

2023-1학기 데이터사이언스입문(01) 중간시험 - 중간 고사

1. 데이터 시각화 요소 중 정확한 순서대로 나열한 것은?
(1점)

- 1) 길이 - 위치 - 각도 - 면적 - 채도
- 2) 각도 - 방향 - 길이 - 면적 - 부피
- 3) 위치 - 길이 - 형태 - 각도 - 색조
- 4) 방향 - 길이 - 면적 - 부피 - 채도
- 5) 위치 - 면적 - 부피 - 채도 - 색조

정답 : 5

해설 :

2. 확률표본추출방법에 대한 설명 중 틀린 것은?
(1점)

- 1) 표본추출 방법은 무작위적 표본추출을 사용한다.
- 2) 표본오차의 추정이 불가능하다.
- 3) 표본추출하는 데 시간이 많이 소요된다.
- 4) 주로 연구대상이 표본으로 추출될 확률이 알려져 있을 때 사용한다.
- 5) 모수추정에 편의가 없다.

정답 : 2

해설 :

3. 데이터사이언스의 학습과 추론과 관련된 내용 중 틀린 것은?
(1점)

- 1) 데이터 사이언스의 추론은 귀납법을 따른다.
- 2) 데이터 사이언스의 학습 방법은 지도 학습, 비지도 학습, 준 지도학습 방법이 있다.
- 3) 입력 데이터가 연속적일 때, regression문제를 지도학습 (supervised learning) 방법으로 학습 가능하다.
- 4) 준지도학습 방법은 지도학습 보다 학습의 난이도가 낮으며, clustering 등의 방법이 있다.
- 5) 데이터사이언스는 다양한 관찰된 데이터를 바탕으로 여기에 내포되어 있는 일반적인 원리를 추론하는 방법을 사용한다.

정답 : 4

해설 :

4. 컴퓨팅 사고에서 중요한 특징 중 하나로 실제 세계의 문제를 해결 가능한 형태로 표현하기 위한 사고 과정을 의미하는 것은?
(1점)

- 1) 추상화
- 2) 자동화
- 3) 이진화
- 4) 일반화
- 5) 패턴인식

정답 : 1

해설 :

5. 데이터에 대한 설명 중 틀린 것은?
(1점)

- 1) 구체적인 데이터를 활용한 주장은 정도의 차이를 명확하게 인식하게 한다.
- 2) 데이터를 활용한 구체적인 표현은 상대방의 인식 변화 및 강화에 효과적이다.
- 3) 정형 데이터가 비정형 데이터보다 더 효과적이며, 의미있는 것이다.
- 4) 우리 일상생활에서 사용하는 감정, 분류, 사실, 소통의 메시지도 데이터의 일부이다.
- 5) 성능 좋은 모델도 중요하지만 더 중요한 것은 양질의 많은 데이터 수집이 더 중요

정답 : 3

해설 :

6. 데이터 시각화가 필요한 이유에 대해 잘 못 설명한 것은?
(1점)

- 1) 많은 양의 데이터를 한 눈에 볼 수 있다.
- 2) 데이터 분석에 대한 전문 지식이 없어도 누구나 쉽게 데이터에서 인사이트를 찾을 수 있다.
- 3) 효과적인 데이터 인사이트 공유로 데이터 기반의 의사결정을 할 수 있다.
- 4) 데이터 시각화는 기술적으로 데이터를 다루는 전문 역량이 없으면 할 수 없다.
- 5) 데이터를 활용한 인사이트 발굴, 스토리텔링에 도움이 된다.

정답 : 4

해설 :

7. 데이터베이스의 설명 중 틀린 것은?
(1점)

- 1) 데이터를 표현하는 표 (Table)에서 열의 개수는 데이터의 수와 같다.
- 2) 데이터베이스 시스템의 주요 기능으로는 정의 기능, 조작 기능, 제어 기능이 있다.
- 3) 데이터베이스 시스템의 단점으로 비용이 많이 든다는 점이 있다.
- 4) 데이터베이스 모델링에 정답은 없지만 모범 답안 형태의 모델링은 존재한다.
- 5) 현실에서 쓰이는 것을 테이블로 변경하는 작업을 데이터베이스 모델링이라고 한다.

