

2023년 36회 기출복원 인쇄용 (49문항)

tbcase73@gmail.com [계정 전환](#)

 비공개



1. 다음 데이터에 대한 설명으로 가장 적절하지 않은 것은?

1점

- ☐ 추론, 예측, 전망, 추정을 위한 근거(basis)로 기능하는 특성을 갖는다.
- ☐ 데이터는 축적된 지식과 아이디어가 결합된 창의적인 산물이다
- ☐ 데이터는 개별 데이터 자체로는 의미가 중요하지 않은 객관적인 사실이다.
- ☐ 다른 객체와의 상호 관계 속에서 가치를 갖는다

2. 빅데이터가 가치 창출 측면에서 기업, 정부, 개인에게 미친 영향으로 옳지 않은 것은?

1점

- ☐ 산업 전체의 생산성이 향상되었다
- ☐ 개인은 맞춤형 서비스를 받거나 적시에 필요한 정보를 얻음으로써 기회비용을 절약하게 되었다.
- ☐ 기업활동의 투명성은 없어지지만 경쟁사보다 강한 경쟁력을 확보하는데 도움이 되었다
- ☐ 비즈니스 모델을 혁신하거나 신사업 발굴에 활용할 수 있게 되었다



3. 사생활 침해 문제를 해결하기 위한 방법으로 가장 적절한 것은 무엇인가?

1점

- ☐ 개인정보 사용자 책임제로 전환
- ☐ 결과기반 책임 원칙 고수
- ☐ 알고리즘 접근 허용
- ☐ 사용자 동의제도 시행

4. 암묵지와 형식지 상호작용의 과정 중 개인의 내재된 경험을 객관적인 데이터로 변환하여 문서나 매체에 저장·가공·분석하는 과정을 무엇이라고 하는가?

1점

- ☐ 표출화
- ☐ 연결화
- ☐ 내재화
- ☐ 공통

5. 다음 비식별화 기법에 대한 설명으로 틀린 것은?

1점

- ☐ 가명처리는 식별할 수 없는 다른 값으로 대체를 의미한다
- ☐ 데이터 마스킹은 개인 정보 식별이 가능한 특정 값을 삭제하는 것이다.
- ☐ 범주화는 단일 식별 정보 해당 그룹의 대푯값으로 변환을 한다.
- ☐ 총계처리는 총합 또는 평균값으로 대체하여 개별 데이터의 값이 보이지 않도록 하는 것이다.



6. 데이터베이스에 대한 설명으로 적절하지 않은 것은?

1점

- ☐ 한 조직의 다수 사용자가 공동으로 이용하고 유지하는 공용데이터이다.
- ☐ DBMS 소프트웨어를 사용하여 데이터베이스를 구축한다.
- ☐ 법률적으로 데이터베이스는 기술을 기반으로 한 일종의 저작물로 인정한다.
- ☐ 데이터베이스내의 모든 데이터는 2차원 테이블로 표현된다.

7. 데이터웨어하우스에 대한 설명으로 가장 적절하지 않은 것은 무엇인가?

1점

- ☐ ETL은 주기적으로 내부 및 외부 데이터베이스로부터 정보를 추출하고 정해진 규약에 따라 정보를 변환한 후에 데이터웨어하우스에 정보를 적재한다
- ☐ 데이터웨어하우스는 전사적 차원보다는 특정 조직의 특정 업무 분야에 초점을 둔 것이다.
- ☐ 데이터웨어하우스에서 관리하는 데이터들은 시간적 흐름에 따라 변화하는 값을 유지한다.
- ☐ 데이터웨어하우스는 기업 내의 의사결정 지원 애플리케이션을 위한 정보를 제공하는 하나의 통합된 데이터 저장 공간을 말한다

8. 다음 데이터 사이언스에 대한 설명으로 가장 부적절한 것은?

1점

- ☐ 데이터 사이언스란 데이터로부터 의미있는 정보를 추출해내는 학문이다
- ☐ 분석 뿐 아니라 이를 효과적으로 구현하고 전달하는 과정까지 포함한 포괄적 개념이다
- ☐ 정형 데이터를 대상으로 통계적 접근법을 사용한다.
- ☐ 과학과 인문학의 교차로에 서 있다고 할 수 있다.



9. 다음 빈 칸에 알맞은 단어는 무엇인가?

