

2020년 A형 - 인쇄용 (50문제)

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



* 필수항목

1. 다음 빅데이터 거버넌스에 대한 설명으로 옳지 않은 것은? *

1점

- ☐ ERD는 운영 중인 데이터베이스와 일치하기 위하여 철저한 변경관리가 필요하다
- ☐ 양질의 데이터가 중요하므로 수명주기보다 데이터 품질관리가 중요하다
- ☐ 산업 분야별, 데이터 유형별, 정보 거버넌스 요소별로 구분하여 작성한다
- ☐ 분석 조직 및 인력에 대해 지속적인 교육과 훈련을 실시해야 한다

2. 다음 중 빅데이터 활용 테크닉 관계가 올바르게 짝지어 진 것은? *

1점

- 회귀분석 : 사용자의 만족도가 충성도에 어떤 영향을 미치는가?
- 연관규칙학습: 호텔에서 고객의 논평을 받아 서비스를 개선하기 위해 활용함
- 유전알고리즘: 응급실에서 의사를 어떻게 배치하는 것이 가장 효율적인가?
- 감정분석: 우유구매자가 기저귀도 같이 구매하는가?

- ☐ 회귀분석, 연관규칙학습
- ☐ 감정분석, 유전알고리즘
- ☐ 감정분석, 연관규칙학습
- ☐ 회귀분석, 유전알고리즘



3. 다음 중 데이터베이스 설계 절차를 나열한 것으로 올바른 것은? *

1점

- ☐ 요구조건분석 - 개념적 설계 - 논리적 설계 - 물리적 설계
- ☐ 개념적 설계 - 논리적 설계 - 요구조건분석 - 물리적 설계
- ☐ 논리적 설계 - 개념적 설계 - 요구조건분석 - 물리적 설계
- ☐ 물리적 설계 - 개념적 설계 - 요구사항분석 - 논리적설계

4. 다음 중 사생활 침해에 따른 문제에 대한 해결책은 무엇인가? *

1점

- ☐ 알고리즘 접근 허용
- ☐ 결과기반 책임 원칙 강화
- ☐ 정보사용자의 책임제로 전환
- ☐ 알고리즘미스트의 육성

5. 다음 중 데이터베이스 구성요소에 설명 중 올바른 것은? *

1점

- 메타데이터 : 데이터에 관한 구조화된 데이터로, 다른 데이터를 설명해 주는 데이터
- 인덱스 : 데이터베이스 분야에 있어서 테이블에 대한 동작의 속도를 높여주는 자료 구조
- 데이터 사전 : 사용자의 의사결정에 도움을 주기 위해 기간시스템의 데이터베이스에 축적된 데이터를 공통의 형식으로 변환해서 관리하는 데이터베이스
- 어트리뷰트 : 테이블의 행을 의미함

- ☐ 메타데이터, 인덱스
- ☐ 어트리뷰트, 인덱스
- ☐ 인덱스, 데이터 사전
- ☐ 메타데이터, 데이터 사전



6. DIKW 단계를 설명하는 것 중 성질이 다른 것은? *

1점

- ☐ A반 학생의 평균 점수는 80점, B반은 82점이다
- ☐ B마트의 소고기 600g의 가격은 54000원이다
- ☐ B마트의 가격이 더 싸다
- ☐ 개인의 구글 하루 평균 방문 빈도는 10회이다

7. 다음 중 빅데이터 본질적인 변화로 옳지 않은 것은? *

1점

- ☐ 표본조사에서 전수조사로 변화
- ☐ 상관관계에서 인과관계로 변화
- ☐ 사전처리에서 사후처리로의 변화
- ☐ 질보다는 양을 중시하게 되었음