정답 : 1

해설 :

8. 표본 조사 방법 중 표본의 크기가 크지 않아도 대표성이 보장되지만 모집단의 각 층에 대한 정보가 필요한 방법은 무엇인가?
(1점)

- 1) 군집 추출법
- 2) 판단 표본 추출법
- 3) 단순 임의 추출법
- 4) 층화 추출법

5) 계통 추출법

정답 : 4

해설 :

9. 데이터마이닝의 특징 중 틀린 것은?
(1점)

- 1) 관측자료는 시간의 흐름에 따라 비계획적으로 축적되며, 자료분석을 염두에 두고 수집된다.
- 2) 대용량의 관측 가능한 자료를 다룬다.
- 3) 일반화에 초점을 두고 있다.
- 4) 경쟁력 확보를 위한 의사결정을 지원하기 위해 활용된다.
- 5) 수리적으로 밝혀지지 않는 경험적 방법에 근거하고 있다.

정답 : 1

해설 :

10. 게슈탈트 이론 중 "적절한 양의 정보가 제공될 때 뇌는 정보의 빈틈을 메꿔 하나의 원전체를 생성"을 의미하는 것은?
(1점)

- 1) 폐합
- 2) 출현
- 3) 구상화
- 4) 대칭
- 5) 다중 안정성

정답 : 1

해설 :

11. 데이터사이언스 프로세스 중 서비스 과정에서 다양한 피드백을 바탕으로 이전 단계를 다시 반복해서 수행할 지 여부를 결정짓는 단계는?
(1점)

- 1) Deployment
- 2) Modeling
- 3) Evaluation
- 4) Data Understanding
- 5) Business Understanding

정답 : 1

해설 :

12. 데이터마이닝의 연관 규칙에 대한 설명 중 틀린 것은?
(1점)

- 1) 목적 변수가 없다.
- 2) 규칙들 간에는 서로 영향을 주지 않기 때문에 하나의 고객이 여러 개의 규칙에 해당 될 수 있다.
- 3) 조건반응으로 표현 가능하다.

4) 비지도 예측의 가장 대표적인 것이 연관 규칙 분석이다.

5) 목표변수, 결과값을 모르는 상태에서 데이터간의 관계를 분석해서 결과값을 도출하는 것을 지도예측이라고 한다.

정답 : 5

해설 :

13. 자료 정리와 관련된 통계학의 기본 내용 중 틀린 것은?

(1점)

1) 데이터가 취하는 값의 범위를 몇 개의 구간으로 나누어 데이터가 해당 구간에 몇 개씩 들어가는지 정리한 표가 도수분포표이다.

2) 도수분포그래프는 도수분포표를 회전후 막대 그래프를 함께 표현하여 자료의 특징을 더 확실히 파악 가능하다.

3) 상대도수분포다각형의 Y값은 -1과 1사이의 값을 가진다.

4) 연속 변량의 경우 도수분포곡선에서 구간폭을 줄이면 히스토그램은 곡선이 된다.

5) 상대 도수는 각 도수를 총 도수로 나눈 실숫값이다.

정답 : 3

해설 :

14. 데이터마이닝의 분류 규칙에 대한 설명 중 옳은 것은?

(1점)

1) F1 score는 데이터 label이 불균형 구조일 때, 모델의 성능을 정확하게 평가할 수 있다.

2) precision값은 높을 수록 recall의 값은 낮을 수록 좋다.

3) 의사결정나무에서 과적합 문제는 고려할 필요가 없다.

4) 의사결정나무의 뿌리 마디는 하나로 고정되어 있다.

5) 의사결정나무를 만들 때 분리기준과 정지규칙을 생략해도 좋다.

정답 : 1

해설 :

15. 관계형 DBMS의 설명 중 틀린 것은?

(1점)

1) 모든 데이터는 테이블에 저장한다.

2) 대용량 데이터를 체계적으로 관리할 수 있다.

3) 매우 복잡한 내부 포인터를 사용한다.

4) RDBMS라고도 한다.

5) 시스템 자원을 많이 차지하여 시스템이 전반적으로 느려지는 단점이 있다.

정답 : 3

해설 :

16. 데이터베이스의 특징이 아닌 것은?

(1점)

1) 지속적인 변화

2) 동시 공유

- 3) 위치에 따른 참조
- 4) 실시간 접근성
- 5) 무결성 유지

정답 : 3

해설 :

17. 활동 과정에서 수집, 처리 저장되었지만 사용 목적이 명확하지 않은 데이터를 의미하는 용어는?
(1점)

- 1) 빅데이터
- 2) 스몰데이터
- 3) 패스트데이터
- 4) 스마트데이터
- 5) 다크데이터

정답 : 5

해설 :

18. 파일 시스템과 DBMS의 특징 중 옳은 것은?
(1점)

- 1) 파일 시스템은 별도의 소프트웨어 설치가 필요 없다.
- 2) DBMS는 프로그램 개발 생산성이 나쁘다.
- 3) DBMS는 데이터의 중복 저장으로 일관성이 결여된다.
- 4) 파일 시스템은 데이터 정의와 프로그램의 독립성이 유지 가능하다.
- 5) 파일 시스템은 데이터의 무결성을 유지할 수 있으며, 데이터 표준 준수에 용이하다.

