

2022년 32회 기출복원 인쇄용 (50문제)

 [tbcase73@gmail.com](#) (공유되지 않음) [계정 전환](#)



1. 다음 SQL의 명령어 중 DML이 아닌 것은 무엇인가?

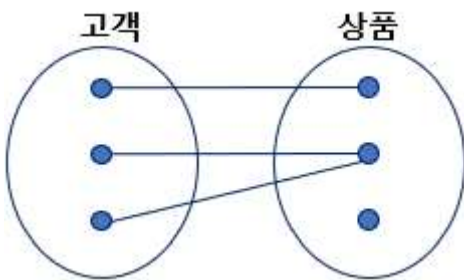
1점

가 - SELECT 나 - UPDATE 다 - INSERT 라 - DELETE 마 - CREATE

- ☐ 가, 나
- ☐ 다
- ☐ 라
- ☐ 마

2. 고객과 상품의 대응관계를 도식화 한 것이다. 대응비 관점에서 고객과 상품 간의 관계가 옳은 것은?

1점



- ☐ 1:1
- ☐ N:1
- ☐ N:M
- ☐ 1:N



3. 다음은 어떤 기업내부 데이터베이스 솔루션에 대한 설명인가?

1점

제조업을 포함한 다양한 비즈니스 분야에서 생산, 구매, 재고, 주문, 공급자와의 거래, 고객 서비스 제공 등 주요 프로세스 관리를 돕는 여러 모듈로 구성된 통합 애플리케이션

- ☐ ERP
- ☐ CRM
- ☐ SCM
- ☐ KMS

4. 다음 중 빅데이터 활용 요소 3가지로 올바른 것으로만 구성된 것은?

1점

- ☐ 데이터, 인프라, 기술
- ☐ 데이터, 인력, 기술
- ☐ 인프라, 프로세스, 인력
- ☐ 프로세스, 인력, 기술

5. 다음 중 딥러닝의 종류가 아닌 것은?

1점

- ☐ SVM
- ☐ DNN
- ☐ CNN
- ☐ RNN



6. 지도 학습은 입력과 출력 모두 주어진 상태에서의 학습이며, 비지도 학습은 입력만 주어진 상태에서의 학습이다. 다음의 보기에서 그 종류가 다른 것은 무엇인가? 1점

가. 회귀 분석 나. 군집 분석 다. 감정 분석 라. 분류 분석

- ☐ 가
- ☐ 나
- ☐ 다
- ☐ 라

7. 빅데이터와 인문학 열풍과 관련된 내용으로 상관 없는 것은? 1점

- ☐ 단순 세계에서 복잡한 세계로의 변화
- ☐ 비즈니스의 중심이 제품생산에서 서비스로 이동
- ☐ 빅데이터 분석 기법 및 방법론의 확대
- ☐ 경제와 산업의 논리가 생산에서 시장창조로 바뀜

8. 데이터 사이언티스트의 역량에 대한 설명이 아닌 것은? 1점

- ☐ 통찰력 있는 분석, Data Technical Skill
- ☐ 설득력 있는 전달
- ☐ 다분야간 협력
- ☐ 데이터 분석 알고리즘으로 인해 피해를 입은 사람을 구제함



9. 다음은 무엇에 대한 설명인가?

0점

구글의 인터넷에 연결된 네스트는 날씨와 기온 정보 그리고 집주인의 평소 온도 설정 DATA를 기반으로 사용자의 Context를 인식해 자동으로 온도를 설정해 주며, 아마존은 Dash라는 작은 장치는 WiFi가 내장된 바코드 인식기로 상품에 인쇄된 바코드를 Dash로 비추게 되면 그 상품을 아마존 장바구니에 저장할 수 있도록 해준다. 나이키의 경우 애플과 제휴하여 스마트한 운동 관리를 할 수 있도록 해주는 서비스로 자리 매김했다. 굳이 우리가 기계를 조작하지 않아도 모든 것이 사람을 위해 알아서 자동으로 돌아가는 세상이 이것이 보여줄 미래이다

내 답변

10. 다음은 어떤 기업내부데이터 솔루션에 대한 설명인가?