8. 다음 중 딥러닝(Deep Learning)과 가장 관련 없는 분석 기법은? *

1점

- ☐ LSTM
- ☐ Autoencoder
- ☐ KNN
- ☐ CNN

9. DIKW 피라미드 계층구조에서 데이터의 가공 및 상관/연관 관계 속에서 의미 * 1점
가 도출된 것을 무엇이라 하는가?

내 답변



10. 다음 보기에서 설명하는 빅데이터 활용 기법을 무엇이라 하는가? *

1점

- 구매를 많이 하는 집단에 속하는지 아닌 집단에 속하는지에 대한 문제 해결에 사용함
- 문서를 분류하거나 조직을 그룹으로 나눌 때 사용함

내 답변

1. 다음 중 CRISP-DM의 모델링 단계의 Task가 아닌 것은 무엇인가? *

1점

- ☐ 모델링 기법 선택
- ☐ 모델 테스트 계획 설계
- ☐ 모델 평가
- ☐ 모델 적용성 평가

2. 다음 중 분석 과제 정의서에 대한 설명 중 옳지 않은 것은? *

1점

- ☐ 분석 과제 정의서는 분석 모델에 적용될 알고리즘과 분석 모델의 기반이 되는 특성 (Feature)이 포함될 필요는 없다.
- ☐ 분석 과제 정의서는 프로젝트를 수행하는 이해관계자가 프로젝트의 방향을 설정하지만, 성공여부를 판별에는 사용할 수 없는 자료이다.
- ☐ 분석 과제 정의서는 소스데이터, 데이터 입수 및 분석의 난이도, 분석 방법 등에 대한 항목이 포함되어야 한다.
- ☐ 분석 과제 정의서는 프로젝트 계획서를 작성하기 위한 중간 결과로써 구성항목으로 도출할 필요가 있다.



3. 빅데이터 분석 방법론에서 단계 간 피드백이 반복적으로 많이 발생할 수 있는 * 1점 단계는?

- ☐ 분석 기획 - 데이터 준비
- ☐ 데이터 준비 - 데이터 분석
- ☐ 시스템구현 - 평가 및 전개
- ☐ 평가 및 전개 - 분석 기획

4. 다음 중 ROI 관점에서의 분석 과제에 대한 우선순위 평가 기준에 대한 설명 중 * 1점 적절하지 않은 것은?

- ☐ 시급성이 높고 난이도가 낮은 분석과제를 일반적으로 우선순위가 높다.
- ☐ 시급성의 판단 기준은 전략적 중요도 및 목표가치가 핵심이다.
- ☐ 난이도는 경영진 또는 실무 담당자의 의사결정에 따라 적용 우선순위를 조정할 수 없다.
- ☐ 난이도는 현 시점에서 과제를 추진하는 것이 비용측면과 범위측면에서 바로 적용하기 쉬운 것인지 또는 어려운 것인지에 대한 판단 기준이다.

5. 다음 중 데이터 거버넌스 체계 요소 중 데이터 표준 용어 설정, 명명규칙 수립, * 1점 메타데이터 구축, 데이터 사전 구축 등의 업무 구성의 요소를 무엇이라 하는가?

- ☐ 데이터 표준화
- ☐ 데이터 관리 체계
- ☐ 데이터 저장소 관리
- ☐ 표준화 활동



6. 다음 데이터 유형 중 정형-반정형-비정형의 순서로 올바르게 연결된 것은? * 1점

- ☐ Demand Forecast – Competitor Pricing – Email Record
- ☐ Twitter Feeds – ERP – Web Log
- ☐ RFID – CRM – Email Record
- ☐ IoT – Facebook comment – Weather data

7. 다음 프로토타이핑(Prototyping) 접근 방법에 관한 설명 중 옳은 것은? * 1점

- ☐ 신속하게 해결책 모형 제시, 상향식 접근방법에 활용 한다.
- ☐ 빠른 결과보다 모델의 안정성에 중점을 둔 기법이다
- ☐ 폭포수 방식처럼 전체적인 플랜을 짜고 문서를 통해 개발한다
- ☐ 대표적인 하향식 접근방법이다

8. 분석기획 발굴의 범위 확장에 관한 설명 중 적절하지 않은 것은? * 1점

- ☐ 역량의 재해석 관점에서는 현재 해당 조직 및 기업이 보유한 역량 뿐만 아니라 해당 조직의 비즈니스에 영향을 끼치는 파트너 네트워크를 포함한 활용 가능한 역량을 토대로 폭넓은 분석 기회를 탐색한다
- ☐ 경쟁자 확대 관점에서는 현재 수행하고 있는 사업 영역의 제품, 서비스에 대해서만 분석 기획 발굴의 폭을 넓혀서 탐색한다.
- ☐ 시장의 니즈 탐색 관점에서는 현재 수행하고 있는 사업에서의 고객 뿐만 아니라 고객과 접촉하는 역할을 수행하는 채널 및 고객의 구매와 의사결정에 영향을 미치는 영향자들에 대한 폭넓은 관점을 바탕으로 분석 기회를 탐색한다
- ☐ 거시적 관점에서는 현재의 조직 및 해당 산업에 폭넓게 영향을 미치는 사회, 경제적 요인을 STEEP 영역으로 나누어 좀 더 폭넓게 기회 탐색을 수행한다



9. 빈 칸에 들어갈 용어는 무엇인가? *

1점

분석 과제 관리 프로세스는 크게 과제 발굴과 (1) 으로 나누어진다. 분석 아이디어와 분석 과제가 확정이 되면 팀을 구성하고 (2) 하고 분석 과제 진행 관리 및 결과를 공유하고 개선하는 절차를 수행한다.

