

데이터의 이해

15회 기출

01. 데이터는 그 형태에 따라 정성 데이터와 정량 데이터로 구분된다. 다음 중 정성 데이터에 속하는 것은?

- ① 풍향 ② 습도 ③ 기상특보 ④ 1시간 강수량

15회 기출

02. 다음 중 암묵지와 형식지의 상호작용과 가장 관련이 없는 것은?

- ① 공통화 ② 내면화 ③ 연결화 ④ 추상화

15회 기출

03. SQL은 다양한 집계함수를 제공하는데, 다음 집계함수 중 어떠한 데이터의 타입에도 사용이 가능한 것은?

- ① AVG ② COUNT ③ SUM ④ STDDEV

14회 기출

04. 다음 중 개인정보 비식별화 기법을 설명한 것으로 가장 부적절한 것은?

- ① 총계처리 - 데이터의 총합 값을 보임으로써 개별 데이터의 값을 보이지 않도록 함
 ② 데이터 마스킹 - 개인 식별에 중요한 데이터 값을 삭제
 ③ 가명처리 - 개인 식별에 중요한 데이터를 식별 할 수 없는 다른 값으로 변경
 ④ 범주화 - 데이터의 값을 범주의 값으로 변환하여 값을 감춤

12회 기출

05. 다음 중 데이터에 대한 설명으로 가장 적절하지 않는 것은 무엇인가?

- ① 양질의 데이터를 확보하지 못하면 잘못된 분석 결과를 얻음
 ② 창의적인 데이터 매시업(Mashup)은 기존에 풀기 어려웠던 문제 해결에 도움
 ③ 비정형 데이터는 데이터 내부에 메타 데이터를 갖고 있으며 일반적으로 파일 형태로 저장
 ④ 공공부문에서 개방하고 있는 대표적인 데이터는 교통 데이터, 물가 데이터, 의료 데이터이다.

06. 다음은 지식에 대한 설명이다. 옳지 않은 것은?

- ① 지식은 크게 암묵지와 형식지로 구분할 수 있으며 서로 상호작용에 있어 중요한 역할을 한다.
- ② 암묵지는 ‘김치담그기’, ‘자전거타기’와 같이 학습적 체험을 통해 개인에게 습득되어 있지만 겉으로는 드러나지 않는 지식을 의미한다.
- ③ 형식지는 교과서, 매뉴얼, 비디오, DB와 같이 형상화된 지식을 의미하는데, 유형의 대상이 있기 때문에 지식의 전달과 공유가 매우 쉽다.
- ④ 암묵지는 사회적으로 중요하지만 다른 사람에게 공유되기 어려우므로 개인에게 축적된 내면화된 지식을 조직의 지식으로 공통화하는 것이 중요하다.

13회 기출

07. 다음 중 그 자체로는 의미가 중요하지 않은 객관적인 사실인 데이터를 가공, 처리하여 얻을 수 없는 것은?

- ① 지혜(Wisdom) ② 정보(Information) ③ 기호(Sign) ④ 지식(Knowledge)

15회 기출

08. 다음 중 지식(Knowledge)에 대한 예시로 가장 적절한 것은?

- ① A사이트보다 B사이트가 다른 물건도 비싸게 팔 것이다.
- ② B사이트보다 가격이 상대적으로 저렴한 A사이트에서 USB를 사야겠다.
- ③ A사이트는 10,000원에, B사이트는 15,000원에 USB를 팔고 있다.
- ④ B사이트의 USB 판매가격이 A사이트보다 더 비싸다.

14회 기출

09. 다음 중 글로벌 기업의 빅데이터 활용사례로 그 연결이 부적절한 것은?

- ① 구글 - 실시간 자동 번역시스템을 통한 의사소통의 불편해소
- ② 라쿠텐 - 이용자의 콘텐츠 기호를 파악하여 새로운 영화를 추천해주는 Cinematch 시스템 운영
- ③ 월마트 - 소셜 미디어를 통해 고객 소비 패턴을 분석하는 월마트랩(Wallmart Labs) 운영
- ④ 자라 - 일일 판매량을 실시간 데이터 분석으로 상품 수요를 예측

10. 다음은 데이터베이스의 특징에 관한 설명이다. 옳지 않은 것은?