1점

()는 거래정보를 하나의 덩어리로 보고 이를 차례로 연결한 거래장부다. 기존 금융회사의 경우 중앙 집중형 서버에 거래 기록을 보관하는 반면, ()는 거래에 참여하는 모든 사용자에게 거래 내역을 보내주며 거래 때마다 이를 대조해 데이터 위조를 막는 방식을 사용한다.

내 답변

10. 다음이 설명하는 것은 무엇인가?

1점

인간의 개입을 최소화하여 인터넷 기반으로 모든 사물을 연결하여 상호 소통하는 지능형 기술로 허기스의 tweet pee, 구글의 google glass, 나이키의 fuel band, 삼성의 갤럭시 워치 등을 예로 들 수 있다.

내 답변

2-1. 다음 중 기업의 데이터 분석 도입의 수준 진단의 대상으로 가장 적절하지 않은 것은? 1점

- ☐ 분석 업무 파악
- ☐ 분석 기법
- ☐ 분석 인력 및 조직
- ☐ 분석 성과 평가



2-2. 분석 성숙도 모델 구성에서 고려하는 분석 성숙도 진단 부문으로 적절하지 않은 것은? 1점

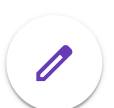
- ☐ 비즈니스 부문
- ☐ 기업 문화 부문
- ☐ 조직의 역량 부문
- ☐ IT 부문

2-3. 다음 분석과제의 특징 중 Accuracy와 Precision에 대한 설명으로 틀린 것은? 1점

- ☐ 분석의 활용적인 측면에서는 Precision이 중요하며, 안정적인 측면에서는 Accuracy가 중요하다.
- ☐ Accuracy와 Precision의 관계는 트레이드 오프가 되는 경우가 많다.
- ☐ Accuracy는 모델과 실제 값의 차이에 대한 것이다.
- ☐ Precision은 모델을 반복했을 때의 편차를 의미한다.

2-4. 다음 중 분석 대상은 명확하지만 분석 방식이 명확하지 않은 경우 수행하는 분석 주제의 유형은 무엇인가? 1점

- ☐ 솔루션(Solution)
- ☐ 통찰(Insight)
- ☐ 최적화(Optimization)
- ☐ 발견(Discovery)



2-5. 분석 과제 정의서에 대한 설명으로 가장 적절한 것은 무엇인가?

1점

- ☐ 프로젝트를 수행 계획 수립 단계에서 전체 업무를 분류하여 구성 요소로 만든 후 각 요소를 평가하고 일정별로 계획하며 그것을 완수할 수 있는 사람에게 할당해주는 역할을 한다.
- ☐ 분석 모델에 적용될 알고리즘과 분석모델의 기반이 되는 Feature가 포함되어야 한다.
- ☐ 이해관계자가 프로젝트의 방향을 설정하고, 성공 여부를 판별할 수 없는 자료이다.
- ☐ 필요한 소스 데이터, 분석 방법, 데이터 입수 난이도, 분석 과정 상세 등의 항목이 포함되어야 한다.

2-6. 분석 마스터 플랜을 수립할 때 적용 범위 및 방식에 대한 고려요소가 아닌 것은 무엇인가?

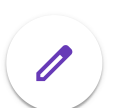
1점

- ☐ 투입 비용 수준
- ☐ 분석 데이터 적용 수준
- ☐ 업무 내재화 적용 수준
- ☐ 기술 적용 수준

2-7. 분석 마스터플랜의 세부 이행계획 수립 시 고려해야 할 데이터 분석 체계에 대한 설명으로 적절한 것은?

1점

- ☐ 분석 마스터플랜의 모든 단계를 반복한다.
- ☐ 프로젝트의 세부 일정계획도 데이터 분석체계를 고려하여 작성한다.
- ☐ 순차적인 정련 과정을 통해 프로젝트의 기간을 단축하는 방식을 주로 사용한다.
- ☐ 데이터 수집 및 확보와 분석 데이터를 준비하는 단계를 반복적으로 진행한다.



2-8. 분석 기획에 대한 설명으로 적절하지 않은 것은 무엇인가?