내 답변

10. 합리적인 의사결정을 방해하는 요소로써 문제의 표현 방식에 따라 동일한 * 1점
사건이나 상황임에도 불구하고 사람들의 선택이나 판단이 달라지는 현상을 무엇이라 하는가?

내 답변

1. 두 집단의 분산이 같은지를 검정할 때 사용되는 검정 통계1.량은 어떤 분포를 * 1점
활용하는 것이 가장 적절한가?

- ☐ t-분포
- ☐ F-분포
- ☐ z-분포
- ☐ 카이제곱 분포



2. 다음 중 데이터의 집중경향치와 산포도에 관한 설명으로 틀린 것은? *

1점

- ☐ 중앙값은 대표적인 집중경향치로 이상값 및 다른 관측값에 의한 영향에 민감하다는 단점이 있다.
- ☐ 평균은 데이터의 총 값을 총 개수로 나눈 값을 의미한다.
- ☐ 분산은 관측값에서 평균을 뺀 값을 제곱하고, 그것을 모두 더한 후 전체 개수로 나누어 구한다.
- ☐ 최빈값은 가장 많이 관측되는 수로, 주어진 자료에서 평균, 중앙값을 구하기 어려운 경우 특히 유용하다.

3. 자료의 척도에 대한 설명으로 부적절한 것은? *

1점

- ☐ 명목척도 : 단순히 측정 대상의 특성을 분류하거나 확인하기 위한 목적으로 사용된다.
- ☐ 서열척도 : 대소 또는 높고 낮음 등의 순위만 제공할 뿐 양적인 비교는 할 수 없다.
- ☐ 등간척도 : 순위를 부여하되 순위 사이의 간격이 동일하여 양적인 비교가 가능하다.
- ☐ 비율척도 : 절대 영점이 존재하지 않으며, 측정값 사이의 비율 계산이 가능한 척도이다.



4. 다음은 어느 문구점의 판매품목 및 거래수에 대한 결과이다. 연필 --> 지우개 * 1점에 대한 향상도는?

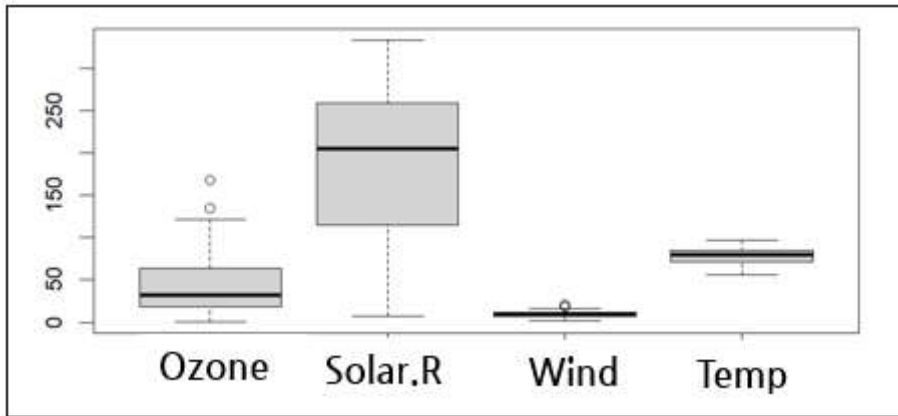
항목	거래수
연필	200
지우개	300
노트	200
연필, 노트	100
지우개, 연필	150
연필, 지우개, 노트	50
전체 거래건수	1000

- ☐ 0.8
- ☐ 3.3
- ☐ 0.2
- ☐ 1.0



5. 다음 airquality 데이터셋에 대한 설명 중 적절하지 않은 것은? *

1점



- ☐ Solar.R 중위수가 가장 크
- ☐ Solar.R 분산이 가장 크다고 할 수 있다
- ☐ Ozone은 이상값이 2개가 존재한다
- ☐ Temp의 데이터 수가 가장 많음을 알 수 있다

6. 오분류표를 활용하여 모델을 평가하는 지표 중 실제값이 True인 것에 대해 예측값이 True인 것을 무엇이라 하는가? *

1점

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



7. 다음 중 모델의 학습 과정에서 학습 및 검증데이터를 나눌 때, 단순히 한 번 나누는 게 아니라 K 번 나누고 각각의 학습 모델의 성능을 비교하여 평균값으로 분류분석 모형을 평가하는 방법을 무엇이라 하는가?

