

DOMANDE SELEZIONE PUBBLICA PER L'ASSUNZIONE A TEMPO INDETERMINATO DI 100 UNITÀ PER LA TERZA AREA FUNZIONALE, FASCIA RETRIBUTIVA F1, PROFILO PROFESSIONALE FUNZIONARIO INFORMATICO - DS50 - PROVA SCRITTA - BUSTA 1

1) Nel linguaggio "R", una lista (list) può contenere elementi ...

- A di tipo e classe anche diversi per ogni elemento.
 - B esclusivamente appartenenti alla stessa classe.
 - C in "R" non esiste il concetto di lista.
-

2) Con riferimento al linguaggio "R", qual è il nome della struttura dati utilizzata per memorizzare dati in formato tabellare, con colonne di tipo diverso?

- A Table.
 - B Data frame.
 - C Dataset.
-

3) Le funzionalità del linguaggio R sono suddivise in una serie di ...

- A Libraries.
 - B Packages.
 - C Objects.
-

4) In Python, i blocchi di codice vengono delimitati attraverso...

- A indentazione.
 - B parole chiave begin / end.
 - C parentesi graffe.
-

5) In Python, se una lista (list) chiamata "Vettore" contiene 8 elementi, quale, tra i seguenti fa riferimento al secondo elemento della lista "Vettore"?

- A Vettore[1].
 - B Vettore[2].
 - C Vettore[-3].
-

6) In quale linguaggio di programmazione la seguente espressione è sintatticamente corretta

`a=[1,2,3]`

- A Sia in R che in Python.
 - B Solo in R.
 - C Solo in Python.
-

7) Che cosa è un entity-relationship diagram (diagramma entità-relazione)?

- A È una tecnica di modellizzazione dei dati che crea una rappresentazione grafica delle entità e delle relazioni tra le entità a livello concettuale.
 - B È una tecnica di modellizzazione dei dati che crea una rappresentazione grafica delle entità e delle relazioni tra le entità a livello fisico.
 - C È una tecnica di modellizzazione dei dati che indica le relazioni tra entità e oggetti esterni.
-

8) In SQL, l'applicazione dell'operatore UNION tra due result set A e B richiede che...

- A A e B abbiano la chiave primaria in comune.
- B A e B abbiano stesso numero di colonne anche di tipo diverso.
- C A e B abbiano stessa struttura dati.

-
- 9) Considerando una tabella IVA che comprende 5.000.000 di righe (una per un valore distinto della colonna codice_fiscale) e con una colonna codice_attivita che assume 1.207 valori distinti, da cosa è costituito il result set ritornato dalla seguente espressione SQL:

```
SELECT COUNT DISTINCT codice_fiscale
FROM IVA
GROUP BY codice_attivita
```

- A 1 riga.
- B 1.207 righe.
- C 5.000.000 righe.

-
- 10) Date due tabelle A e B di un database relazionale, l'operazione di INNER JOIN tra A e B restituisce...

- A solo le colonne di A e B che hanno una chiave in comune.
- B solo le righe di A e B che hanno una chiave in comune.
- C tutte le righe di A e B.

-
- 11) Quale dei seguenti comandi deve essere usato, in SQL, per eliminare un attributo (ATT) da una tabella (RT)?

- A alter table RT drop column ATT;
- B alter table RT delete column ATT;
- C alter table ATT drop column RT;

-
- 12) In SQL, con riferimento alla tabella in figura, che effetto avrà questa espressione?

```
DELETE FROM Indirizzo
WHERE Citta = 'Genova';
```

Indirizzo				
Cognome	Nome	Via	Civico	Citta
Rossi	Paolo	Adige	33	Genova
Verdi	Antonio	Po	55	Torino
Bianchi	Luigi	Piave	44	Roma
Viola	Dario	Isonzo	66	Roma

- A Verrà eliminata la prima riga della tabella.
- B Verrà eliminata la tutta la tabella Indirizzo.
- C Verrà eliminato il valore del campo Città della prima riga della tabella.

-
- 13) Cosa si deve creare, in un database relazionale, per creare una relazione "molti a molti" tra due tabelle?

- A Una tabella.
- B Una vista.
- C Un nuovo attributo in ciascuna delle due tabelle.

14) Nel campo dell'inferenza statistica cosa si intende per ipotesi alternativa?

- A Una caratteristica della popolazione che non è associabile a nessun individuo.
- B L'ipotesi posta a test in alternativa all'ipotesi nulla.
- C L'ipotesi messa a test la cui probabilità è uguale a quella dell'ipotesi nulla.

15) Nel campo dell'inferenza statistica cosa rappresenta l'errore di seconda specie?

- A L'errore che si commette nel calcolo della probabilità di due eventi indipendenti.
- B L'errore che si commette accettando l'ipotesi alternativa quando questa è falsa.
- C L'errore che si commette accettando l'ipotesi nulla quando questa è falsa.

16) Nel campo dell'inferenza statistica quale delle seguenti NON è una proprietà degli stimatori?

- A Ammissibilità.
- B Correttezza.
- C Efficienza.

17) In statistica, quale metodo è utilizzato per valutare l'indipendenza di due variabili aleatorie?

- A Test t di Student.
- B Metodo del chi-quadro.
- C Metodo di Gauss.

18) Quale caratteristica di una distribuzione denota la seguente espressione:

$$\frac{1}{n} \sum_{i=1}^n x_i.$$

- A Media geometrica.
- B Media aritmetica.
- C Covarianza.

19) Quali dei seguenti metodi esamina i dati attuali e storici allo scopo di prevedere eventi futuri?

