

## ADsP 33회 기출문제 답안

### 【객관식 정답】

01	①	11	④	21	②	31	②
02	③	12	①	22	④	32	②
03	③	13	④	23	①	33	②
04	④	14	①	24	③	34	③
05	①	15	①	25	①	35	②
06	②	16	③	26	④	36	①
07	①	17	①	27	④	37	①
08	①	18	②	28	③	38	③
09	②	19	①	29	④	39	③
10	①	20	④	30	④	40	④

### 【주관식 정답】

01	데이터 웨어하우스 (Data Warehouse)
02	데이터 사이언스 (Data Science)
03	통찰(Insight)
04	나선형 모델(Spiral Model)
05	특이도(Specificity)
06	3개
07	제 1종 오류
08	오즈(Odds)
09	소프트맥스(Softmax) 함수
10	기울기 소실(Gradient Vanishing)

01. 사물인터넷(IoT)은 인터넷을 기반으로 모든 사물을 연결하여 사물과 사물, 사물과 사람 간의 정보를 상호 소통하는 지능형 기술 및 서비스이다. 빅데이터 관점에서 사물인터넷은 사물에서 나오는 데이터를 활용해 더욱 지능화 된 기기 활용을 할 수 있도록 데이터를 수집하여야 되므로 모든 사물에서 데이터를 추출할 수 있어야 된다.
02. 책임원칙 훼손의 통제방안은 결과 기반 책임 원칙 고수이며 데이터 오용은 알고리즘 접근의 허용이다.
03. NoSQL(비관계형) 데이터베이스에는 HBase, MongoDB, Redis 등이 속한다. MySQL은 관계형 데이터베이스에 해당한다.
04. 데이터베이스의 일반적인 특징으로는 통합된 데이터, 저장된 데이터, 공유 데이터, 변화되는 데이터로 변화되는 데이터는 데이터베이스에 저장된 내용은 곧 데이터베이스의 현 시점에서의 상태를 나타낸다.
05. 소프트 스킬에는 통찰력 있는 분석, 설득력 있는 전달, 다분야간 협력 등이 있다. 이론적 지식은 하드 스킬에 포함이 된다.
06. 데이터의 크기는 작은 것부터 큰 것까지 순서대로 페타바이트(PB), 엑사바이트(EB), 제타바이트(ZB), 요타바이트(YB)이다.
07. 빅데이터 분석 활용의 효과로 서비스 산업의 확대되지만, 제조업의 경우에도 생산성 향상, 불량률 감소 등에 대한 성과를 거둘 수 있기에 제조업이 축소되는 것은 아니다.
08. 빅데이터 시대에는 과거에는 표본조사였으나 현재는 전수조사로 바뀌었다.

09. 분석과제의 적용 우선순위 기준을 '시급성'에 둔다면 III - IV - II 이다.

10. 분석과제의 우선순위 평가에서 시급성은 전략적 중요도와 목표가치를 평가하고 난이도는 데이터획득/저장/가공비용과 분석 적용 비용, 분석수준이 평가요소이다.

11. 하향식 접근법의 데이터 분석 기획 단계는 문제 탐색(Problem Discovery) → 문제 정의(Problem Definition) → 해결 방안탐색(Solution Search) → 타당성 검토(Feasibility study)이다.

12. 정확도(Accuracy)는 True를 True, False를 False라고 예측하는 지표이며, 정밀도(Precision)는 True라고 예측한 것 중에서 실제 True인 것의 비율이다.

13. 분석과제 우선순위 고려요소는 전략적 중요도, 비즈니스 성과/ROI, 실행 용이성이 있다.

14. 최상위 계층은 단계로 구성되고 마지막 계층은 스텝으로 구성된다.

15. 분석 방법론의 분석 기획 단계에서는 비즈니스 이해 및 범위 설정, 프로젝트 정의 및 계획수립, 프로젝트 위험계획 수립을 주요 업무로 한다. 필요데이터 정의는 데이터 준비 단계에서 수행하는 Task이다.

16. 상향식 접근법은 비지도 학습(Unsupervised Learning)방식을 수행한다.

17. 이산형 확률변수의 기댓값은  $E(X) = \sum xf(x)$  이다.

18. 이동평균법은 시계열자료에서 계절변동과 불규칙변동을 제거하여 추세변동과 순환변동만 가진 시계열 자료로 변환하는 평활법이다.

19. 스피어만 상관계수는 서열척도의 자료를 대상으로 한다.

20. F-statistic 우측의 p-value를 보면 회귀식의 유의성을 확인할 수 있다. p-value < 2.2e-16으로 매우 작으므로 회귀식이 유의하다고 할 수 있다.

21. 활성화 함수는 입력 신호의 총합을 출력 신호로 변환하는 함수로 종류로는 계단, 시그모이드, ReLU, Softmax 등이 있다.

