdivision of the hierarchical methods are A. Agglomerative and divisive B. Partition and divisive C. Partition and agglomerative D. Distance based and density based Answer: A

UNIT-04	
Unit-04	1. Which of the following is not a component of Relational Marketing a. Organization b. BI and Data Mining c. Technology d. Fund Answer: d
	2. Which of the following is true: a. Intensity of Relation is low for B2C b. Intensity of Relation is low for B2B c. Intensity of Relation is high for B2C d. Does not depend on whether it is B2B or B2C. Answer: a
	3. Which of the following is not included in Cycle of Relational Marketing Analysis: a. Collecting information on Customers b. Identifying their needs c. Paying them d. Planning based on knowledge Answer: c
	4. Which of the following is not a stage in "Lifetime of a Customer" a. Acquisition b. Cross/Up Selling c. Retention d. Bargaining Answer: d
	5. Which of the following is not part of Web Mining: a. Content Mining b. Structure Mining c. Database Mining d. Usage Mining Answer: c
	6. Which of the following is False for Supply Chain a. It is network of connected and interdependent organizational units b. Strong Coordination is required c. It improves flow of materials if it is effective d. Suppliers are given priority

	Answer: d
	7. Which of the following is not an optimization model? a. Extra Capacity b. Maximum Fixed Cost c. Backlogging d. Multiple Plants Answer: b 8. It is a managerial Policy whose purpose is to maximize profit through an optimal balance between demand and supply. a. Management insight b. Economic System c. Sales System d. Revenue Management System Answer: d
	9. CCR Model stands for: a. Charnes-Cooper-Rhodes b. Charley-Common-Rules c. Challenging-Common-Rules d. Cooper-Common-Rules Answer: a
	10. The _____ expresses relationship between the Inputs utilized and Outputs Produced: a. Efficiency Function b. Effective Frontier c. Efficient Frontier d. Effective Fact Answer: c
	11. The Relationship Marketing is all about a. Creating database value b. Travelling programs c. Maintaining relationship with customer d. Loyalty based on behaviour Ans: C

	12. The goal of an organization should be to a. Maintaining existing customers only b. Get more customers and also get more return customers c. Get more potential customers d. Just selling the product Ans : B
	13.Which of the following is not a component of relational marketing strategy? a. Strategy b. Data Mining c. Technology d. Customers Answer: D
	14.Which of the following is not a decision-making option for a relational marketing strategy? a. Prices b. Distribution Channels c. Product d. Sales Processes Ans: C
	15. Which of the following is not a decision-making option for a relational marketing strategy? A. Prices B. Distribution Channels C. Product D. Sales Processes Ans: C
	16. Which of the following statements is true? a. A relationship marketing is a collection of software applications. b. A relationship marketing is a coherent project where the various company departments are called upon to cooperate and integrate the managerial culture and human resources c. A relationship marketing is a coherent project where the various company departments are called upon to work using CRM tools d. A relational marketing creates a true data culture in an organization. Answer : B

	17. Which of the following is not an actor of an enterprise relationship system? a. Employees b. Customers c. Suppliers d. Competitors Answer: D
	18. Which of the following are the customers of high value? a. B2B b. B2C c. B2C2B d. B2B2C Answer: A
	19.The implementation of _____ actions gives rise to low revenue per unit transactions. a. Low-cost b. High-cost c. Min-cost d. Max-cost Answer: A
	20.The companies well acquainted with fast delivery at _____ must involve in the direction of increased customization, by introducing more options and variants of products and services offered. a. High-cost b. Low-cost c. Zero-cost d. Min-cost Answer: B
	21. _____ analyses for relational marketing purposes are a powerful tool for identifying the segments to be targeted with customized products. a. Data b. Data-mining c. Sales d. Market Answer: B

	22.Which of the following databases is not used to store the data into the data mart for relational marketing analysis? a. Salesforce database b. Data Warehouse c. OLTP database d. Marketing database Answer: C
	23. Which of the following are the indirect methods that are used to acquire customers data? a. Telephonic Conversations b. Display Advertisement Boards c. Sending email to customers d. Talk with Sales Agents Answer: B
	24. Revenue management is a managerial policy whose purpose is to maximize profits through an optimal balance between _____. A. Price and product B. Demand and supply C. Raw material D. Employee and organization Answer: B
	25. The term _____ refers to the possibility that a portion of the demand due in a given period may be satisfied in a subsequent period, incurring an additionalpenalty cost. A. Sales B. Sales process C. Backlog D. Execution Answer: C
	26._____ is concerned with optimising financial results and is especially popular in industries like hospitality, which must contend with high fixed costs and a perishable inventory. a. Revenue management b.Revenge management c.Revenue manager d.Rename management answer: A

