

PART 1. 데이터의 이해_Test

데이터 준전문가

ADSP, Advanced Data Analytics semi-Professional

류영표 강사

ryp1662@gmail.com

Test

수업시간에 다뤘던 문제들에 대한 출제이다.

객관식 8문제, 주관식 2문제

1. 다음 중 구조 관점의 데이터 유형 중 아래에서 설명하는 것은 무엇인가?

스키마(형태) 구조 형태를 가지고 XML, HTML, 웹 로그, 시스템 로그, 알람 등과 같이 메타데이터를 포함하며, 값과 형식에서 일관성을 가지지 않는 데이터

- ① 정형데이터
- ② 비정형데이터
- ③ 반정형 데이터
- ④ 스트림 데이터



2. 다양한 데이터 유형 중 정형 데이터-반정형 데이터-비정형데이터 순서로 가장 알맞은 것은?

- ① 인스타그램 게시물 - 기상청 날씨 데이터 - 웹 로그 데이터
- ② 물류창고제고 데이터 - XML - 이메일 전송 데이터
- ③ CRM 데이터 - 카카오톡 대화 데이터 - Twitter 상태 메시지
- ④ RFID - IOT 센서 데이터 - 동영상 데이터



3. 다음은 DIKW(Data, Information, Knowledge, Wisdom hierarchy)에 대한 설명이다.

가장 적절한 설명은 무엇인가?

- ① 지식(Knowledge)은 근본 원리에 대한 깊은 이해를 바탕으로 도출되는 창의적 아이디어로 설명할 수 있다.
- ② 정보(Information)는 상호 연결된 정보 패턴을 이해하고, 이를 토대로 예측한 결과물이라 할 수 있다.
- ③ 지혜(Wisdom)는 데이터의 가공 및 상관관계 간 이해를 통해 패턴을 인식하고, 그 의미를 부여한 데이터라고 할 수 있다.
- ④ 데이터(Data)는 존재형식을 불문하고 타 데이터와의 상관관계가 없는 가공하기 전의 순수한 수치나 기호를 의미한다고 할 수 있다.



4. 다음은 데이터베이스의 일반적인 특징에 관한 설명이다. 가장 부적절한 것은 무엇인가?

- ① 데이터베이스는 통합된 데이터(integrated data)이다.
- ② 데이터베이스는 저장된 데이터(stored data)이다.
- ③ 데이터베이스는 공용 데이터(shared data)이다.
- ④ 데이터베이스는 변화되지 않는 데이터(unchanged data)이다.



5. 빅데이터 출현 배경 중 거대한 데이터의 분석 비용 문제를 해결해 준 것은 무엇인가?

- ① 디지털 기술
- ② 클라우드 컴퓨팅 기술
- ③ 하드 드라이브 가격의 하락
- ④ SNS 확산



6. 커피를 사는 사람들이 탄산음료도 많이 구매하는 지를 알아보기 위해 사용되는 분석은?

- ① 회귀 분석(Regression Analysis)
- ② 기계 학습(Machine learning)
- ③ 유전 알고리즘(Genetic algorithm)
- ④ 연관 규칙 학습(association rule learning)



7. 빅데이터 활용에 필요한 기본적인 3요소로 가장 적절한 것은?

- ① 데이터, 기술, 인력
- ② 데이터, 기술, 프로세스
- ③ 기술, 인력, 프로세스
- ④ 데이터, 인력, 프로세스



8. 다음 중 빅데이터 출현 배경에 관한 설명으로 부적절한 것은?

- ① 개별 기업의 데이터 축적 및 데이터 활용에 대한 니즈 증가
- ② 데이터 저장 기술 발전과 저장 비용 감소
- ③ 인터넷, SNS와 사물네트워크의 확산으로 데이터 생산량 증가
- ④ 수집 관리 및 분석에 용이한 형태로 데이터 구조의 정형화



9. 빅데이터의 가치 선정이 어려운 이유로 올바른 것은?

- ① 분석기술의 발전 한계
- ② 데이터 다목적성의 축소
- ③ 데이터가 기존에 없던 가치를 창출
- ④ 빅데이터의 등장



10. 인터넷의 진화는 수많은 센서들이 인터넷으로 연결되는 사물인터넷(IOT) 시대로 나아가고 있다. 미래의 빅데이터 관점으로 볼 때 사물인터넷과 가장 관련이 큰 것은?