22. 비표본오차는 표본오차를 제외한 모든 오차로서 조사 과정에서 발생하는 모든 부주의나 실수, 알 수 없는 원인 등 모든 오차를 의미하며 조사대상이 증가하면 오차가 커진다.

23. 2번은 다차원척도법, 3번은 군집분석, 4번은 연관규칙분석에 대한 설명이다.

24. F1의 정의는  $2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$  이다. Recall과 Precision을 대입하면  $2 \times \frac{\frac{2}{5} \times \frac{2}{5}}{\frac{2}{5} + \frac{2}{5}} = 0.40$ 이다.

25.  $\text{Recall} = \frac{TP}{TP + FN} = \frac{30}{30 + 70} = \frac{3}{10}$  이다.

26. k개의 초기 중심값은 임의로 선택이 가능하므로 한번 군집을 형성되어도 군집 내 객체들은 다른 군집으로 이동이 될 수 있다.
27. 연관규칙의 분석 결과로는 품목 간 구체적인 영향을 줄수 있는지 알 수 없다.
28.  $P(\text{질병})=0.1$ , 진단 시 질병 가진 사람= $P(\text{양성})=0.2$ , 질병을 가지고 있는 사람 중 질병이라 진단 받은 사람= $P(\text{양성} \cap \text{질병})=0.9$ 이며, 구해야할 값은  $P(\text{질병}|\text{양성})$ 으로 베이즈 정리에 따라  $P(\text{양성} \cap \text{질병})=0.09$ 이다. 그러므로  $P(\text{질병}|\text{양성})=P(\text{양성} \cap \text{질병})/P(\text{양성})=0.09/0.2=0.45$ 이다.
29. A, B의 유클리드 거리는  $\sqrt{(185-180)^2 + (70-75)^2} = \sqrt{50}$  이다.
30. 주성분의 개수는 고윳값, 누적 기여율, Scree Plot를 통해 확인할 수 있다. 고유치 분해 가능여부는 주성분 개수를 정하는 방법이 아니다.
31. SOM은 역전파 알고리즘 등을 이용하는 인공신경망과 달리 단 하나의 전방 패스를 사용함으로써 속도가 매우 빠르다. 따라서 실시간 학습처리를 할 수 있는 모형이다.
32. 분리변수의 P차원 공간에 대한 현재 분할은 이전 분할에 영향을 받는다.
33. 부스팅은 예측력이 약한 모형들을 결합하여 강한 예측모형을 만드는 방법이다.
34. 향상도는  $\frac{2000}{\frac{5000}{\frac{3000}{5000} \times \frac{2500}{5000}}} = \frac{4}{3}$ 로 1보다 큰 값이 나와 연관성이 높다고 할 수 있다.
35. 새로운 변수가 추가될 때 기존 변수의 중요도가 약해질 수 있으므로 그러한 변수를 제거하는 방법이 단계별 선택법 (Stepwise selection)이다.
36. 생성된 모델이 훈련 데이터에 최적화되어 있기 때문에 테스트 데이터의 작은 변화에 민감하게 반응한다.
37. 구간추정은 모수의 참값이 포함되어 있다고 추정되는 구간을 결정하는 것이지만, 실제 모집단의 모수가 신뢰구간에 꼭 포함되어 있는 것은 아니다.
38. 산점도와 head, summary값으로 나무 종류별로 둘레에 유의한 차이가 있는지 알 수 없다.
39. Loadings에서 Comp.2의 값을 보면 head2와 breath2는 음의 값을 가짐을 보아 양의 상관관계를 가진다고 할 수 없다.
40. 동일 모집단에서 동일한 방법과 개수로 표본을 다시 추출하더라도 동일한 신뢰구간이 나오지 않을 수 있다.

## 【 단답형 】

---

단답형 01. 데이터 웨어하우스(Data Warehouse)

단답형 02. 데이터 사이언스(Data Science)

단답형 03. 통찰(Insight)

단답형 04. 나선형 모델(Spiral Model)

단답형 05. 특이도(Specificity)

단답형 06. height가 60에서 직선을 갖게 되면 총 3개의 군집으로 분류된다.

단답형 07. 제 1종 오류는 귀무가설이 실제로 참이어서 채택해야 함에도 불구하고 표본의 오차 때문에 이를 채택하지 않는 오류를 말한다.

단답형 08. 오즈(Odds)는 성공할 확률이 실패할 확률의 몇배인지를 나타낸다.

단답형 09. 소프트맥스(Softmax) 함수는 세 개 이상으로 분류하는 다중 클래스 분류에서 사용되는 활성화 함수이다.

단답형 10. 기울기 소실(Gradient Vanishing) 문제는 신경망의 활성화함수의 도함수 값이 계속 곱해지다 보면 가중치에 따른 결과값의 기울기가 0이 되어 버려서, Gradient Descent를 이용할 수 없게 되는 문제이다.