1점

- ☐ 해당 문제 영역에 대한 전문성 역량 및 통계학적 지식을 활용한 분석 역량과 분석 도구인 데이터 및 프로그래밍 기술 역량에 대한 균형 잡힌 시각을 가지고 방향성 및 계획을 수립해야 한다
- ☐ 성공적인 분석을 하기 전 중요 사전 작업이다.
- ☐ 상향식 분석은 분석 기획 전 탐색적 데이터 분석 수행을 한다
- ☐ 실제 분석을 수행에 앞서 분석을 수행할 과제의 정의 및 의도했던 결과를 도출할 수 있도록 이를 적절하게 관리할 수 있는 방안을 사전에 계획하는 일련의 작업이다.

2-9. 빅데이터의 4V는 빅데이터의 3V에 무엇이 추가된 것인가?

1점

내 답변

2-10. 다음 빈칸에 공통으로 들어갈 단어로 알맞은 것은 무엇인가?

1점

분석 과제 우선순위 평가 기준에는 ()과 난이도가 있으며 ()의 경우 전략적 중요도와 목표가치, 난이도의 경우 데이터 획득/저장/가공 비용, 분석 적용 비용, 분석 수준에 따라 판단하게 된다.

내 답변

3-1. 다음 중 군집분석 기법으로 적절하지 않은 것은 무엇인가?

1점

- ☐ PAM
- ☐ DBSCAN
- ☐ 실루엣 지수(Silhouette Coefficient)
- ☐ 퍼지(Fuzzy) Clustering



3-2. Wage 데이터셋에 대한 anova 분석 결과 해석의 내용 중 틀린 것은?