- ① 데이터베이스는 동일한 내용의 데이터가 중복되어 있지 않다는 것을 의미하는 통합된 데이터이며 데이터 중복은 관리상의 복잡한 부작용을 초래할 수 있다.
- ② 데이터베이스는 자기 디스크나 자기 테이프 등과 같이 컴퓨터가 접근할 수 있는 저장 매체에 저장되는 것을 의미하므로 저장된 데이터이며 기본적으로 컴퓨터 기술을 바탕으로 한 것이다.
- ③ 데이터베이스는 여러 사용자가 서로 다른 목적으로 데이터를 공동으로 이용하므로 공용 데이터이며 대용량화되고 구조가 복잡한 것이 보통이다.
- ④ 데이터베이스는 새로운 데이터의 삽입, 기존 데이터의 삭제, 갱신으로 항상 변화하면서도 항상 현재의 정확한 데이터를 유지해야 하므로 고정된 데이터이다.

11. 다음 중 데이터에 관한 구조화된 데이터로서 다른 데이터를 설명해 주는 데이터로 정의되는 것은?

- ① 데이터모델 ② 메타데이터 ③ 백업데이터 ④ 데이터마트

12. 다음 중 주요 데이터 분석 기술에 대한 설명으로 가장 부적절한 것은?

- ① OLAP - 다차원의 데이터를 대화식으로 분석하기 위한 기술
- ② Business Intelligence - 데이터 기반 의사결정을 지원하기 위한 리포트 중심의 도구
- ③ Business Analytics - 의사결정을 위한 통계적이고 수학적 분석에 초점을 둔 기법
- ④ Deep Learning - 대용량 데이터에서 의미있는 정보를 추출하여 의사결정에 활용하는 기술

13. 다음의 용어와 설명의 연결이 가장 부적절한 것은?

- ① Analytics - 의사결정을 위한 통계적이고 수학적 분석에 초점을 둔 기법
- ② BI(Business Intelligence) - 데이터 기반 의사결정을 지원하기 위한 리포트 중심의 도구
- ③ OLTP - 다차원의 데이터를 대화식으로 분석하기 위한 소프트웨어
- ④ Data Mining - 대용량 데이터로부터 의미있는 관계, 규칙, 패턴을 찾는 과정

14. 다음 중 기업내부 데이터베이스인 고객관계관리(CRM)에 대한 설명으로 적절한 것은 무엇인가?

- ① 부품의 설계, 제조, 유통 등의 공정 포함
- ② 외부 공급업체와의 정보시스템 통합으로 시간과 비용 최적화
- ③ 기업의 내부 고객들만을 대상으로 한 정보시스템
- ④ 단순한 정보의 수집에서 탈피, 분석 중심의 시스템 구축 지향

15. 기업내부 데이터베이스 중 기업 전체가 경영자원을 효과적으로 이용하기 위해 통합적으로 관리하고 경영의 효율화를 기하기 위한 수단으로 정보의 통합을 위해 기업의 모든 자원을 최적으로 관리하기 위한 기업 경영정보시스템은?

- ① CRM(Customer Relationship Management)
- ② SCM(Supply Chain Management)
- ③ KMS(Knowledge Management System)
- ④ ERP(Enterprise Resource Planning)

16. 다음 중 사회기반 구조로서의 데이터베이스에 대한 설명으로 가장 부적절한 것은?

- ① 물류, 무역, 조세 등 사회간접자본 차원에서 정보망을 통해 유통, 이용된 정보가 데이터베이스로 구축
- ② 지리, 교통 부문에서 데이터베이스가 보다 고도화되어 데이터베이스를 구축
- ③ 인터넷의 보편화로 데이터베이스가 사회 전반의 인프라로 자리매김
- ④ 의료, 교육, 행정 부문에서는 데이터베이스 구축과 활용이 활성화되지 못함

17. 다음 중 유통부분의 대한 내용으로 부적절한 것은?

- ① 2000년 이후 전반적인 IT 변화 환경에 맞물려 CRM과 SCM 구축이 이뤄졌다.
- ② 전자문서 교환이 본격화되면서 유통부분의 주요 기업들이 많은 예산을 투입하여 SCM을 구축하기 시작했다.
- ③ 최근 전자태그가 등장하면서 유통부문에서는 전자태그를 지원하는 대용량 데이터베이스를 지원하는 플랫폼이 요구되는 상황이다.
- ④ 2000년대 초반에는 EAI, ERP, e-CRM 등과 같이 데이터베이스간의 정보 공유 및 통합이나 고객 정보의 전략적 활용이 주된 테마였다.

18. 다음 중 빅데이터에 대한 정의를 설명한 것으로 가장 부적절한 것은?