1점

제조, 물류, 유통업체 등 유통공급망에 참여하는 모든 업체들이 협력을 바탕으로 정보 기술 (Information Technology)을 활용, 재고를 최적화하기 위한 솔루션이다

내 답변

1. 분석 기회 발굴의 범위 확장에서 '거시적 관점'이 아닌 것은?

1점

- ☐ 사회, 기술
- ☐ 경제
- ☐ 환경, 정치
- ☐ 고객



2. ROI의 4V 중 Return에 해당하는 것은?

1점

- ☐ Volume
- ☐ Variety
- ☐ Velocity
- ☐ Value

3. 다음 중 데이터 표준화에 대한 설명으로 올바른 것은?

1점

- ☐ 메타데이터와 데이터 사전의 관리 원칙을 수립한다
- ☐ 데이터 표준 용어 설정, 명명 규칙수립, 메타데이터 구축, 데이터 사전 구축 등의 업무로 구성된다
- ☐ 메타데이터 및 표준 데이터를 관리하기 위한 전사 차원의 저장소를 구성한다
- ☐ 데이터 거버넌스 체계를 구축한 후 표준 준수 여부를 주기적으로 점검하고 모니터링을 실시한다

4. 빅데이터 분석 방법론의 분석기획 단계의 Task 중에서 '비즈니스 이해 및 범위 설정' 단계에서 프로젝트에 참여하는 관계자들의 이해를 일치시키기 위해 작성하는 것은 ?

1점

- ☐ SOW(Statement of Work)
- ☐ WBS(Work Breakdown Structure)
- ☐ EDA(Exploratory Data Analysis)
- ☐ ETL(Extract Transformation Loading)



5. 기업에서 사용하는 데이터의 가용성, 유용성, 통합성, 보안성을 관리하기 위한 정책과 프로세스를 다루며 프라이버시, 보안성, 데이터품질, 관리규정 준수를 강조하는 것을 무엇이라고 하는가? 1점

- ☐ 데이터 표준화 활동
- ☐ 분석 마스터 플랜
- ☐ 데이터 거버넌스
- ☐ 메타 데이터

6. 빅데이터 분석 방법론의 순서로 올바른 것은? 1점

- ☐ 분석 기획 - 데이터 준비 - 데이터 분석 - 시스템 구현 - 평가 및 전개
- ☐ 데이터 준비 - 분석 기획 - 데이터 분석 - 시스템 구현 - 평가 및 전개
- ☐ 데이터 준비 - 분석 기획 - 데이터 분석 - 평가 및 전개 - 시스템 구현
- ☐ 분석 기획 - 데이터 준비 - 데이터 분석 - 평가 및 전개 - 시스템 구현

7. 하향식 접근방식에 대한 설명으로 틀린 것은? 1점

- ☐ 하향식 접근 방식의 데이터 분석 기획의 단계는 문제 탐색, 문제 정의, 해결 방안 탐색, 타당성 검토 순서이다
- ☐ 분석 유즈 케이스는 향후 데이터 분석 문제로의 전환 및 적합성 평가에 활용하도록 한다
- ☐ 하향식 접근 방식은 유사/동종 사례 벤치마킹을 통한 분석 기회 발굴을 하는 내부 참조 모델 기반 문제 탐색을 한다
- ☐ 분석 유즈 케이스는 풀어야 할 문제에 대한 상세 설명 및 해당 문제를 해결했을 때 발생하는 효과를 명시한다



8. 분석 준비도 프레임워크에 대한 내용이 아닌 것은?

1점

- ☐ 분석 인력 및 조직, 분석 인프라
- ☐ 분석 업무 파악, 분석 데이터
- ☐ 분석 성과, 분석 문화
- ☐ 분석 인력 및 조직, 분석 문화

9. 다음 분석 성숙도 단계 중 괄호에 들어갈 단어는 무엇인가?