	27 .CRM stands for_____. a. Customer relationship management b. Customer relationship manager c. Customer relay management d. Custom relationship management Answer: A
	28. _____is the amount of divisions ordered from a task. a. Efficiency b. Density c. frequency d. threshold Answer: a
UNIT-05	
	Q1. Which among these are the key stages for managing organizational transformation a) Break with the past b) Manage the future c) Both d) Accept current Ans: a
	Q2. _____ represent how you increase the ability of individuals within the organization to influence others with the knowledge. a) People b) Processes c) Technology d) Culture Ans: a
	Q3. It addresses how you choose, configure and utilise tools and automation to enable knowledge management. a) People b) Processes c) Technology d) Culture Ans: c
	Q4. It directs how you transform organizational structures to facilitate and encourage cross discipline awareness and expertise. a) Binding b) Context c) Structure d) Association Ans: c

	Q5. Characteristics of expert systems a)High Performance b)Demonstrating c)Advising d)Diagnosing Ans: a
	Q6. Capabilities of expert systems. a) Reliable b) Demonstrating c) High Responsive d) Understandable Ans: b
	Q7. In-capabilities of expert systems. a) Advising b) Diagnosing c) Interpreting input d) Refining their own knowledge Ans: d
	Q8. Strategy followed for finding cause or reasons. a) Backward Chaining b) Forward Chaining c) Facts d) Decisions Ans: a
	Q9. Strategy followed for working on conclusion, results or effects. e) Backward Chaining f) Forward Chaining g) Facts h) Decisions Ans: b
	Q.10 Levels in ES technology a) Shells b) Design c) Both d) None Ans:a
	Q.11 Knowledge Management Activity aims at **a)** Total turning test **b)** The rational agent approach **c)** To build knowledge infrastructure **d)** Thinking humanly Answer: c

	Q.12 Knowledge is _____ resource. A. Replaceable B. Draining C. Exhaustible D. Depleting Correct answer A
	Q.13 Who is the person who is responsible for making knowledge management effort in an organization . A. CIO B. CEO C. CKO D. CKE Correct answer C
	Q.14 Human use content memory and thinking whereas, robots are using the _____, designed by scientists. A. Knowledge B. Data C. built-in instruction D. High Level Information Correct answer C
	Q.15 Which of the following is not a component of ES . A. Knowledge Base B. Inference Engine C. User Interface D. High-level Data Correct answer D
	Q.16 _____ is required to exhibit intelligence. A. Data B. Knowledge C. Information D. High-quality data Correct answer B

	Q.17 Forward Chaining Strategy is used by _____ to recommend a solution. A. Inference Engine B. Knowledge Engine C. Expert Engine D. Forward Engine Correct answer A
	Q.18 In which of the following steps of expert system development, the knowledge should be represented in IF-THEN-ELSE rules form? A. System Design B. Expert System Development and Completion C. Prototype Development D. Problem Domain Identification Correct answer C
	Q.19 Who introduced the term “Artificial Intelligence” A. Arthur Samule B. Marvin Lee Minsky C. Jhon McCarthy D. E. F. Codd Correct answer C
	Q.20 KDD Stands for : A. Knowledge Discovery Data B. Knowledge Discovery in Database C. Knowledge Database Discovery D. Knowledge Data Discovery Answer D
	Q.21 which one of the following is not a type of Knowledge A. Declarative Knowledge B. Procedural Knowledge C. Tactic Knowledge D. Collective Knowledge Correct answer D
	Q.22 What kind of signal is used in speech recognition? a)Electromagnetic signal b) Electric signal c) Acoustic signal

	d) Radar Answer: c
	Q.23 A _____ is nothing but an expert system without knowledge base. a)Tools b)Expert System c) shell d) knowledge Answer: c
	Q.24 Data, information, and past experience combined together are termed as _____. a) Inference b) Acquisition c) vision d) knowledge Answer: d
	Q.25 The advantage of AI over Natural Intelligence are a) Fabulous speed b) Less baised c) Error prone d) a & b Answer: d
	Q.26 Which is the key area in which Knowledge Management is applied a) Technological Advances b) Inference Engine c) Globalization of Business d) a & c Answer: d
	Q.27 Which is the key area in which Knowledge Management is applied a) Technological Advances b) Inference Engine c) Globalization of Business d) a & c Answer: d