- ① 대규모 데이터에서 저비용으로 가치를 추출, 초고속으로 수집 및 분석하기 위한 아키텍처이다.
- ② 용량은 방대하지만 구조가 단순한 데이터세트의 집합이다.
- ③ 일반적인 데이터베이스 소프트웨어로 저장, 분석할 수 있는 범위를 초과하는 규모를 빅데이터라 정의한다.
- ④ 데이터의 양, 수집, 처리 속도가 급격히 증가하면서 나타난 현상이다.

19. 빅데이터로 인한 새로운 도전과 기회를 잘 요약한 특징으로 가트너그룹의 더그래니가 정리한 빅데이터 세가지 측면으로 부적절한 것은?

- ① 양(Volume) ② 다양성(Variety) ③ 속도(Velocity) ④ 유효성(Vaild)

20. 다음은 빅데이터의 출현 배경에 대한 설명이다. 옳지 않은 것은 ?

- ① 여러 거대 기업들이 온, 오프라인을 통해 사용자와 소비자의 다양한 정보를 수집 분석하여 경영과 전략에 활용하게 되었다.
- ② 기업이 보유한 데이터는 급기야 1페타바이트(PB)이상으로 늘어나고 있으며, 보유한 데이터에 숨어있는 가치를 발굴해 새로운 성장 동력원으로 만들 수 있는 환경이 되었다.
- ③ 컴퓨터 기술의 발전으로 저장 기술의 다양화와 발전은 가격 상승을 유도했고 분석기법의 발전을 통해 적은 데이터에서도 새로운 인사이트를 발견할 수 있게 발전되었다.
- ④ 디지털화의 급진전, 인터넷의 발전과 모바일 시대의 진전에 따른 클라우드 컴퓨팅의 보편화도 빅데이터의 출현 배경에서 직간접적으로 영향을 미쳤다.

21. 다음 중 빅데이터의 수집, 구축, 분석의 최종 목적으로 가장 적절한 것은?

- ① 새로운 통찰과 가치를 창출
- ② 데이터 중심 조직 구성
- ③ 초고속 데이터 처리 기술 개발
- ④ 데이터 관리 비용 절감

22. 빅데이터의 기능 중 ‘공동 활용의 목적으로 구축된 유, 무형의 구조물 역할을 수행한다.’라는 것에 해당하는 내용은 무엇인가?

- ① 산업혁명 시대의 석탄, 철
- ② 21세기의 원유
- ③ 렌즈
- ④ 플랫폼

23. 다음 중 빅데이터가 만들어 내는 변화로 가장 부적절한 것은?

- ① 사전처리에서 사후처리 시대로의 변화
- ② 대면조사에서 표본조사로의 변화
- ③ 데이터의 질보다 양의 중요도 증가
- ④ 인과관계에서 상관관계의 중요도 증가

24. ‘특정 그룹의 편중된 의견으로 인해 왜곡된 결과를 초래하는 문제가 빅데이터의 도입으로 해결되고 있다.’와 관련 있는 변화는 무엇인가?

- ① 사후처리로의 변화
- ② 전수조사로의 변화
- ③ 인과관계로의 변화
- ④ 상관관계로의 변화

25. 다음 중 데이터의 가치 측정이 어려운 이유로 적절하지 않은 것은 무엇인가?

- ① 데이터 재사용의 일반화로 특정 데이터를 언제 누가 사용했는지 알기 힘들기 때문이다.
- ② 빅데이터 전문 인력의 증가로 다양한 곳에서 빅데이터가 활용되고 있기 때문이다.
- ③ 분석기술의 발전으로 과거에 분석이 불가능했던 데이터를 분석할 수 있게 되었기 때문이다.
- ④ 빅데이터는 기존에 존재하지 않던 새로운 가치를 창출하기 때문이다.

26. 클라우드 분산 병렬처리 컴퓨팅 중 빅데이터와 같은 대용량 데이터의 처리 비용을 획기적으로 줄인 방식으로 적절한 것은?

- ① 맵부치 ② API ③ Ngram ④ 맵리듀스

14회 기술

27. 다음 중 감성 분석(Sentimental Analysis)에 대한 설명으로 가장 부적절한 것은?

- ① 특정 주제에 대한 사용자의 긍정·부정 의견을 분석한다.
 ② 주로 온라인 쇼핑몰에서 사용자의 상품평에 대한 분석이 대표적 사례이다.
 ③ 사용자간의 소셜 관계를 알아내고자 할 때 이용한다.
 ④ 사용자가 사용한 문장이나 단어가 분석 대상이 된다.

12회 기술

28. 아래와 같은 비즈니스 문제가 있다면, 각 문제를 해결하기 위해 주로 사용되는 기법과 연결이 적절하지 않은 것은?