1점

도입 단계 - () 단계 - 확산 단계 - 최적화 단계

내 답변

10. 합리적인 의사결정을 방해하는 요소로써 문제의 표현 방식에 따라 동일한 사건이나 상황임에도 불구하고 사람들의 선택이나 판단이 달라지는 현상을 무엇이라 하는가?

1점

내 답변

1. 다음 중 R에서 서로 다른 데이터 타입을 담을 수 있는 구조는 무엇인가?

1점

- ☐ matrix
- ☐ list
- ☐ array
- ☐ vector



2. 박스플롯(boxplot)에서 중앙에 있는 값 50%를 무엇이라고 하는가?

1점

- ☐ 범위
- ☐ 1 사분위수 범위
- ☐ IQR(사분위소 범위)
- ☐ 3 사분위수 범위

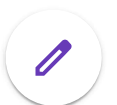
3. 박스플롯(boxplot)에서 중간에 있는 선은 무엇인가?

1점

- ☐ median
- ☐ mean
- ☐ IQR
- ☐ outlier

4. Boxplot에서 상한(최댓값)과 하한(최솟값)은 얼마인가? $Q1(1\text{사분위수}) = 4$, 1점
 $Q3(3\text{사분위수}) = 12$

- ☐ 하한 = -8, 상한 = 24
- ☐ 하한 = -6, 상한 = 22
- ☐ 하한 = -4, 상한 = 20
- ☐ 하한 = -2, 상한 = 18



5. 1종 오류의 최대 허용 한계에 관한 확률을 나타내는 것은?

1점

- ☐ 유의수준
- ☐ 기각역
- ☐ 검정통계량
- ☐ 2종 오류

6. 표본조사에 대한 설명이 부적절한 것은?

1점

- ☐ 표본 오차(Sampling Error)는 모집단의 일부인 표본에서 얻은 자료를 통해 모집단 전체의 특성을 추론함으로써 생기는 오차를 의미한다.
- ☐ 표본 편의(Sampling Bias)는 표본추출방법에서 기인하는 오차를 의미한다.
- ☐ 표본 편의는 확률화(Randomization)에 의해 최소화하거나 없앨 수 있다.
- ☐ 비표본 오차(non-sampling error)는 표본크기가 증가함에 따라 감소한다.

7. 다음 중 비모수 검정의 종류가 아닌 것은?

1점

- ☐ Runs Test
- ☐ 카이제곱검정
- ☐ Wilcoxon Signed Rank Test
- ☐ Sign Test



8. 파생 변수에 대한 설명으로 옳지 않은 것은?

1점

- ☐ 요약 통계량을 이용한 변수 생성을 통해 데이터 분석이 가능하다
- ☐ 데이터에 포함된 나이 연속형 변수를 이용해 연령대 범주형 파생변수를 추가했다
- ☐ 파생 변수를 해당 데이터셋을 대표하는 변수로 사용할 수 있다
- ☐ 파생 변수는 매우 주관적일 수 있으므로 논리적 타당성을 갖추어 개발해야 한다

9. 다음이 설명에 해당 하는 것은 무엇인가?

1점

공분산행렬 또는 상관관계수 행렬을 사용해 모든 변수들을 가장 잘 설명하는 변수를 찾는 방법으로, 상관관계가 있는 변수들을 선형 결합에 의해 상관관계가 없는 새로운 변수를 만들고 분산을 극대화하는 변수로 축약하는 방법으로 새로운 변수들은 변수들의 선형결합으로 이루어져 있다

- ☐ 요인 분석
- ☐ 회귀 분석
- ☐ 주성분 분석(PCA)
- ☐ 다차원척도법

10. 다음 주성분 분석에 대한 설명 중 적절하지 않은 것은?

1점

- ☐ 가장 분산이 적은 것을 제1주성분으로 설정한다.
- ☐ 주성분 분석은 상관관계가 있는 변수들을 결합해 상관관계가 없는 변수로 분산을 극대화하는 변수로 선형결합을 해 변수를 축약하는데 사용하는 방법이다
- ☐ 공분산 행렬은 변수의 특정단위를 그대로 반영한 것이고, 상관행렬은 모든 변수의 측정단위를 표준화한 것이다
- ☐ 공분산행렬을 이용한 분석의 경우 변수들의 측정 단위에 민감하다



11. 시간의 흐름에 따라 관측된 데이터를 무엇이라 하는가?