- ① 인공지능(AI)
- ② 데이터화(Datafication)
- ③ 스마트 데이터(Smart Data)
- ④ 지능적 서비스(Intelligence Service)



11. 다음 중 사생활 침해를 막기 위해 개인정보를 무작위 처리하는 등 데이터가 본래 목적 외에 가공되고 처리되는 것을 방지하는 기술은 무엇인가?

- ① 정규화
- ② 난수화
- ③ 익명화
- ④ 일반화



12. 아래 빅데이터 활용을 위한 기본 테크닉 중 어떤 사례를 활용하는가?

A 마트는 금요일 저녁에 맥주를 사는 사람은 기저귀도 함께 구매했다는 사실을 발견하고, 두가지 상품을 가까운 곳에 진열하기로 결정했다.

- ① 회귀분석
- ② 연관성분석
- ③ 유형분석
- ④ 구문분석



13. 아래 보기에서 데이터의 양의 표시 단위를 작은것부터 큰 순서로 나열하시오.

- a. 엑사바이트(EB)
- b. 페타바이트(PB)
- c. 요타바이트(YB)
- d. 제타바이트(ZB)



14. 조직이나 기업의 인적 자원이 축적되고 있는 개별적인 지식을 체계화하여 공유함으로써 경쟁력을 향상시키기 위한 기업 정보시스템을 무엇이라 하는가?



15. 인터넷으로 연결된 기계마다 통신 장치를 있는 환경에서 사람 또는 기계끼리 자동으로 통신하는 기술로써 사물과 사람, 사물과 사물 간의 상호 소통하는 방식을 무엇이라 하는가?



16. 아래 빈 칸에 알맞은 데이터베이스 용어는?

데이터는 ()는 데이터베이스 내의 데이터에 대한 정확성, 일관성, 유효성, 신뢰성을 보장하기 위해 데이터 변경 혹은 수정시 여러 가지 제한을 두어 데이터의 정확성을 보증하는 것을 말한다.



17. 아래 빈 칸에 알맞은 용어는?

() 프로세스는 기업 시스템 내의 축적된 데이터 중 A 시스템에 데이터를 추출(Extract), 데이터웨어하우스 등에서 이용하기 쉬운 형태로 변환(Transform)해, 대상이 되는 B 시스템에 적재(Load) 또는 이들 일련의 처리 지원하는 이렇게 처리된 데이터는 시각화(Visualization) 작업을 통해 기업 의사결정권자들을 위한 비즈니스 리포로 만들어 진다.



Test

복원문제에 대한 출제이다.

객관식 8문제, 주관식 2문제

1. 데이터베이스의 특징으로 가장 부적절한 것은?

- ① 데이터베이스는 여러 사용자가 서로 다른 목적으로 데이터를 공동으로 이용할 수 있도록 구성되어 있다.
- ② 데이터베이스는 통합된 데이터(integrated data)다.
- ③ 데이터베이스는 변화하는 데이터로 데이터의 삽입, 삭제, 갱신을 하더라도 항상 현재의 정확한 데이터를 유지해야 한다.
- ④ 데이터베이스는 검색기능을 가지고 있으므로 다양한 방법으로 필요한 정보를 검색할 수 있다.



2. 다음 중 데이터 사이언티스트(data scientist)에게 요구되는 소프트 역량(soft skill)이 아닌 것은?

- ① 이론적 지식
- ② 창의적 사고
- ③ 커뮤니케이션 기술
- ④ 비주얼라이제이션을 활용한 설득력



3. 다음 중 빅데이터 분석에 경제성을 제공해 준 결정적인 기술로 가장 적절한 것은?

- ① 저장장치 비용의 지속적인 하락
- ② 텍스트 마이닝
- ③ 클라우드 컴퓨팅
- ④ 스마트폰의 급속한 확산



연습문제

4. 인터넷 등 각종 경로로 정보를 수집하는 구글은 이미 지난 2010년에 서비스 사용자가 1시간 뒤에 어떤 일을 할지 87% 정확도로 예측할 수 있는 데이터와 분석 신뢰도를 확보하고 있다고 했다. 또, 여행사실을 트위터한 사람의 집을 강도가 노리는 고전적 사례도 발생했다. 이러한 사례를 통해 알 수 있는 빅데이터 시대의 위기 요인으로 적절한 것은?

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



5. 사물끼리 정보를 주고 받는 사물인터넷 시대를 빅데이터 관점에서 바라볼 때 다음 중 사물인터넷의 의미로 가장 적절한 것은?