- ① 맥주를 사는 사람은 콜라도 같이 구매하는 경우가 많다 - 연관규칙학습
 ② 고객의 만족도가 충성도에 어떤 영향을 미치는가? - 회귀분석
 ③ 친분관계가 승진에 어떤 영향을 미치는가? - 소셜 네트워크 분석
 ④ 택배차량을 어떻게 배치하는 것이 비용측면에서 가장 효율적인가? - 유형분석

6회 기술

29. '특정인과 다른 사람이 몇 촌 정도의 관계인가?' 를 파악할 때 사용되는 기법으로 적절한 것은?

- ① 연관 규칙 학습 ② 유전 알고리즘 ③ 사회관계망 분석 ④ 기계 학습

14회 기술

30. 다음 핀테크 분야에서 빅데이터 활용이 가장 핵심적인 분야인 것은?

- ① 크라우드 펀딩(Crowd Funding)
 ② 신용평가(Credit Rating)
 ③ 간편결제(Simple Payment)
 ④ 블록체인(Block Chain)

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

- ① LSTM(Long Short-Term Memory)
- ② Autoencoder
- ③ K-NN(K Nearest Neighborhood)
- ④ RNN(Recurrent Neural Network)

32. 최근에 딥러닝(Deep Learning)에 대한 관심이 전 세계적으로 높아지고 있다. 딥러닝을 활용하기 위해 다양한 오픈소스가 개발되어 제공되고 있다. 다음 중 이와 가장 관련이 없는 것은?

- ① Caffe ② Tensorflow ③ Anaconda ④ Theano

33. 다음 중 빅데이터 시대에 발생할 수 있는 위기 요인으로 가장 부적절한 것은?

- ① 재산권 침해 ② 데이터 오용 ③ 책임원칙 훼손 ④ 사생활 침해

34. 다음 중 빅데이터 시대 위기 요인 중 사생활 침해 문제를 해결하기 위해 개인 정보를 사용하는 자가 적극적인 보호 장치를 강구하게 하는 방법으로 가장 적절한 것은?

- ① 알고리즘에 대한 접근을 허용해 부당함을 반증할 방법을 명시해 공개할 것을 주문
- ② 개인정보 제공자가 허락하는 동의제의 도입
- ③ 개인정보를 사용하는 사람이 직접 책임지는 책임제의 도입
- ④ 정보 사용자에게 수집된 내용을 공개하고 접근하는 권리 부여

35. 애플 CEO였던 스티브 잡스는 새로운 제품을 개발할 때 사람들의 의견을 묻지 않는다고 했다. 왜냐하면 사람들이 일반적으로 필요로 하는 것은 현실에 대한 인식에 바탕으로 두고 있기 때문이다. 따라서 포드가 자동차를 만들려고 했을 때 사람들의 의견을 물었다면 사람들은 더 빠른 말이 필요하다는 대답을 했을 것이라고 비유를 들었다. 이러한 예를 통해 알 수 있는 빅데이터 시대의 위기 요인으로 적절한 것은?

- ① 소셜 네트워크 ② 사생활 침해 ③ 책임 원칙 훼손 ④ 데이터 오용

36. 다음 중 데이터화(datafication) 현상에 큰 영향을 미치는 기술로 적절한 것은?

- ① 사물인터넷(Internet of Things)
- ② 인공지능(Artificial Intelligence)
- ③ 가상현실(Virtual Reality)
- ④ 3D 프린팅(3D-Printing)

37. 다음은 빅데이터의 위기요인을 통제할 수 있는 방안에 대한 설명이다. 옳지 않은 것은?

- ① 동의에서 책임으로 : 현재 개인정보의 활용 동의를 통한 가공, 유통됨에 따른 사생활침해에 대비해 개인정보 사용자의 책임제로 변환하여 사용자의 적극적인 보호를 유도하고자 한다.
- ② 결과기반 책임 원칙 고수 : 사전적인 예측을 통해 판단보다는 행동 결과를 보고 판단하는 원칙을 고수함으로써 기업들이 담합할 확률이 높다고 먼저 처벌하지 말고, 사용자의 신용등급이 낮다고 불이익을 당할 가능성을 최소화하고자 한다.
- ③ 데이터 오용 방지 : 데이터를 통한 결과가 모든 것을 예측할 수 없기 때문에 가능하면 데이터 분석을 하지 말고 사용자의 판단에 따르고자 한다.
- ④ 알고리즘 접근 허용 : 데이터 오용의 위기에 대처하기 위해 활용되고 있는 알고리즘이 어떤 식으로 계산되는지 알고리즘을 소개하여 명시함으로써 문제가 발생한 피해자를 구제하고자 한다.