- A Analisi predittiva.
- B Analisi non supervisionata.
- C Analisi stocastica.

-
- 20) **I modelli di apprendimento non supervisionato...**
- A vengono utilizzati per trovare nei dati forme di organizzazione e di rappresentazione a partire da un dataset etichettato.
 - B vengono utilizzati per trovare nei dati forme di organizzazione e di rappresentazione precedentemente note.
 - C vengono utilizzati per trovare nei dati forme di organizzazione e di rappresentazione precedentemente sconosciute.
-
- 21) **I metodi di apprendimento con rinforzo...**
- A includono le tecniche di clustering.
 - B non prevedono una alimentazione preliminare con dati etichettati.
 - C prevedono una alimentazione preliminare con dati etichettati.
-
- 22) **Quale tipologia di machine learning richiede necessariamente un set di dati etichettati?**
- A Apprendimento con rinforzo.
 - B Apprendimento non supervisionato.
 - C Apprendimento supervisionato.
-
- 23) **Quale dei seguenti affermazioni rispetto all'uso dei metodi di apprendimento con rinforzo è corretta?**
- A L'agente inizialmente adotta scelte spesso errate.
 - B L'agente necessita di un ambiente su cui apprendere pre-caricato con dati storici.
 - C L'agente fin da subito adotta scelte sempre corrette.
-
- 24) **Quale tra i seguenti NON è un tipo di modello di apprendimento supervisionato?**
- A Classificazione.
 - B Ad agente e stato.
 - C Regressione.
-
- 25) **La regressione lineare nell'apprendimento supervisionato consiste...**
- A nel calcolo di una relazione che massimizza la correlazione lineare tra i predittori.
 - B nella ricerca della retta che congiunge i dati di input del dataset.
 - C nella ricerca di una relazione lineare fra i predittori e la variabile di risposta.
-
- 26) **Nel metodo di regressione lineare cosa rappresentano i residui?**
- A Lo scostamento tra il valore reale ed il valore predetto.
 - B L'insieme dei punti scartati dal set di apprendimento.
 - C I coefficienti dell'equazione lineare che non sono presi in considerazione dal modello predittivo.
-
- 27) **Perché nei metodi per verificare la bontà di un modello di regressione lineare si considerano i quadrati dei residui?**
- A Per ridurre la distorsione della stima.
 - B Per evitare che residui positivi e negativi si compensino tra loro.
 - C Per approssimare l'errore ad una parabola.
-
- 28) **Nella fase di pre-elaborazione dei dati, quale delle seguenti affermazioni è corretta in riferimento alla presenza di dati mancanti?**
- A Tutti i dati mancanti devono essere rimpiazzati con un dato stimato.
 - B Tutti i record contenenti dati mancanti devono essere necessariamente esclusi.
 - C Si possono modificare i dati inserendo al posto dei dati mancanti delle medie o dei valori stimati.
-

- 29) **Un albero di classificazione si usa quando la variabile risposta è...**
- A continua.
 - B positiva.
 - C categorica.
-
- 30) **Nell'algoritmo K-Means un centroide rappresenta...**
- A il valore medio tra tutti i dati analizzati.
 - B il valore medio tra i cluster.
 - C il punto medio o baricentro di un cluster.
-
- 31) **L'algoritmo di backpropagation...**
- A non richiede dati etichettati.
 - B richiede un'interazione con l'ambiente.
 - C viene applicato in presenza di dati etichettati.
-
- 32) **Quale delle seguenti è un esempio di stopword?**
- A Un qualsiasi verbo.
 - B L'articolo "il".
 - C Un nome di città.
-
- 33) **Per quali tra queste applicazioni non risulta appropriato l'utilizzo di tecniche di Social Network Analysis?**
- A Analizzare la serie storica delle vendite in relazione ad un insieme di clienti.
 - B Realizzare un modello di propagazione di un'epidemia.
 - C Identificare gruppi di soggetti che effettuano transazioni commerciali tra di loro.
-
- 34) **Nell'ambito del text mining quale delle seguenti operazioni è frequentemente eseguita nella fase di pre-processing dei testi?**
- A Trasformazione del testo in minuscolo.
 - B Imputazione dei valori mancanti.
 - C La lettura automatizzata del testo da un'immagine.
-
- 35) **Come si chiama l'applicazione del text mining finalizzata all'analisi delle opinioni sui social network?**
- A Social analysis.
 - B Posts analysis.
 - C Sentiment analysis.
-
- 36) **Come si misura l'importanza di un nodo in un grafo?**
- A Attraverso misure di densità.
 - B Attraverso misure di dimensione.
 - C Attraverso misure di centralità.
-
- 37) **Quale delle seguenti affermazioni sulla qualità dei dati è falsa?**
- A Accuratezza, completezza e tempestività sono dimensioni di qualità.
 - B Gli aspetti relativi alla qualità non sono mai misurabili in termini quantitativi.
 - C Le dimensioni di qualità sono definite in appositi standard.
-
- 38) **Nell'ambito della qualità dei dati, la dimensione di consistenza di un dataset si riferisce...**
- A all'aderenza rispetto a un dizionario di valori predefiniti.
 - B alla coerenza rispetto ad un insieme di regole semantiche del dominio di riferimento.
 - C alla percentuale di dati valorizzati rispetto alla popolazione di riferimento.

39) Quali delle seguenti caratteristiche non attiene alla qualità dei dati?

- A Completezza.
- B Consistenza.
- C Distorsione.

40) Un dataset in cui il 50% dei valori è mancante, è caratterizzato da un basso valore di...

- A consistenza.
- B fruibilità.
- C completezza.