- ① 모든 것의 데이터화(datafication)
- ② 서비스 지능화(intelligent service)
- ③ 분석 고급화(advanced analytics)
- ④ 정보 공유화(information sharing)



6. 개인의 사생활 침해를 방지하고 통계 응답자의 비밀사항은 보호하면서 통계자료의 유용성을 최대한 확보 할 수 있는 데이터변환 방법은 무엇인가?



7. 빅데이터 시대에는 데이터를 많이 확보했거나 확보 할 수 있는 기업이 혁신을 시도하거나 경쟁력과 생산성 향상을 도모하기에 유리하다. 다음 보기 중 이러한 속성에 부합되기 어려운 기업 분류는?

- ① 신용카드회사
- ② 여행회사
- ③ B2B기업
- ④ 이동통신사



8. 데이터와 정보의 차이를 구분하는 것은 중요하다. 다음 중 정보에 대한 예로 부적절한 것은?

- ① 평균 구매액
- ② 주문 수량
- ③ 베스트셀러
- ④ 우량 고객



9. 영화 '마이노리티 리포트'에 나오는 것처럼 범죄 예측 프로그램에 의해 범행을 저지르기 전에 체포될 수 있는 사례를 통해 알 수 있는 빅데이터 시대의 위기 요인으로 적절한 것은?

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



10. 전략적 분석을 통해 놀라운 성과를 얻은 미국의 최대 카지노 관련 회사인 하라스엔터테인먼트의 회장인 러브먼은 분석 기반 경영이 도입되지 못하는 이유를 이야기하였다. 보기에서 그 내용이 아닌 것은?

- ① 기존 관행을 그냥 따를 뿐 중요한 시도를 하지 않는다.
- ② 경영진이 의사결정을 할 때 직관적으로 결정했을 때 성과가 나오는 것을 경영진의 진정한 재능이라고 생각한다.
- ③ 분석적 실험을 능숙하게 해내는 사람이 많지 않고, 적절한 방법조차 제대로 익히지 못한 사람이 분석업무를 한다.
- ④ 사람들은 아이디어를 낸 사람이 누군지 보다는 아이디어 자체에 관심을 더 많이 가지고 있다.



11. 다음 개인정보 비식별화 기술 중 아래에서 설명하고 있는 것으로 가장 적절한 것은?

개인정보의 주요 식별요소를 다른 값으로 대체하여 개인 식별을 어렵게 만드는 기술

- ① 가명처리(Pseudonymisation)
- ② 데이터 삭제(data Reduction)
- ③ 범주화(data suppression)
- ④ 데이터 마스킹(data masking)



12. 인터넷상의 서버에서 데이터 저장, 처리, 네트워크, 콘텐츠 사용 등 서로 다른 물리적인 위치에 존재하는 컴퓨팅 자원을 가상화 기술을 통해 IT 관련 서비스를 한 번에 제공하는 혁신적인 컴퓨팅 기술은 무엇인가?



13. 다음 중 빅데이터 현상이 출현하게 된 배경과 가장 거리가 먼 것은?

- ① 의료정보 등 공공데이터의 개방 가속화
- ② M2M, IOT와 같은 통신 기술의 발전
- ③ 하둡 등 분산처리 기술의 발전
- ④ 트위터, 페이스북 등 SNS의 급격한 확산



14. 기업 내부 데이터베이스를 기반으로 다양한 정보시스템이 구축·활용되고 있다. 고객 관련 데이터베이스를 분석하여 고객 개개인에게 적합한 차별적 제품 및 서비스를 제공함으로써 고객과의 관계를 지속적으로 강화해 나가기 위해 구축하는 정보시스템은 다음 중 무엇인가?

- ① CRM 시스템
- ② SCM 시스템
- ③ ERP 시스템
- ④ KMS 시스템



15. 다음 중 데이터의 유형이 다른 하나는?

- ① 개인 페이스북에 올린 어느 회사 제품에 대한 사용 후기글
- ② 어느 기계에서 작동하는 동안 발생한 소음을 데시벨 단위로 기록한 센서 데이터
- ③ 어느 포털 사이트에서 하루 동안 언급된 모든 검색어
- ④ 콜센터에 접수된 어느 고객의 제품 불만사항을 녹음한 음성파일



16. 다음 개인정보 비식별화 기술 중 아래에서 설명하고 있는 것으로 가장 적절한 것은?

은행에서 대출 심사를 할 때, 소득, 카드 사용액, 나이 등 해당 고객의 개인적인 정보를 바탕으로 그 고객이 대출 상환을 잘 하는 집단에 속할지 그렇지 않은 집단에 속할지는 예측 할 수 있다.