38. 데이터 사이언스는 데이터 처리와 관련된 IT 영역, 분석적 영역, 그리고 비즈니스 컨설팅 영역을 포괄하고 있다. 다음 중 세 개의 영역과 다른 영역에 속하는 하나는?

- ① 데이터 시각화
- ② 데이터 웨어하우징
- ③ 분산 컴퓨팅
- ④ 파이썬 프로그래밍

39. 다음 중 데이터 사이언스에 대한 설명으로 가장 부적절한 것은?

- ① 데이터 사이언스의 핵심 구성요소는 Analytics, IT, 그리고 Domains Expertise이다.
- ② 데이터 사이언스가 기존 통계학과 다른 점은 총체적 접근법을 사용한다는 점이다.
- ③ “데이터 사이언스는 과학과 인문의 교차로에 서 있다.”는 말은 학문간 통일의 중요성을 의미한다.
- ④ 데이터에서 의미있는 정보를 추출해내는 학문으로 데이터마이닝과 개념상으로 유사하다.

40. 다음 중 데이터 사이언티스트의 필요 역량으로 가장 부적절한 것은?

- ① 설득력있는 스토리텔링
- ② 통찰력있는 분석
- ③ 네트워크 최적화
- ④ 다분야 간 협력을 위한 커뮤니케이션

41. 다음 중 빅데이터와 데이터 사이언스에 대한 설명으로 가장 부적절한 것은?

- ① 데이터 사이언스는 정형화된 실험데이터를 분석 대상으로 한다.
- ② 미래 가치 패러다임과 변화에서 빅데이터 분석 활용 능력은 핵심적인 역할을 할 것이다.
- ③ 데이터 사이언스에서 시각화와 효과적인 커뮤니케이션은 매우 중요한 요소이다.
- ④ 기업에서의 빅데이터 분석은 기업의 분석 문화에 결정적으로 영향을 받는다.

42. 고객테이블(CUSTOMERS)로부터 나이(AGE)가 20~30대인 고객정보(NAME, GENDER, SALARY)를 추출하기 위해 아래와 같은 SQL문을 작성하려고 한다. 다음 (가) 안에 들어갈 적절한 구문을 채워 쓰시오.

```
SELECT NAME, GENDER, SALARY
FROM CUSTOMERS
WHERE AGE ( 가 ) 20 AND 39
```

()

43. 아래에서 설명하고 있는 (가)와 (나)적절한 용어를 쓰시오.

데이터 사이언티스트가 갖춰야 할 역량은 빅데이터의 처리 및 분석에 필요한 이론적 지식과 기술적 숙련에 관련된 능력인 (가) skill 과 데이터 속에 숨겨진 가치를 발견하고 새로운 발전 기회를 만들어 내기 위한 능력인 (나) skill 로 나누어진다.

()

44. 아래 (가) 안에 들어갈 용어를 기입하시오.

(가)는 데이터의 가공 및 상관관계 간 이해를 통해 패턴을 인식하고 그 의미를 부여한 것이며, 지식을 도출하기 위한 재료가 된다.

()

45. 아래 데이터 분석과 관련된 기술을 설명한 것이다. (가) 에 들어갈 용어를 기입하시오.

기업의 의사결정 과정을 지원하기 위한 주제 중심으로 통합적이며 시간성을 가지는 비휘발성 데이터의 집합을 (가)라고 한다.

()

46. 아래에서 설명하고 있는 (㉠)은 무엇인가?

지난 몇 년간 여러 사일로 대신 하나의 데이터 소스를 추구하는 경향이 생겼다. 전사적으로 쉽게 인사이트를 공유하는 데 도움이 되기 때문이다. 다시 말해 별도로 정제되지 않은 자연스러운 상태의 아주 큰 데이터 세트인 (㉠)을/를 기업들이 구현하는 것은 2017년 새롭게 등장한 트렌드가 아니다. 그러나 2017년은 이를 적절히 관리해 운영하는 첫해가 될 전망이다.

()

47. 형태와 형식이 정해져 있지 않고 언어 또는 문자로 기술되는 데이터를 무엇이라고 하는가?

()

48. 아래는 데이터의 양을 표현하는 단위를 값이 작은 것에서부터 큰 것 순으로 나열한 것이다. (㉠)에 적합한 단위는 무엇인가?

기가바이트(GB) < 테라바이트(TB) < (㉠) < 엑사바이트(EB)

()

