Mid-Term Exam

Started: 4 Mar at 14:02

Quiz instructions

The exam on **modules 1, 2, 3, 4, 5, and 6.**

The exam will be available on **Monday March 04, 2024 from 2:00 PM to 6:00 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.**

First Part: MCQs

Question 1 1 pts

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

- ○ Clustering
- ○ Classification
- ○ Association
- ◉ Visualization

Question 2 1 pts

Information is transformed into _____to make decisions.

- ○ Data
- ○ Information
- ○ File
- ◉ Knowledge

Question 3 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 pts

Which is not the Phase of data Mining Process

○
Data preprocessing

○
Data Discarding

○
Feature Selection

○
Prediction and interpretation

Question 5 1 pts

_____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.

○
Business information warehouse

○
Best practice

○
Business intelligence

○
Data mart

Question 6 1 pts

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

○
Web mining

○
Invisible data mining

○
All the three

○
Scalable data mining methods

Question 7 1 pts

Which is the Application of Data Mining?

○
Both

○
Fraud Detection

○
None

○
Risk Analysis

Question 8 1 pts

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

○
Nominal

○
Binary

○
Numerical

○
Ordinal

Question 9 1 pts

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

○
Ratio

○
Numerical

○
Nominal

○
Ordinal

Question 10 1 pts

_____represent the real problem situations.

○
Tools

○
Information

○
Data

○
Models

Question 11 1 pts

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

○
Reduction

○
Outliers

○
Noise

○
Inconsistencies

Question 12 1 pts

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

○
Training data

○
testing data

○
missing data

○
resulting data

Question 13 1 pts

The purpose of feature selection, also called _____.

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

Question 14 1 pts

Data by itself is not useful unless

- It is massive
- It is properly stated
- It is processed to obtain information
- It is collected from diverse sources

Question 15 1 pts

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

- Training dataset
- Result dataset
- Duplicate dataset
- Original dataset

Question 16 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

○
Accuracy = N/(TP + TN)

Question 17 1 pts

On which learning methods the Data Mining method is based?

○
basic learning methods

○
deductive learning methods

○
inductive learning methods

○
comprehensive learning methods

Question 18 1 pts

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

○
Partially incorrect

○
Correct

○
Partially correct

○
Incorrect

Question 19 1 pts

In data mining, what is the purpose of Interpretation?

○
to express the rules and criteria for easy understanding

○
to identify irregular patterns in the data

○
to determine useful patterns in the data

○
All the three statements

Question 20 1 pts

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

○
[0,1]

○
[0, ∞)

○
[-1, 1]

○
[0, 1)

Question 21 1 pts

The confusion matrix for a binary classifier gives

○
True Positives, true negatives

○
True Positives, true negatives, false Positives, false negatives

○
True negatives

○
False Positives, false negatives

Question 22 1 pts

The precision is the proportion of _____ positive examples.

○
Correctly classified

○
Occasionally classified

○
Actually classified

○
Misclassified

Question 23 1 pts

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

○
All the three

○
Good classification performance

○
Computational efficiency

○
Simplicity

Question 24 1 pts

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

○
a set of predicted variables

○
a set of rules

○
a set of targeted results

○
a set of misclassified variables

Question 25 1 pts

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

○
Since they can provide complex forms of relationships.

○
Because they usually required large datasets to be initiated.

○
Due to its drawback of handling the presence of correlated attributes.

○
All the three

Question 26 1 pts

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

○
Not mutually exclusive rules

○
Not exhaustive rules

○
Exhaustive rules

○
Mutually exclusive rules

Question 27 1 pts

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

○
All the records are used instead of relying on just the matching records

○
Numerical variables need not to be converted into categorical

○
Predictors should also be categorical

○
Naïve Bayes classifiers are highly scalable

Question 28 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

○
deep learning or machine learning

○
machine learning or knowledge discovery in databases

○
machine learning or knowledge recovery in databases.

Question 29 1 pts

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

○
Pivot Table

○
Pie Chart

○
Conditional Probability

○
Regression

Question 30 1 pts

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

○ 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.

○ Determine what classes they all belong to and which is more prevalent

Question 31 1 pts

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

○ Sigmoid function

○ Sine

○ Log function

○ Cosine

Question 32 1 pts

K- Nearest Neighbor Classifier is know as:

○ Lazy learner

○ All the three

○ Instance-based learner

○ Local classifier

Question 33 1 pts

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

○ Linear, numeric

Linear, binary

Nonlinear, binary

Nonlinear, numeric

Question 34 1 pts

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

Misclassification rate will be very high.

All the three

The classifier be capable of overfitting

The neighborhood may include points from other classes

Question 35 1 pts

Logistic regression is applicable for:

Clustering

Association

Prediction

Classification

Question 36 1 pts

In logistic regression the logit is:

The cube root of the sample size

A logarithm of a digit

The natural logarithm of the odds

○
An instruction to record the data

Question 37 1 pts

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

○
All the three

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

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

○
Proximity computations normally require the presence of all attributes

Question 38 1 pts

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

○
All the three

○
Due its characteristics of producing decision boundaries of arbitrary shapes

○
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

Second Part: Short Questions

Question 39 3 pts

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

Edit View Insert Format Tools Table

12pt Paragraph B I U A T²

p 0 words

Question 40 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)?

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 41 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.

Edit View Insert Format Tools Table

12pt Paragraph

p 0 words

Question 42 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)

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

Saved at 14:04 Submit quiz