1점

- ☐ 주성분분석
- ☐ 회귀분석
- ☐ 시계열 자료
- ☐ 군집 분석

12. 다음 시계열 데이터에 대한 설명 중 가장 옳지 않은 것은?

1점

- ☐ 시계열 데이터의 모델링은 다른 분석 모형과 같이 탐색 목적과 예측 목적으로 나눌 수 있다.
- ☐ 짧은 기간 동안의 주기적인 패턴을 계절변동이라 한다.
- ☐ 잡음은 무작위적인 변동이지만 일반적으로 원인은 알려져 있다.
- ☐ 시계열분석의 주목적은 외부인자와 관련해 계절적인 패턴 추세와 같은 요소를 설명할 수 있는 모델을 결정하는 것이다.

13. 다음 시계열 자료의 정상성(stationary)에 대한 설명 중 가장 부적절한 것은? 1점

- ☐ 모든 시점에 대해 일정한 평균을 가진다.
- ☐ 모든 시점에 대해 일정한 분산을 가진다.
- ☐ 공분산은 단지 시차에만 의존하고 시점 자체에는 의존하지 않는다.
- ☐ 데이터에 포함된 잡음은 독립성을 갖는다.



14. 다음이 설명하는 데이터 마이닝 기법은 무엇인가?

1점

미리 정의된 기준이나 예시에 의해서가 아닌 레코드 자체가 가진 다른 레코드와의 유사성에 의해 그룹화되고 이질성에 의해 세분화 되는 것으로 데이터 마이닝이나 모델링의 준비단계로서 사용되는 기법이다.

- ☐ 군집 분석
- ☐ 분류 분석
- ☐ 기술(description)
- ☐ 연관 분석

15. 다음 중 오분류표의 평가지표 중 True로 예측한 관측치 중 실제 True인 지표를 1점 무엇이라 하는가?

- ☐ Precision
- ☐ Accuracy
- ☐ F1
- ☐ Sensitivity



16. 아래 오분류표를 이용하여 구한 민감도(sensitivity) 값은 얼마인가?

1점

Confusion matrix		예측값	
		TRUE	FALSE
실제값	TRUE	40	60
	FALSE	60	40

- ☐ 0.4
- ☐ 0.3
- ☐ 0.15
- ☐ 0.5

17. 아래의 오분류표에서 특이도 값을 구하는 식을 구하시오.

1점

confusion matrix		예측값	
		T	F
실제값	T	TP	FN
	F	FP	TN

- ☐ $TP / (TP + FN)$
- ☐ $(TP + TN) / (TP + FP + FN + TN)$
- ☐ $TN / (TN + FP)$
- ☐ $TP / (TP + FP)$



18. 거리를 활용한 측도에 대한 설명으로 틀린 것은?

1점

- ☐ 유클리드는 두 점 사이의 거리로, 가장 직관적이고 일반적인 거리의 개념이다
- ☐ 맨해튼 거리는 두 점의 좌표 간의 절대값 차이를 구하는 것이다
- ☐ 마할라노비스는 변수의 표준화를 고려하고, 변수 간의 상관성을 고려하지 않는다
- ☐ 표준화, 마할라노비스 거리는 통계적 거리의 개념이다

19. K-fold 교차검증에 대한 설명으로 옳바르지 않은 것은?

1점

- ☐ $k=2$ 인 경우, LOOCV라고 한다.
- ☐ 주어진 데이터를 가지고 K 번 반복적으로 성과를 측정해 그 결과를 평균한다.
- ☐ K-fold 는 데이터를 K 개로 분할하는 것을 의미한다.
- ☐ 데이터가 충분하지 않은 경우 주로 사용한다.

20. 비계층적 군집분석인 K-means 군집분석의 수행순서는?