- ☐ K-fold
- ☐ 홀드아웃
- ☐ 붓스트랩
- ☐ 군집분석

8. 다음 두 점의 맨해튼 거리는 무엇인가?

1점

A(10, 6) B(3, 4)

- ☐ 12
- ☐ 5
- ☐ 9
- ☐ 10

9. 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.



10. 다음 주성분 분석 결과에 대한 설명 중 옳바르지 않는 것은? *

1점

```
> fit <-prcomp(USArrests, scale=TRUE)
> summary(fit)
Importance of components:

                PC1      PC2      PC3      PC4
Standard deviation  1.5749  0.9949  0.59713  0.41645
Proportion of Variance 0.6201 0.2474 0.08914 0.04336
Cumulative Proportion 0.6201 0.8675 0.95664 1.00000
```



```
> fit$rotation

                PC1      PC2      PC3      PC4
Murder    -0.5358995  0.4181809 -0.3412327  0.64922780
Assault   -0.5831836  0.1879856 -0.2681484 -0.74340748
UrbanPop  -0.2781909 -0.8728062 -0.3780158  0.13387773
Rape      -0.5434321 -0.1673186  0.8177779  0.08902432
```

- ☐ 주성분 분석에 대한 결과로 첫 번째 주성분으로 전체 데이터 분산의 62%를 설명할 수 있다
- ☐ 두 번째 주성분은 4개의 변수 모두와 양의 관계에 있음을 보여준다
- ☐ 변수들이 선형결합으로 이루어진 주성분은 서로 독립이며 기존 자료보다 적은 수의 주성분들로 기존 자료의 변동을 설명한다.
- ☐ 위 주성분 분석에서는 상관행렬을 사용하였다



11. 아래 표는 불순도 측정 결과이다. 지니 지수는 얼마인가? *

1점



- ☐ 12/25
- ☐ 13/25
- ☐ 3/5
- ☐ 2/5

12. 데이터마이닝의 상품에 관한 이해를 증가시키기 위한 것으로 데이터의 특징 * 1점
및 의미를 표현 및 설명하는 기능을 무엇이라 하는가?

- ☐ 분류(Classification)
- ☐ 예측(Forecast)
- ☐ 추정(Estimate)
- ☐ 기술(Description)

13. 다음 중 신뢰구간에 대한 설명 중 가장 적절하지 않은 것은? *

1점

- ☐ 구간추정은 모수의 참값이 포함되어 있으리라고 추정되는 구간을 결정하는 것이다
- ☐ 관측치의 크기가 커지면 신뢰구간의 길이는 줄어든다
- ☐ 신뢰수준 95%란 신뢰구간에 미지의 수가 포함되지 않을 확률이 95%를 의미한다
- ☐ 점추정의 정확성을 보완하는 방법이다



14. 다음은 Default 데이터를 사용한 고객에 대한 채납 여부(default)에 대한 분석 결과에 대한 설명으로 옳지 않은 것은? * 1점

```
> model = glm(default ~ ., data=Default, family=binomial)
> summary(model)

Call:
glm(formula = default ~ ., family = binomial, data = Default)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4691  -0.1418  -0.0557  -0.0203   3.7383

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.087e+01  4.923e-01 -22.080  < 2e-16 ***
studentYes   -6.468e-01  2.363e-01  -2.738  0.00619 **
balance       5.737e-03  2.319e-04  24.738  < 2e-16 ***
income        3.033e-06  8.203e-06   0.370  0.71152
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1571.5  on 9996  degrees of freedom
AIC: 1579.5

Number of Fisher Scoring iterations: 8
```

default -> 고객의 채무불이행 확률
 studentYes : 학생여부(Yes/No)
 balance : 채무잔액, income : 연수입

- ☐ balance는 default에 통계적으로 유의미한 영향을 주는 변수이다
- ☐ 로지스틱 회귀분석을 사용했다.
- ☐ income은 default에 통계적으로 유의미한 영향을 주는 변수이다
- ☐ studentYes의 값이 Yes 일 때, 채무불이행(default) 될 확률이 낮다.



