

1번) 데이터웨어하우스에 대한 설명으로 가장 적절하지 않은 것?

--> 데이터웨어하우스는 전사적 차원보다는 특정 조직의 특정 업무 분야에 초점

2번) 개인정보 비식별화 기법을 설명한 것으로 부적절한 것?

--> 데이터 마스킹은 개인정보 식별이 가능한 특정 데이터값을 삭제

3번) 데이터베이스에 대한 설명으로 가장 적절하지 않은 것?

--> 데이터베이스 내의 모든 데이터는 2차원 테이블로 표현

4번) 사생활 침해 문제를 해결하기 위한 방법으로 가장 적절한 것?

--> 개인정보 사용자 책임제도 도입

5번) 개인의 내재된 경험을 객관적인 데이터로 변환하여 문서나 매체에 저장·가공·분석하는 과정?

--> 표출화

6번) 빅데이터가 가치 창출 측면에서 기업, 정부, 개인에게 미친 영향?

--> 개인은 맞춤형 서비스를 받거나 적시에 필요한 정보를 얻음으로써 기회비용을 절약

7번) 데이터 사이언스에 대한 설명으로 가장 부적절한 것?

--> 정형 데이터를 대상으로 총체적 접근법을 사용

8번) 데이터에 대한 설명으로 가장 적절하지 않은 것?

--> 데이터는 축적된 지식과 아이디어가 결합된 창의적 산물

단1) 사물인터넷(IoT)

단2) 블록체인

9번) 분석 대상은 명확하지만 분석 방식이 명확하지 않은 경우 수행하는 분석 주제의 유형?

--> 솔루션

10번) 분석 마스터플랜의 세부 이행계획 수립시 고려해야 할 데이터 분석체계?

--> 프로젝트의 세부 일정계획도 데이터 분석체계를 고려하여 작성한다.

11번) 분석과제의 특징 중 Accuracy와 Precision 설명으로 틀린 것?

--> 분석의 안정성-Accuracy, 활용-Precision이 중요

12번) 분석 기획에 대한 설명으로 적절하지 않은 것?

--> 상향식 분석은 분석 기획 전 탐색적 데이터 분석 수행

13번) 분석 마스터 플랜을 수립할 때 적용 범위 및 방식에 대한 고려요소가 아닌 것?

--> 투입 비용 수준

14번) 분석 정의서에 대한 설명으로 가장 적절한 것?

--> 소스 데이터, 분석 방법, 난이도 등의 분석 과정을 기록한 문서

15번) 분석 수준 진단의 대상으로 가장 적절하지 않은 것?

--> 분석 성과에 대한 조사

16번) 분석 성숙도 모델 구성에서 고려하는 분석 성숙도 진단 부문으로 적절하지 않은 것?

--> 기업 문화 부문

단1) Value

단2) 시급성

17번) Wage R 코드 분석 결과 중 틀린 것?

--> age와 wage는 양의 상관관계를 가진다.

18번) 웹 데이터의 수집을 위해 웹페이지의 구조를 분석하여 데이터를 자동으로 수집하는 방법?

--> 웹 크롤링(Web Crawling)

19번) 오른쪽 꼬리가 긴(right-skewed)분포를 나타내는 자료의 평균과 중앙값의 관계?

--> 평균이 중앙값보다 큰 경향을 보인다.

20번) 앙상블 모형의 특징으로 가장 적절하지 않은 것?

--> 각 모형의 상호 연관성이 높을수록 정확도가 향상된다.

21번) 다차원척도법에 대한 설명으로 가장 적절하지 않은 것?

--> 개체들 사이의 유사성과 비유사성을 측정하여 차원을 축소하기 위해 사용한다.

22번) 이상치 판정 방법 중 가장 부적절한 것?

--> $Q2(\text{중위수}) + 1.5 \cdot IQR$ 보다 크거나 $Q2(\text{중위수}) - 1.5 \cdot IQR$ 작은 데이터를 이상치로 규정

23번) 표본추출의 방법으로 틀린 것?

--> 집단추출법

24번) R의 데이터 구조에서 숫자형, 문자형, 논리형을 모두 합쳐 하나의 벡터를 구성하였을 경우 합쳐진 벡터의 형식 ?

--> 문자형 벡터

25번) 신경망 모형에서 출력값이 여러 개이고 목표치가 다범주인 경우에 사용하는 활성화 함수?

--> 소프트맥스

26번) 표본들이 서로 관련된 경우 짝지어진 두 관찰치의 크고 작음을 표시하여 그 두 분포의 차이에 대한 가설을 검증하는 방법

--> 부호 검정(sign test)

27번) 인공신경망 모형에서 활성화 함수인 시그모이드(sigmoid) 함수의 결과값?

--> $0 \leq y \leq 1$

28번) 분류 모형의 평가를 위해 사용되는 방법으로 틀린 것?

--> 덴드로그램

29번) 자료의 측정수준에 대한 설명으로 부적절한 것

--> 비율척도는 ... 사칙연산이 가능하고 ... 혈액형, 학력 등이 해당된다.

30번) 군집분석에 대한 설명으로 적절하지 않은 것?

--> 군집분석은 집단 간 이질성과 집단 내 동질성이 모두 낮아지는 방향으로 군집을 만든다.

31번) 빅데이터 분석 프로세스에서 모델링 단계에 해당하지 않는 과정?

--> 수행방안 설계

32번) 코드 실행 결과에 대한 설명으로 적절한 것?

--> 회귀모형은 유의수준 5%하에서 통계적으로 유의미

33번) 아래 수식에 해당하는 데이터 간의 거리 계산 방식?

--> 맨하튼 거리

34번) 군집분석 기법으로 적절하지 않은 것?

--> Silhouette Coefficient

35번) 의사결정나무의 특징으로 틀린 것?

--> 비정상적인 잡음 데이터에 대해서는 민감하게 분류한다.

36번) 연관분석에 대한 특징 틀린 것?

--> 분석을 위한 계산이 복잡하다는 단점이 있다.

37번) 데이터 분할에 대한 설명으로 적절하지 않은 것?

--> 검증용 데이터는 학습 과정에서 사용되지 않는다.

38번) 변수 가공에 대한 설명으로 적절하지 않은 것?

--> 구간화의 개수가 감소하면 정확도는 높아지지만 속도가 느려진다.

39번) 모형평가 방법으로 적절하지 않은 것?

--> 엔트로피

40번) 데이터 전처리 과정에 대한 설명으로 맞는 것?

--> 데이터 특성을 파악하고 통찰을 얻기 위한 방법을 데이터 EDA라고 한다.

단 5번) AR 모형

단 6번) 랜덤 포레스트

단 7번) 6/19

단8) $a=3$, $b=8$

단9) 95.25

단10) 0.2