17. 데이터 사이언티스트가 효과적인 분석 모델 개발을 위해 고려해야 하는 사항으로 가장 부적절한 것은?

- ① 분석모델이 예측할 수 없는 위험을 살피기 위해 현실세계를 돌아보고 분석을 경험과 세상에 대한 통찰력과 함께 활용한다.
- ② 가정들과 현실의 불일치에 대해 끊임없이 고찰하고 모델의 능력에 대해 항상 의구심을 가진다.
- ③ 분석의 객관성에 의문을 제기하고 분석 모델에 포함된 가정과 해석 개입 등의 한계를 고려한다.
- ④ 넓은 시각에서 모델 범위 바깥의 요인들을 판단할 수 있도록 가능한 많은 과거 상황 데이터를 모델에 포함한다.



18. 다음 중 일반적으로 통용되고 있는 빅데이터의 정의와 거리가 가장 먼 것은?

- ① 빅데이터는 일반적인 데이터베이스 소프트웨어로 저장,관리, 분석할 수 있는 범위를 초과하는 규모의 데이터이다.
- ② 빅데이터는 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처이다.
- ③ 빅데이터는 데이터의 양(Volume), 데이터 유형과 소스 측면의 다양성(Variety), 데이터 수집과 처리 측면에서도 속도(Velocity)가 급격히 증가하면서 나타난 현상이다.
- ④ 빅데이터는 기존의 작은 데이터 처리 분석으로는 얻을 수 없었던 통찰과 가치를 하둡(Hadoop)기반으로 대용량의 분산처리 기술을 통해 창출하는 새로운 방식이다.

19. 머신러닝 알고리즘은 크게 지도학습(Supervised learning)과 비지도학습(Unsupervised learning) 으로 나눌 수 있다. 이러한 측면에서 보기 중 나머지와 성격이 다른 것은?

- ① 군집분석
- ② 분류분석
- ③ 감성분석
- ④ 회귀분석

20. 다음 중 사용자 정의 데이터나 멀티미디어 데이터 등 데이터 구조를 표현 및 관리할 수 있는 데이터베이스 관리 시스템(DBMS)으로 적절한 것은?

- ① 관계형 DBMS
- ② 객체지향 DBMS
- ③ 네트워크 DBMS
- ④ 계층형 DBMS

21. 아래는 기업 내부에서 사용하는 데이터베이스의 활용에 대한 설명이다. (가)에 들어갈 알맞은 용어는?

(가)(은) 는 기업이 외부 공급업체 또는 제휴업체와 통합된 정보시스템으로 연계하여 시간과 비용을 최적화시키기 위한 것으로, 자재구매, 생산 · 재고, 유통 · 판매, 고객 데이터로 구성 됨.



22. 분석의 전형적인 의사결정 오류를 로직오류와 프로세스 오류로 나눠 보았을 때, 각 오류에 대한 설명 중 가장 부적절한 것은?

- ① 부정확한 가정을 하고 테스트를 하지 않은 것은 로직 오류이다.
- ② 결정에서 분석과 통찰력을 고려하지 않는 것은 프로세스 오류이다.
- ③ 데이터 수집이나 분석이 너무 늦어 사용할 수 없게 되는 것은 로직 오류이다.
- ④ 대안을 진지하게 고려하지 않는 것은 프로세스 오류이다.



23. 데이터웨어하우스는 기업 내의 의사결정지원 어플리케이션을 정보 기반을 제공하는 하나의 통합된 데이터 저장 공간을 말한다. 다음 중 데이터웨어하우스의 고유한 특성이 아닌 것은?

- ① 데이터웨어하우스에서는 데이터의 지속적 갱신에 따른 무결성 유지가 무엇보다 중요하다.
- ② 데이터웨어하우스의 데이터들은 전사적 차원에서 일관된 형식으로 정의된다.
- ③ 데이터웨어하우스에서 관리하는 데이터들은 시간의 흐름에 따라 변화하는 값을 저장한다.
- ④ 데이터웨어하우스에서는 특정 주제에 따라 데이터들이 분류, 저장, 관리된다.



24. 다음 중 데이터베이스를 관리하여 응용프로그램들이 데이터베이스를 공유하며 사용할 수 있는 환경을 제공하는 소프트웨어는?