15. 다음과 같은 방법을 무엇이라 하는가? *

1점

데이터 전처리 방법 중 데이터를 일정범위로 Feature scaling 범위 0~1사이로 적용해주고 원 데이터의 분포를 유지하는 정규화 방법

- ☐ Robust Normalization
- ☐ Min-Max Normalization
- ☐ Standardization
- ☐ Smooting

16. 다음 중 결측값(missing value) 처리에 대한 대치법(imputation)에 관한 설명으로 틀린 것은? * 1점

- ☐ complete case analysis는 불완전 자료는 모두 무시하고 완전하게 관측된 자료만으로 표준적 통계기법에 의해 분석하는 방법을 말한다.
- ☐ 평균대치법(mean imputation)은 관측 또는 실험되어 얻어진 자료의 적절한 평균값으로 결측값을 대치해서 불완전한 자료를 완전한 자료로 만든 후, 완전한 자료를 마치 관측 또는 실험되어 얻어진 자료라 생각하고 분석하는 방법을 말한다.
- ☐ 단순확률대치법(single stochastic imputation)은 평균대치법에서 추정량 표준오차의 과소추정 문제를 보완하고자 고안된 방법이다.
- ☐ 다중대치법은 추정량의 과소추정이나 계산의 난해성 문제를 보완하는 방법이다.

17. 다음 중 앙상블 모형이 아닌 것은? *

1점

- ☐ 배깅(bagging)
- ☐ 랜덤 포레스트(Random Forest)
- ☐ 시그모이드(sigmoid)
- ☐ 부스팅(boosting)



18. 계층적 군집의 응집형 군집 방법과 관련이 없는 것은?

1점

- ☐ 최단연결법
- ☐ 와드연결법
- ☐ k-중앙값군집
- ☐ 평균연결법

19. 귀무가설이 사실일 때 기각하는 1종 오류 시 우리가 내린 판정이 잘못되었을 때 실제 확률은? * 1점

- ☐ 유의수준
- ☐ p-value
- ☐ 기각역
- ☐ 대립가설

20. 표본조사에 대한 설명이 부적절한 것은? *

1점

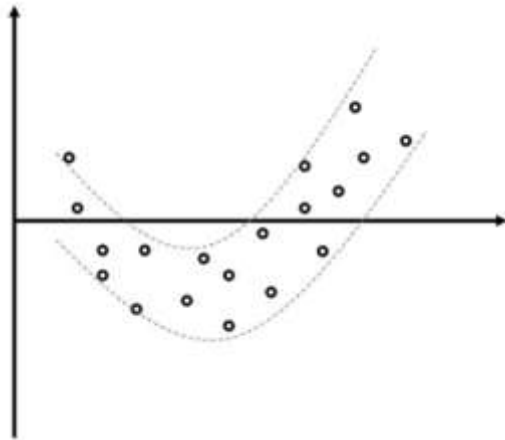
- ☐ 표본 오류(Sampling Error)는 모집단을 대표하지 못하는 표본을 추출하여 발생하는 오류이다.
- ☐ 표본편의(Sampling Bias)는 표본추출방법에서 기인하는 오차를 의미한다.
- ☐ 표본편의는 확률화(Randomization)에 의해 최소화하거나 없앨 수 있다.
- ☐ 비표본오차(non-sampling error)는 표본크기가 증가함에 따라 감소한다.



21. 독립변수간 상관관계가 높아 많은 문제점을 발생하는 현상으로 회귀계수의 분산을 증가시켜 불안정하고 해석하기 어렵게 만들게 되는 것을 무엇이라 하는가? 1점

- ☐ 다중공선성
- ☐ 공분산행렬
- ☐ 후진제거법
- ☐ 전진선택법

22. 회귀분석의 잔차 분석 결과가 다음과 같을 때 해결 방안은? * 1점



- ☐ x^2 항을 모형에 추가
- ☐ 데이터를 좀 더 수집함
- ☐ 변수 제거
- ☐ 중요 변수 선택



23. Q1, Q3가 다음과 같을 때 이상값 판단을 위한 하한 및 상한은 얼마인가?

1점

Q1 = 4, Q3 = 12

- ☐ 하한 = -8, 상한 = 24
- ☐ 하한 = 24, 상한 = -8
- ☐ 하한 = -8, 상한 = 16
- ☐ 하한 = -4, 상한 = 16

