

ANALIZA DANYCH

próbny test kompetencyjny

IMIĘ:																		
NAZWISKO:																		
Nr indeksu:																		
Moduły:	Podstawy matematyki (zad. 1-5)	Statystyczna analiza danych (zad. 6-9)	Podstawy informatyki (zad. 10-13)	Analiza danych biznesowych (zad. 14-17)	Metody eksplo-racji danych (zad. 18-21)	Bazy danych (zad. 22-25)												
Wyniki:																		

Poniższy test wielokrotnego wyboru należy wypełnić długopisem, wpisując we wszystkie rubryki odpowiednio: **T** (tak) lub **N** (nie). Błędne zaznaczenie należy otoczyć kółkiem i wpisać po lewej stronie poprawioną odpowiedź. W ramach każdego zadania, za zaznaczenie wszystkich trzech poprawnych odpowiedzi uzyskuje się 1 punkt, za zaznaczenie dwóch poprawnych odpowiedzi - 0,5 punktu, za zaznaczenie jednej dobrej odpowiedzi - 0,25 punktu, za brak dobrych odpowiedzi - 0 punktów. Brak odpowiedzi jest traktowany jak odpowiedź błędna. Za test można uzyskać maksymalnie 25 punkty. Powodzenia.

1. Zdaniem równoważnym do: *Jeśli X jest ojcem Y i Y jest ojcem Z, to Z nie jest synem X.* jest:

Jeśli X nie jest ojcem Z, to X nie jest ojcem Y lub Y nie jest ojcem Z.

Jeśli Z jest synem X, to X nie jest ojcem Y i Y nie jest ojcem Z.

Jeśli X jest ojcem Z, to X nie jest ojcem Y lub Y nie jest ojcem Z.

2. Dana jest funkcja $f(x) = e^{x^2-4x}$. Wówczas:

$f'(2) = 0$, więc jest spełniony warunek konieczny istnienia ekstremum w punkcie $x = 2$.

$f'(2) = \lim_{h \rightarrow 0} \frac{e^{(2+h)^2-4(2+h)} - e^{-4}}{h}$.

tangens kąta nachylenia stycznej do wykresu funkcji f w punkcie $x = 2$ do osi OX wynosi e^{-4} .

3. Całka $\int_{-1}^1 (x^3 - x) dx$ jest równa:

$2 \int_0^1 (x^3 - x) dx$.

-1.

polu powierzchni figury ograniczonej krzywymi $y = x^3$ i $y = x$.

4. Dany jest układ równań:

$$\begin{cases} ax - 3y + z = 3 \\ 2y + 5z = -2 \\ 4z = 0 \end{cases}$$

Gdy $a \neq 0$, to układ jest układem Cramera.

Gdy $a = 0$, to układ ma jedno rozwiązanie: $x = 0$, $y = -1$, $z = 0$.

Gdy $a = 0$, to układ jest niesprzeczny, ponieważ jest spełnione założenie tw. Kronecera-Capellego.

5. Rzucamy symetryczną monetą do momentu wyrzucenia orła. Niech X oznacza liczbę wykonanych rzutów. Wtedy
- zmienna losowa X ma rozkład Poissona.
 - prawdopodobieństwo zdarzenia, że $X = 3$ wynosi $\frac{1}{8}$.
 - zmienna losowa X ma rozkład geometryczny.
6. Wariancją liczb x_1, x_2, \dots, x_n nazywamy liczbę $\frac{1}{n} \sum_{i=1}^n (x_i - m)^2$, gdzie m jest średnią arytmetyczną liczb x_1, x_2, \dots, x_n . Dane są liczby 1, 1, 3, 5, 6, 6. Wiadomo, że ich wariancja wynosi $\frac{41}{9}$. Wtedy
- wariancja liczb x_1, x_2, \dots, x_n wynosi $(\frac{1}{n} \sum_{i=1}^n x_i^2) - m^2$.
 - średnia arytmetyczna liczb 1, 1, 3, 5, 6, 6 jest większa od ich wariancji.
 - wariancja liczb 3, 3, 9, 15, 18, 18 wynosi 41.
7. Niech X_1, X_2, \dots, X_n będzie próbą prostą. Estymatorem nieobciążonym wariancji jest statystyka
- $\frac{1}{n-1} \sum_{i=1}^n (X_i - \mathbb{E}X_i)^2$.
 - $\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}X_i)^2$.
 - $\frac{1}{n} \sum_{i=1}^n X_i$.
8. W wykresie ramka-wąsy
- pozioma linia wewnątrz skrzynki prezentuje wartość średniej.
 - długość ramki jest wyznaczona za pomocą wariancji obserwowanej zmiennej liczbowej.
 - długość ramki jest wyznaczona za pomocą średniej oraz mediany dla obserwowanej zmiennej liczbowej.
9. W modelu liniowym regresji wielokrotnej wybieramy zmienne w modelu, np. za pomocą metody eliminacji i dołączania. Metoda dołączania polega na wybieraniu zmiennej, dla której p-wartość odpowiadającego jej testu t
- jest największą p-wartością mniejszą od zadanej z góry liczby.
 - jest największą p-wartością większą od zadanej z góry liczby.
 - jest najmniejszą p-wartością mniejszą od zadanej z góry liczby.
10. Klasyczny komputer składa się z:
- procesora, magistrali i zasilania.
 - procesora, pamięci operacyjnej i urządzeń wejścia/wyjścia.
 - taśmy, głowicy i młyna.
11. Liczba (-27) zakodowana na 8 bitach w formacie
- znak moduł ma postać 10011011.
 - uzupełnień do dwóch ma postać 11100101.
 - nie może być zakodowana w żadnym z tych formatów, ponieważ jest liczbą ujemną.