	Q.28 Which of the following is not a benefits of Expert Systems? a) Availability b) Speed c) Time d) Less Error Rate Answer: c
	Q.29 What is the form of Knowledge representation? a) IF-THEN b) IF-THEN-ELSE c) IF-ELSE d) ELSE Answer: b
	Q.30 Which of the following is not a Capabilities of Expert Systems? a) Advising b) Demonstrating c) Explaining d) Expanding Answer : d
	Q.31 The challenges faced by Knowledge Management System are_____ a) Psychology b) Communication and Collaboration c) Control theory and cybernetics d) Computer Engineering Answer: b
	Q.32 Knowledge Management Activity aims at a) Total turning test b) The rational agent approach c) To build knowledge infrastructure d) Thinking humanly Answer: c
	Q.33 Lengthy information is presented in _____ or graphical format and stored in appropriate form in summarizing step of knowledge management process. (a) Unorganised (b) Tabular

	(c) Sequential (d) unstructured Answer: b
	Q.34 In decision making step of Knowledge Management process the _____ is used for decision making. (a)Data (b) Knowledge (c) Information (d) Metadata Answer: B
	Q.35 People are ultimately the _____ of knowledge. (a) Storage (b) Gainer (c) Holders (d) Acquire Answer: C
	Q.36 Learning organisation is an organisation characterised by a deep commitment to learning and _____ with the intention of continuous improvement. (a) Education (b) Understanding (c) Training (d) Improving Answer: A
	Q.37 _____ is a way of making a computer, a computer-controlled robot, or a software think intelligently, in the similar manner the intelligent humans think. (a) Machine Intelligence (b) Artificial Intelligence (c) Linguistic Intelligence (d) Naturalistic Intelligence Answer: B
	Q.38 Data, information and past experience combined together are termed as _____. (a) Intelligence (b) Knowledge (c) Expert systems (d) Data driven systems Answer: b

	Q.39 First basic step for knowledge management process is_____ A. summarizing information B. Data collection C. Data analysis D. Data classification Answer:B
	Q.40 Expert systems are capable of _____. A. assisting human in decision making B. refining their own knowledge C. possessing human capabilities D. substituting human Answer: A
	Q.41 Knowledge Management process has basically six steps, they are collecting, organizing, summarizing, analysing, synthesizing and _____ of data. (a) Decision making (b) Knowledge processing (c) Planning (d) Arranging Answer: A
	Q.42 The three key stages for managing organisational transformation for critical success factor are break with the past, managing the past and _____. (a) Invest in present (b) Invest in the future (c) Managing the present (d) Outlining the future Answer: B
	Q.43 _____ are one of the prominent research domains of Artificial Intelligence. (a) Knowledge Management Systems (b) Data driven Systems (c) Expert Systems (d) Cognitive Systems Answer:B

	Q.44 _____ is a strategy of an expert system to answer the question, "What can happen next?" (a) Forward Chaining (b) Backward Chaining (c) Total Task Chaining (d) Backward Reasoning Answer: A
	Q.45 Following step is not involved in knowledge management process. A. Decision making B. Information Analyzing C. Application Development D. Synthesizing Answer: C
	Q.46 In data collection procedure of knowledge management process, certain data collection points define the _____ of certain routine reports. A. Idea B. Knowledge C. Data D. Summary Answer: D
	Q.47 knowledge management process has _____ basic steps. A. 4 B. 6 C. 7 D. 8 Answer: B
	Q.48 The components of Expert system includes_____ A. knowledge expert,inferenceengine,user interface B. knowledge base,inferenceengine,end interface C. knowledge base,interfaceengine,user interface D. knowledge base,inferenceengine,user interface Answer: D
	Q.49 Forward and Backward channing are the strategies of _____ A. Inference Engine B. Knowledge management process C. Data mining process D. Effective communication process Answer: A
	Q.50 Knowledge engineers acquire information from subject expert by_____. A. Recording,observing his/her work,Interviewing B. Recording,observing his/her communication,Interviewing C. Recording,observing his/her work,observation

	D. Reading,observing his/her work,Interviewing Answer: A