1점

```
> aov <- aov(wage ~ age, data=data)
> summary(aov)
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|------|---------|---------|---------|------------|
| age | 1 | 199870 | 199870 | 119.3 | <2e-16 *** |
| Residuals | 2998 | 5022216 | 1675 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- ☐ age와 wage에 대한 선형관계를 알 수 없다.
- ☐ age와 wage는 양의 상관관계이다
- ☐ 관측치는 3000 개 이다
- ☐ MSE는 1675 이다

3-3. 웹 데이터의 수집을 위해 웹페이지의 구조를 분석하여 데이터를 자동으로 수집하는 방법을 무엇이라고 하는가?

1점

- ☐ FTP
- ☐ 웹 크롤링(Web Crawling)
- ☐ Streaming)
- ☐ Open API

3-4. 신경망 모형에서 출력값이 여러 개이고 목표치가 다범주인 경우에 사용하는 것으로 각 범주에 속할 사후 확률을(posterior probability) 제공하는 활성화 함수는 무엇인가?

1점

- ☐ 항등 함수
- ☐ ReLU
- ☐ sigmoid
- ☐ softmax



3-5. 시그모이드 함수의 범위로 알맞은 것은?

1점

- ☐ 0 ~ 1
- ☐ -1 ~ 1
- ☐ -1 ~ 0
- ☐ 0.5 ~ 1

3-6. 다음의 수식에 해당하는 데이터 간의 거리 계산 방식은 무엇인가?

1점

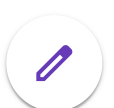
$$\sum_{i=1}^n |x_i - y_i|$$

- ☐ 유클리드 거리
- ☐ 맨해튼 거리
- ☐ 민코프스키 거리
- ☐ 마할라노비스 거리

3-7. 연관분석에 대한 특징 틀린 것은?

1점

- ☐ 조건반응(if then)으로 표현되는 연관분석의 결과를 이해하기 쉽다.
- ☐ 비목적성 분석 기법이다.
- ☐ 대표적인 알고리즘으로 Aprior가 있다.
- ☐ 분석을 위한 계산이 복잡하다는 단점이 있다.



3-8. 다음 중 표본들이 서로 관련되어 있는 경우, 짝지어진 두 개의 관찰치들의 크고 작음을 +와 -로 표시하여 그 개수를 가지고 두 그룹의 분포 차이가 있는가에 대한 가설을 검증하는 방법은 무엇인가? 1점

- ☐ Sign Test
- ☐ Chi-Square Test
- ☐ ANOVA Test
- ☐ 스피어만 상관계

3-9. 자료의 척도에 대한 설명으로 적절하지 않은 것은? 1점

- ☐ 비율척도 - 사칙연산이 모두 가능하고, 혈액형, 학력 등이 해당된다.
- ☐ 구간척도 - 덧셈, 뺄셈이 가능하고 절대 0점을 포함하지 않는 온도가 이에 해당된다.
- ☐ 서열척도 - 연산이 불가능하고 메달과 같이 범주간 순서가 있는 것이 이에 해당된다.
- ☐ 명목척도 - 단순히 측정 대상의 특성을 분류하거나 확인하기 위한 목적으로 사용된다.

3-10. 소득순위처럼 정규분포가 아닌 오른쪽 꼬리가 긴 분포(Positive skewed)에서 평균과 중앙값의 관계로 알맞은 것은? 1점

- ☐ 중앙값이 평균보다 크다
- ☐ 평균이 중앙값보다 크다
- ☐ 평균과 중앙값의 관계에 변화가 없다
- ☐ 평균은 중앙값의 제곱과 같다



3-11. R에서 숫자형, 문자형, 논리형 벡터를 하나로 합친 벡터를 구성하는 경우 합쳐진 벡터의 형식은 무엇인가?

1점

- ☐ 숫자형 벡터
- ☐ 문자형 벡터
- ☐ 논리형 벡터
- ☐ 데이터프레임

3-12. 다음 중 빅데이터 분석 프로세스에서 모델링 단계에 해당하지 않는 항목은 무엇인가?

1점

- ☐ 데이터 분할
- ☐ 데이터 모델링
- ☐ 모델 적용 및 운영 방안
- ☐ 수행방안 설계

3-13. 다음 중 모형 성과 평가 방법으로 적절하지 않은 것은?

1점

- ☐ 결정계수
- ☐ 실루엣 지수
- ☐ 엔트로피(Entropy)
- ☐ ROC 그래프



3-14. 다음 중 분류 모형 평가에 활용하지 않는 것은 무엇인가?

1점

- ☐ 텐드로그램
- ☐ 오분류표
- ☐ ROC 그래프
- ☐ Kappa 지수

3-15. 다차원척도법에 대한 설명으로 가장 적절하지 않은 것은 무엇인가?

1점

- ☐ 개체들의 거리는 유클리드(Euclidean) 거리와 유사도를 이용하여 구한다.
- ☐ 관측 대상의 상대적 거리의 정확도를 높이기 위해 적합 정도를 스트레스 값(Stress Value)로 나타낸다.
- ☐ 스트레스 값은 0에 가까울 수록 적합도가 좋음을 나타낸다.
- ☐ 개체들 사이의 유사성과 비유사성을 측정하여 차원을 축소하기 위해 사용한다.

3-16. 앙상블 모형의 특징으로 옳바르지 않은 것은?

1점

- ☐ 성능을 분산시키기 때문에 과대적합(overfitting) 감소 효과가 있다.
- ☐ 각 모형의 상호연관성이 높을수록 정확도 또한 높아진다.
- ☐ 여러 개의 모형의 결과를 종합하여 정확도를 높이는 방법이다.