- ① DBMS
- ② ERD
- ③ SQL
- ④ Data Dictionary



25. 암묵지와 형식지의 상호작용 관계를 가장 적절하게 표현한 것은 무엇인가?

- ① 내면화 → 연결화 → 표출화 → 공통화
- ② 표출화 → 공통화 → 내면화 → 연결화
- ③ 공통화 → 표출화 → 연결화 → 내면화
- ④ 내면화 → 연결화 → 표출화 → 공통화



26. 아래에 설명하는 (가) 는 무엇인가?

(가) 는 인터넷을 기반으로 모든 사물을 연결해 사람과 사물, 사물과 사물 간의 정보를 상호 소통하는 지능형 기술 및 서비스이며, 사물에서 생성되는 Data를 활용한 분석을 통해 마케팅 등에 활용할 수 있다.



27. 다음 중 빅데이터 시대에 발생할 수 있는 위기 요인 중 사생활 침해 문제를 해결하기 위한 방법으로 가장 적절한 것은?

- ① 알고리즘 접근 허용
- ② 결과기반 책임 원칙 고수
- ③ 데이터 오용 방지
- ④ 정보 사용자 책임제로 변환



연습문제

28. 다음은 데이터베이스의 구성요소들을 설명한 것이다. 각 설명에 해당하는 구성요소를 가장 적절하게 나열한 것은?

- (A) 데이터에 관한 구조화된 데이터로, 다른 데이터를 설명해 주는 데이터
- (B) 데이터베이스 내의 데이터를 신속하게 정렬하고 탐색하게 해주는 구조

- ① (A) - 메타데이터, (B) - 인덱스
- ② (A) - 데이터모델, (B) - 트리거
- ③ (A) - 백업데이터, (B) - 저장된 절차
- ④ (A) - 스키마구조, (B) - 데이터마트

29. 다음 중 빅데이터 가치 산정이 어려운 이유로 부적절한 것은?

- ① 기존 데이터를 재사용, 재조합을 통해 새로운 가치를 창출할 수 있다.
- ② 특정데이터를 언제, 어디서, 누가 활용하느냐에 따라 가치를 산정할 수 있다.
- ③ 기존에 없던 가치를 창출함에 따라 그 가치를 측정하기 어려워졌다.
- ④ 기존 사업자에게 경쟁우위를 가져다 준다.



30. 데이터사이언스는 데이터처리와 관련된 IT영역, 분석적 영역, 그리고 비즈니스 컨설팅 영역을 포괄하고 있다. 다음 중 세 개의 영역과 다른 영역에 속하는 하나는?

- ① 데이터웨어하우징
- ② 데이터 시각화
- ③ 분산컴퓨팅
- ④ 파이썬 프로그래밍



31. 다음 내용이 설명하고 있는 빅데이터의 역할로 알맞은 것은?

이것은 비즈니스 측면에서 일반적으로 ‘공동 활용의 목적으로 구축된 유무형의 구조물’을 의미한다. 각종 사용자 데이터나 M2M센서 등에서 수집된 데이터를 가공·처리·저장해두고 이 데이터에 접근할 수 있도록 API(Application Program Interface)를 공개한다. 그러면 다양한 서드파티(3rd-party)사업자들이 비즈니스에 필요한 정보를 추출해 활용하게 되고 빅데이터는 그 자체로 이 역할을 수행하게 된다.

32. 데이터의 특징 중 부적절한 것은?

- ① 데이터란 용어는 1646년 영국 문헌에 처음으로 등장한 것으로 알려져 있다.
- ② 데이터 추론, 예측, 전망, 추정을 위한 근거로서의 특성을 갖는다.
- ③ 데이터는 암묵지와 형식지의 유형으로 구성되어 있다.
- ④ 데이터는 객관적 사실이라는 존재적 특성을 갖는다.



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

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



34. 다음 중 데이터에 관한 구조화된 데이터로서 다른 데이터를 설명해주는 데이터로 정의하는 것은?

- ① 데이터 모델
- ② 메타데이터
- ③ 백업데이터
- ④ 데이터마트



연습문제

35. 아래는 빅데이터가 만들어 내는 본질적인 변화이다. (A)와 (B)에 들어갈 말을 쓰시오.

(A) 은(는) 어떤 현상에 대하여 현상을 발생시킨 원인과 그 결과 사이의 관계를 말하고, (B)은(는) 어떤 두 현상이 관계가 있음을 말하지만 어느 쪽이 원인인지 알 수 없다.



Thank you.

ADSP/ 류영표 강사
ryp1662@gmail.com