정답 : 1

해설 :

19. 가설 검정의 오류에 대한 설명 중 옳은 것은?
(1점)

- 1) 귀무 가설이 맞음에도 불구하고 귀무 가설을 기각하는 오류를 α error라고 한다.
- 2) H_0 가 옳지 않은데도 H_0 를 채택하게 되는 오류를 제 1종 오류라고 한다.
- 3) 현재 표본으로부터 얻어진 정보로는 대립가설이 사실임을 입증할 만한 충분한 증거가 없음을 의미하는 것이 제 2종 오류이다.
- 4) 일반적으로 유의수준 α 를 미리 정해 놓고 β 를 최대화하는 검정의 방법을 사용한다.
- 5) 입증하여 주장하고자 하는 가설을 귀무가설이라 한다.

정답 : 1

해설 :

20. 데이터사이언스 프로세스 중 Data Understanding 과정 에서 데이터의 질 (Quality)이 중요하다는 용어로 데이터사이언스, AI 분야에서 많이 사용되는 문구는 무엇인가?
(1점)

- 1) Garbage-in Garbage-out

- 2) Quality of Experience
- 3) On Premise and Private Cloud
- 4) Transformation and Cleaning
- 5) Feature Engineering

정답 : 1

해설 :

21. 빅데이터의 4V 특성에 포함되는 것이 아닌 것은?
(1점)

- 1) Value
- 2) Volume
- 3) Velocity
- 4) Variety
- 5) Veracity

정답 : 1

해설 :

22. 표본조사방법 비확률추출 방법은?
(1점)

- 1) 할당 표본 추출법
- 2) 단순 임의 추출법
- 3) 계통 추출법
- 4) 층화 추출법
- 5) 군집 추출법

정답 : 1

해설 :

23. 데이터 마이닝과 전통적인 통계의 비교 중 전통적인 통계에 대한 설명은?
(1점)

- 1) 현실적인 noisy한 데이터에 대한 가정이 없다.
- 2) 알고리즘이 비선형성에 기반을 두고 있다.
- 3) 제안된 가설에 대한 검증이 주 목적이다.
- 4) 모수 추정이 주 목적이 아니다.
- 5) 궁극적인 목적은 예측에 중점을 두고 있다.

정답 : 3

해설 :

24. DBMS의 특징 중 틀린 것은?
(1점)

- 1) 데이터가 추가, 수정, 삭제할 때마다 정렬과 인덱스 업데이트가 수행된다.
- 2) CRUD는 Create, Recover, Update, Delete의 약자이다.
- 3) DBMS는 쿼정렬 또는 힙정렬을 이용해 검색을 빠르게 한다.
- 4) DBMS 가 설치되어 데이터를 가진 쪽을 서버, 외부에서 데이터 요청하는 쪽을 클라이언트라 한다.
- 5) DBMS는 이진 검색 또는 B-Tree 검색 등을 이용하여 데이터가 추가, 수정, 삭제할 때마다 인덱스를 업데이트한다.

정답 : 2

해설 :

25. Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Data Science (DS) 용어 중 더 큰 범위 순으로 정렬한 것은?
(가장 넓은 범위의 의미를 가진 용어가 먼저 기술)
(1점)

- 1) ML - AI -DL - DS
- 2) DS - AI - ML - DL
- 3) AI - DS - ML -DL
- 4) DS - ML - AI - DL
- 5) AI - DS -DL - ML

정답 : 2

해설 :

26. 컴퓨팅 사고의 4가지 구성요소는?
(1점)

- 1) 분해, 추론, 패턴인식, 학습
- 2) 알고리즘, 추상화, 데이터 수집, 평가
- 3) 분해, 패턴인식, 알고리즘, 추상화
- 4) 2진수, 캐시, 버퍼, 병렬처리
- 5) 분해, 머신러닝, 자동화, 추상화

정답 : 3

해설 :

27. 추론 통계 (inferential statistics)에 대한 설명 중 틀린 것은?
(1점)

- 1) 수집한 데이터를 바탕으로 추론 예측하는 통계 기법이다.
- 2) 추리통계 , 추측통계 라고도 불리며 , 기술통계가 선행되어야 한다.
- 3) 실제 모집단의 통계적 특성과 차이가 있을 수 있으며 , 이를 표집오차 라고 한다.
- 4) 수집한 데이터를 요약 , 묘사 , 설명하는 통계 기법이다.
- 5) 추론 통계는 추정과 가설검정으로 나눌 수 있다.