- ☐ Bagging, Boosting 등 다양한 방법의 앙상블 기법이 존재한다.



3-17. 이상치 관련한 설명으로 가장 옳지 않은 것은?

1점

- ☐ DBSCAN 군집을 실행해 군집에 포함되지 않은 것을 이상치로 한다.
- ☐ ESD 방법에서는 평균 - 3*표준편차보다 작거나, 평균 + 3*표준편차보다 큰 데이터를 이상치로 규정한다.
- ☐ 기하평균을 이용하는 경우 기하평균 - 2.5*표준편차 보다 작거나, 기하평균 + 2.5*표준편차보다 큰 데이터를 이상치로 규정한다.
- ☐ IQR을 사용하는 방식의 경우 $Q2(\text{중위수}) + 1.5 \cdot IQR$ 보다 크거나 $Q2(\text{중위수}) - 1.5 \cdot IQR$ 작은 데이터를 이상치로 규정한다.

3-18. 군집분석에 대한 설명으로 적절하지 않은 것은?

1점

- ☐ 집단별 특성이 유사할 경우 안정성이 높다
- ☐ 유사성을 이용하여 몇 개의 집단으로 그룹화하는 분석이다
- ☐ 군집분석은 집단 간 이질성과 집단 내 동질성이 모두 낮아지는 방향으로 군집을 만든다
- ☐ 비계층적 군집분석 기법의 경우 사용자가 사전 지식 없이 그룹의 수를 정해주는 일이 많기 때문에 결과가 잘 나오지 않을 수 있다.

3-19. 의사결정나무의 특징으로 알맞지 않은 것은?

1점

- ☐ 상관성이 높은 변수가 있어도 크게 영향을 받지 않는다.
- ☐ 비정상적인 잡음 데이터에 대해서는 민감하게 분류한다.
- ☐ 목적 변수가 이산형(범주형)인 경우와 연속형인 경우 모두 사용할 수 있다.
- ☐ 설명력이 좋으며, 과대적합에 취약한 특징이 있다.



3-20. 데이터마이닝을 위한 데이터 분할과 관련된 설명 중 알맞지 않은 것은?

1점

- ☐ 데이터는 학습용, 검증용, 평가용 데이터로 분할하여 사용할 수 있다.
- ☐ 검증용 데이터(validation data)는 학습과정에서 사용되지 않는다.
- ☐ 검증용 데이터는 훈련에 사용되지 않는다.
- ☐ 데이터 수가 적을 때는 교차 검증을 사용한다

3-21. 확률적 표본추출 방법이 아닌 것은?

1점

- ☐ 단순 무작위 추출법
- ☐ 계통 추출법
- ☐ 집단 추출법
- ☐ 층화 추출법



3-22. 코드 실행 결과에 대한 설명으로 적절한 것은?

1점

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

- ☐ 회귀모형은 유의수준 5%하에서 통계적으로 유의미하다.
- ☐ 모든 회귀계수들이 유의미한다.
- ☐ 설명력은 7.165 이다
- ☐ 데이터의 개수는 41개 이다

3-23. 데이터 전처리 과정에 대한 설명으로 올바른 것은 무엇인가?

1점

- ☐ 결측치는 연산에 아무런 방해가 되지 않으므로 그대로 두어도 무방하다.
- ☐ 데이터 특성을 파악하고 통찰을 얻기 위한 방법을 데이터 EDA라고 한다.
- ☐ 모든 분석의 이상치는 시간이 오래 걸리더라도 모두 찾아내어 제거한다
- ☐ 데이터 변환을 통해 정규분포 형태의 데이터로 만들면 데이터가 왜곡되어 올바른 학습이 되지 않는다.

3-24. 변수 가공에 대한 설명으로 적절하지 않은 것은?

1점

- ☐ 구간화의 개수가 감소하면 정확도는 높아지지만 속도가 느려진다.
- ☐ log, sqrt를 취하면 큰 값을 작게 만들 수 있다 - 오른쪽 꼬리 긴 분포에 사용
- ☐ 제곱, exp를 취하면 작은 값을 크게 만들 수 있다 - 왼쪽 꼬리 긴 분포에 사용
- ☐ MinMax Normalization을 하면 값이 0 ~ 1 사이의 범위로 변경된다

3-25. 다음이 설명하는 시계열 모형은 무엇인가?

1점

“자기자신의 과거자료로 설명하는 모형으로 백색잡음의 현재값과 자기자신의 과거값의 가중합으로 선형성을 표현하는 정상시계열 모형이다.”

내 답변

3-26. 다음이 설명하는 앙상블 모형의 종류는 무엇인가?

1점

“배깅(bagging)에 랜덤 과정을 추가한 방법으로 노드 내 데이터를 자식 노드로 나누는 기준을 정할 때 모든 예측 변수에서 최적의 분할을 선택하는 대신, 설명변수의 일부분만을 고려함으로 성능을 높이는 방법을 사용한다.”

내 답변

3-27. 다음 오분류표를 사용하여 F1-score 를 구하시오.

1점

| | | 예측 | | |
|----|-------|------|-------|-----|
| | | TRUE | FALSE | 합계 |
| 실제 | TRUE | 30 | 70 | 100 |
| | FALSE | 60 | 40 | 100 |
| | 합계 | 90 | 110 | 200 |

내 답변



3-28. 이산확률변수 X 가 가능한 값으로 1, 2, 4가 있다.
 $P(X=1) = 0.3$ 이고 기댓값이 2.7 일 때 $P(X=2)$ 는 무엇인가?

1점

내 답변

3-29. 우등반에 들어가기 위해서는 어느 시험에서 상위 2% 안에 들어야 한다. 해당 시험 점수의 평균이 85점이고 표준편차가 5일 때, 우등반에 들어가기 위한 최소 시험 점수는? (단, $P(Z \leq 2.05) = 0.98$)

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