24. 연관분석에 대한 설명이 아닌 것은? *

1점

- ☐ 연관규칙(Association rule)은 항목들 간의 '조건-결과' 식으로 표현되는 유용한 패턴이다
- ☐ 장바구니 분석이라고도 한다.
- ☐ Apriori 알고리즘과 FP Growth 가 대표적이다.
- ☐ 연관규칙 측정지표에는 정확도, 정밀도, 민감도 등이 있다



25. 다음 중 다중회귀분석에서 종속변수(Fertility)에 가장 유의한 변수는? *

1점

```
> temp <- lm(Fertility~., data=swiss)
> summary(temp)
```

Call:

```
lm(formula = Fertility ~ ., data = swiss)
```

Residuals:

Min	1Q	Median	3Q	Max
-15.2743	-5.2617	0.5032	4.1198	15.3213

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	66.91518	10.70604	6.250	1.91e-07	***
Agriculture	-0.17211	0.07030	-2.448	0.01873	*
Examination	-0.25801	0.25388	-1.016	0.31546	
Education	-0.87094	0.18303	-4.758	2.43e-05	***
Catholic	0.10412	0.03526	2.953	0.00519	**
Infant.Mortality	1.07705	0.38172	2.822	0.00734	**

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.165 on 41 degrees of freedom

Multiple R-squared: 0.7067, Adjusted R-squared: 0.671

F-statistic: 19.76 on 5 and 41 DF, p-value: 5.594e-10

내 답변



26. 아래 오분류표를 이용하여 특이도 값을 구하시오. *

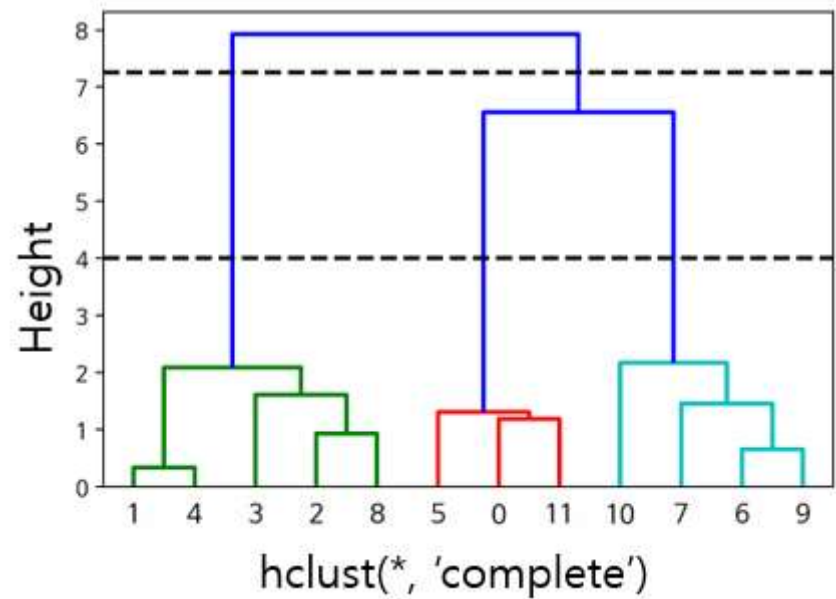
1점

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

내 답변

27. 다음은 덴드로그램의 결과이다. Height가 4일 때 군집의 수는? *

1점



내 답변

28. 로지스틱 회귀분석에서 $\exp(x_1)$ 의 의미는 x_1 이 한 단위 증가할 때마다 성공 * 1점의 ()가 몇 배 증가하는지 나타낸다. 빈 칸에 알맞은 용어는?

내 답변



29. 시계열분석에서 시계열의 수준과 분산에 체계적인 변화가 없고, 주기적 변동이 없다는 것으로 미래는 확률적으로 과거와 동일하다는 것을 의미하는 용어는? * 1점

내 답변

30. 연관규칙 분석기법의 대표적 알고리즘으로 가장 빈번한 항목 집합을 찾기 위한 접근 방식으로, 이해하기 쉽고, 전체 데이터를 스캔한다. 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