1점

가) 초기 군집의 중심으로 k 개의 객체를 임의로 선택한다
 나) 각 자료를 가장 가까운 군집 중심에 할당한다
 다) 각 군집 내의 자료들의 평균을 계산하여 군집의 중심을 갱신한다
 라) 군집 중심의 변화가 거의 없을 때까지 나와 다를 반복한다

- ☐ 가-나-다-라
- ☐ 나-가-다-라
- ☐ 다-나-가-라
- ☐ 라-가-나-다



21. 여러 분포가 확률적으로 선형 결합된 형태로, 데이터가 k 개의 모수적 모형의 가중합으로 표현되는 모집단 모형에서 나왔다는 가정하에, 추정된 k 개의 모형 중 어느 모형으로부터 나왔을 확률이 높은지에 따라 군집 분류를 수행하는 것은 무엇인가? 1점

- ☐ 평균연결법
- ☐ 혼합분포군집
- ☐ SOM
- ☐ DBSCAN

22. 교차판매/물건배치 등에 이용되는 분석 기법은?

1점

- ☐ 연관분석
- ☐ 주성분분석
- ☐ 회귀분석
- ☐ SOM



23. A -> B의 신뢰도는?

1점

장바구니	item
1	A, B, C
2	A, C
3	B, C
4	A, B
5	A, C, D

- ☐ 0.75
- ☐ 0.65
- ☐ 0.5
- ☐ 0.4



24. 다음은 어느 마트의 판매품목에 대한 결과이다. 커피 -> 우유에 대한 향상도 1점
는?

장바구니	item
1	커피, 빵
2	커피, 계란, 우유
3	계란, 커피, 빵, 우유
4	계란, 우유
5	커피, 우유
6	커피, 빵, 계란
7	우유, 빵

- ☐ 0.6
- ☐ 1.25
- ☐ 0.43
- ☐ 0.84

25. 시계열에 영향을 주는 일반적인 요인을 시계열에서 분리해 분석하는 방법을 1점
무엇이라 하는가?

내 답변



26. 시계열 모형의 여러 종류 중 아래에서 설명하는 것은 무엇인가?

1점

가) 시계열 모델 중 자기 자신의 과거 값을 사용하여 설명하는 모형임
 나) 백색 잡음의 현재 값과 자기 자신의 과거 값의 선형 가중합으로 이루어진 정상 확률 모형
 다) 모형에 사용하는 시계열 자료의 시점에 따라 1차, 2차, ..., p차 등을 사용하나
 정상시계열 모형에서는 주로 1, 2차를 사용함

내 답변

27. 학습 데이터에 너무 잘 맞게 학습되어 학습 데이터에 대한 성능은 매우 높지만 평가 데이터에 대한 성능은 낮은 상태를 무엇이라고 하는가?

1점

내 답변

28. 다음이 설명하는 기업 내부 데이터베이스 솔루션은 무엇인가?

1점

전사적으로 구축된 데이터웨어 하우스로부터 특정 주제, 부서 중심으로 구축된 소규모 단일 주제의 데이터웨어 하우스로 재무, 생산, 운영과 같이 특정 조직의 특정 업무 분야에 초점을 두고 있다.

내 답변

29. 사람 뇌의 뉴런을 모방하여 분류 및 예측에 사용할 수 있는 모형으로 입력층, 은닉층, 출력층 3개의 층으로 구성되어 있고, 각 층에 뉴런이 여러 개 포함되어 있는 것은 무엇인가?

1점

내 답변



30. 모형의 성능을 평가하기 위한 척도로 얼마나 예측이 잘 이루어졌는지를 나타내기 위해 임의로 나눈 각 등급별로 반응검출율, 반응률, 향상도 등의 정보를 산출하여 나타내는 것은 무엇인가? 1점

내 답변

제출

양식 지우기

Google Forms를 통해 비밀번호를 제출하지 마세요.

이 콘텐츠는 Google이 만들거나 승인하지 않았습니다. [악용사례 신고](#) - [서비스 약관](#) - [개인정보처리방침](#)

Google 설문지