12. W wyniku realizacji fragmentu kodu napisanego w Pythonie

```
for i in range(4,13,3):  
    print(i)
```

- zostaną wyświetlone cztery liczby.
- suma wyświetlonych liczb jest równa 21.
- ostatnią wyświetloną liczbą jest liczba parzysta.

13. Cechą charakterystyczną programowania obiektowego jest:

- dziedziczenie, czyli mechanizm umożliwiający dziedziczenie funkcjonalności.
- hermetyzacja, która umożliwia dostęp do danych tylko za pomocą wyodrębnionych metod.
- podział programu na podprogramy (moduły), które komunikują się za pomocą interfejsów.

14. Formuła =ILOCZYN(D\$4;D\$5;D\$6;D\$7) wprowadzona w komórce D9 da ten sam wynik co formuła:

- =ILOCZYN(\$D\$4:\$D\$7) wprowadzona w komórce E11.
- =ILOCZYN(\$D4:\$D7) wprowadzona w komórce E11.
- =D4*D5*D6*D7 wprowadzona w komórce E11.

15. O ważonej wykładniczo średniej kroczącej można stwierdzić, że

- jest bardziej zależna od ostatnich cen niż od wcześniejszych.
- reaguje szybciej na zmiany ceny niż średnia krocząca prosta o tej samej liczbie okresów.
- wyprzedza cenę.

16. Wskazania oscylatorów ROC (RATE OF CHANGE), MOM (MOMENTUM)

- są ważniejsze na końcu trendu niż na jego początku.
- informują tylko o sile, a nie o kierunku trendu.
- informują o fazie trendu.

17. W następującym wyrażeniu języka DAX: Kupno = SUM(Produkty[Zakupy]), słowo Produkty **nie** oznacza:

- tabeli, do której istnieje odwołanie.
- początku formuły.
- nazwy miary.

18. Dana jest metryka $d(X, Y) = \max_{i=1,2,\dots,n} |x_i - y_i|$. W metryce d

- kołem jednostkowym na płaszczyźnie jest kwadrat o boku długości 1.
- kołem jednostkowym na płaszczyźnie jest kwadrat o wierzchołkach (1,1); (-1,1); (-1,-1); (1,-1).
- kołem jednostkowym w przestrzeni jest sześciąt o boku długości 2.

19. Grupowanie

- jest metodą uczenia nadzorowanego.
- wymaga wyboru miary odległości pomiędzy obserwacjami.
- wymaga stosowania metod wyłącznie hierarchicznych.

20. Zmienna A przyjmuje wartości $0, 0, 0, 1, 1, 1, 1, 1$. Ile wynosi entropia podziału węzła według zmiennej A ?

- $-(\frac{5}{8} \log_2 \frac{5}{8} + \frac{3}{8} \log_2 \frac{3}{8})$.
- $-(\frac{5}{8} \log_2 \frac{3}{8} + \frac{5}{8} \log_2 \frac{3}{8})$.
- $-(0 \log_2 \frac{3}{8} + 5 \log_2 \frac{5}{8})$.

21. CRISP-DM

- jest metodą uczenia z nauczycielem w eksploracji danych.
- jest metodą uczenia bez nauczyciela w eksploracji danych.
- jest schematem ilustrującym proces eksploracji danych.

22. W każdej tabeli relacyjnej bazy danych:

- istnieje przynajmniej jeden klucz kandydujący.
- może istnieć więcej niż jeden klucz obcy.
- może istnieć klucz kandydujący, składający się ze wszystkich pól tabeli.

23. Wśród klauzul instrukcji **SELECT** wyróżniamy:

- FROM**.
- CHECK**.
- WHERE**.

24. Twierdzenie CAP

- stwierdza, iż w systemach działających w rozproszeniu (na wielu partycjach) nie możemy jednocześnie mieć spełnionego warunku spójności (consistency) i dostępności (availability).
- stwierdza, iż w systemach NoSQL mamy zawsze możliwość tworzenia kopii danych (copy), operacje są atomowe (atomic), a wszystkie dane przenośne (portability).
- dotyczy związku pomiędzy kopiowaniem i partycjonowaniem (CAP - CopyAvoidPartition, tworzenie kopii uniemożliwia podział na partycje).

25. Zagadnienia Big Data charakteryzują następujące czynniki (wszystkie z grupy):

- price, capacity, consistency.
- volume, velocity, variety, veracity, value.
- space, size, speed, sensibility.