정답 : 4

해설 :

28. 지도학습과 비지도학습에 대한 설명 중 옳은 것은?
(1점)

- 1) 군집분석은 지도 학습 방법 중 하나 이다.
- 2) 지도학습 중 범주형 데이터는 regression, 연속형 데이터는 classification으로 예측 가능하다.
- 3) 비지도학습은 비교적 정확한 학습이 가능한 장점이 있지만 사용 가능한 데이터의 한계/부족한 단점이 있다.
- 4) 일반적으로 비지도 학습은 지도학습 대비 학습의 난이도가 높다.
- 5) 연관규칙은 지도 학습 방법 중 하나 이다.

정답 : 4

해설 :

29. 비즈니스 인텔리전스에 대한 설명 중 틀린 것은?
(1점)

- 1) 최종 사용자 질의 및 보고를 포괄하는 의미이다.
- 2) 경영진과 경영 분석가들이 데이터를 통해 합리적 의사결정을 내릴 수 있도록 데이터를 수집, 저장, 처리, 분석하는 일련의 기술, 응용시스템을 의미한다.
- 3) 데이터웨어하우스의 의미를 포함한다.
- 4) 데이터 질의 및 보고도구의 요소를 포함한다.
- 5) 비즈니스 인텔리전스는 데이터 마이닝의 일부로 볼 수 있다.

정답 : 5

해설 :

30. 컴퓨팅 사고의 사례 중 앞으로 사용이 예상되는 것을 미리 가져다 놓은 것의 의미를 가지는 개념은?
(1점)

- 1) 버퍼
- 2) 캐시
- 3) 병렬처리
- 4) 추상화
- 5) 패턴인식

정답 : 2

해설 :

31. 다음 중 빅데이터가 만들어 내는 변화와 가장 거리가 먼 것은?
(1점)

- 1) 가치가 있을 것이라고 예상되는 특정한 정보만 모아서 처리하는 것이 아니라 최대한 많은 데이터를 모으고 그 데이터를 다양한 방식으로 조합해 숨은 정보를 찾아내는 방식이 중요해진다.
- 2) 데이터의 규모가 증가함에 따라 사소한 몇 개의 오류 데이터는 분석결과에 영향을 미치지 않기 때문에 데이터셋에 포함하여 분석해도 상관없는 경우가 많아진다.
- 3) 인과관계의 규명 없이 상관관계 분석결과만으로도 인사이트를 얻고 이를 바탕으로 수익을 창출할 수 있는 기회가 점차 늘어나고 있다.
- 4) 데이터의 양이 증가하고 유형이 복잡해짐에 따라 수많은 데이터 중에서 분석에 필요한 데이터를 선정하기 위한 정교한 표본조사 기법의 중요성이 대두되고 있다.

5) 최근 빅데이터의 중요성과 더불어 스몰데이터의 중요성도 함께 강조되고 있다.

정답 : 4

해설 : 데이터의 양이 증가하고 유형이 복잡해짐에 따라 수많은 데이터 중에서 분석에 필요한 데이터를 선정하기 위한 정교한 표본조사 기법의 중요성이 대두되고 있다. 내용은 빅데이터와는 거리가 멀다.

32. 데이터마이닝에서 사용하는 데이터의 설명 중 옳은 것은?
(1점)

- 1) 데이터에 대한 데이터를 의미하는 용어가 데이터 마트 이다.
- 2) 정보분석을 위해 구축된 것 /수집된 데이터를 요약, 가공하여 저장하는 것을 운영계 데이터베이스라 한다.
- 3) 메타 데이터는 최근의 데이터를 저장하며 대량의 데이터 저장에 용이하다.
- 4) 하나의 데이터마이닝 주제 또는 고객분석을 위해 통합된 데이터로 구성된 보조적인 데이터 저장소를 데이터마트라고 한다.
- 5) 데이터웨어하우스는 특정한 목적의 사용자 그룹을 위해 특정 주제영역의 데이터로 구성된다.

정답 : 4

해설 :

33. 상관 계수의 설명 중 틀린 것은?
(1점)

- 1) 상관계수는 계산식은 복잡하나 1 ~ 1 사이의 값을 가진다.
- 2) 상관계수가 0 일 때 두 변수 사이에 관계가 없다는 것을 의미한다.
- 3) 상관계수의 +, -는 방향을 의미이다.
- 4) 상관계수가 절대값 1 에 가까울 수록 힘이 세다는 의미이다.
- 5) 데이터들이 퍼져 있으면 상관계수가 -1에 가까워 진다.

정답 : 5

해설 :