

ADsP 35회 기출문제 답안

【객관식 정답】

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

【주관식 정답】

01	데이터베이스(Database)
02	유전자 알고리즘 (Generic Algorithm)
03	하향식 접근 방식
04	집중 구조
05	다차원 척도법
06	후진 제거법 (Backward Elimination)
07	홀드아웃 방법
08	0.4
09	포아송 분포(Poisson Distribution)
10	최단 연결법

01. SQL은 데이터베이스를 사용할 때 데이터베이스에 접근할 수 있는 데이터베이스의 하부 언어이다.
02. 데이터 사이언티스트의 필요 역량은 하드 스킬과 소프트 스킬이 있으며, 소프트 스킬 중 통찰력 있는 분석, 설득력 있는 전달, 다분야간 협력이 있다.
03. 사생활 침해와 관련하여 익명화(anonymization) 기술 발전이 필요하다.
04. 알고리즘으로 부당한 피해를 보는 사람을 방지하기 위해서 생겨난 작업이다.
05. 사물인터넷의 적용으로 사람의 개입이 최소화 되고 기기에서 수집하는 데이터가 실시간으로 수집될 것이다.
06. 정부는 이익을 목적으로 하지 않고 공익을 목적으로 한다. 개인 정보를 활용하여 이익을 목적으로 하는 것은 사기업이다.
07. 빅데이터 시대 위기 요인은 사생활 침해, 책임 원칙 훼손, 데이터 오용이 있다.
08. DBMS는 다수의 사용자들이 DB 내의 데이터를 접근할 수 있도록 해주는 소프트웨어이다.
09. 데이터 거버넌스 체계 중 데이터 관리 체계는 데이터 정확성 및 활용의 효율성을 위하여 표준데이터를 포함한 메타 데이터와 데이터 사전의 관리 원칙을 수립한다. 빅데이터의 경우 데이터 양의 급증으로 데이터의 생명 주기 관리방안을 수립하지 않으면 데이터 가용성 및 관리비용 증대 문제에 직면하게 될 수 있다.

10. 데이터 거버넌스의 구성 요소인 원칙(Principle), 조직(Organization), 프로세스(Process)는 유기적으로 조합하고 효과적으로 관리하여, 데이터를 비즈니스 목적에 부합하도록 하고 최적의 정보 서비스를 제공할 수 있도록 한다.
11. 가치는 비즈니스 효과이다.
12. 분석 준비도를 측정하기 위한 요소는 분석업무, 인력 및 조직, 분석기법, 분석 데이터, 분석 문화, IT 인프라가 있다.
13. 사분면 영역에서 난이도와 시급성을 모두 고려할 때 가장 우선적인 분석 과제 적용이 필요한 영역은 난이도 : 쉬움, 시급성 : 현재를 나타내는 3사분면이다.
14. 분산구조는 분석조직 인력들을 현업부서로 직접 배치하여 분석업무를 수행하며, 전사차원의 우선순위에 따라 수행한다. 분석결과에 따른 신속한 Action이 가능하며, 베스트 프랙티스 공유가 가능하다.
15. 분석을 통한 상향식 접근법(Bottom Up Approach)의 프로세스는 프로세스 분류 → 프로세스 흐름 분석 → 분석 요건 식별 → 분석 요건 정의이다.
16. 분석 방법론 기획단계에서 프로젝트 위험 대응 계획을 수립할 때 예상되는 위험에 대해 회피(Avoid), 전이(Transfer), 완화(Mitigate), 수용(Accept)으로 구분하여 위험관리 계획서를 작성한다.
17. lasso 회귀모형에서는 사용하는 규제 방식을 L1 규제(Penalty)라고 한다.
18. 맨하탄 거리를 구하면 $|180-175|+|65-70|=10$ 이다.
19. 혼합분포 군집모형은 군집의 크기가 너무 작으면 추정 정도가 떨어지거나 어려울 수 있다.
20. 분류 모델링은 데이터가 어느 종류에 속하는지를 판별하는 모델이다.
21. $1*0.5 + 2*0.3 + 3*0.2 + 4*0 = 1.7$ 이다.
22. 분석 결과를 보면 EM 알고리즘을 통해 모수를 추정하는 과정에서 반복횟수 2회 만에 로그-가능도 함수가 최대가 됨을 알 수 있다.
23. X의 기댓값은 $\frac{1}{3} \times 1 + \frac{1}{6} \times 2 + \frac{1}{2} \times 3 = \frac{13}{6}$ 이다.
24. 정밀도는 $\frac{TP}{TP+FP} = \frac{30}{30+60} = \frac{1}{3}$ 이다.
25. 학생여부에 따른 Balance가 커지는거에 대한 유의적인 차이는 회귀분석 결과에서 확인하기 어렵다.
26. 단순회귀분석에서 결정계수는 상관계수의 제곱과 같다.
27. 연속형 목표변수의 분리 기준은 분산분석에서의 F-통계량, 분산 감소량이 있다.

28. SOM에서 입력층의 뉴런은 경쟁층에 있는 뉴런들과 완전연결(Fully Connected)되어 있다.
29. 모형의 적합성을 파악하기 위해서는 F-통계량으로 파악한다.
30. 산점도를 확인했을 때, horsepower이 커질수록 mpg가 작아지는 걸로 보아 상관계수가 음의 상관을 가질 것으로 말할 수 있어 관계를 파악하기에 유용하게 사용할 수 있다.
31. 지지도는 전체 거래 중 항목 A와 항목 B를 동시에 포함하는 거래의 비율로 정의한다.
32. 데이터 마이닝 추진단계는 순서대로 목적 설정, 데이터 준비, 가공, 기법적용, 검증이다.
33. 시그모이드 함수는 S자형 곡선 또는 시그모이드 곡선을 갖는 수학 함수로 로지스틱 회귀분석에서 사용된다.
34. 로지스틱 회귀모형은 반응변수가 범주형인 경우에 적용되는 모형이다.
35. DBSCAN은 데이터 형태가 모호하거나 다른 임의의 모양일 때 k-means보다 성능이 더 좋다.
36. 데이터 분할 방법으로는 홀드아웃방법, 교차검증, 부스트랩 등이 있다.
37. 특이도는 $\frac{TN}{TN+FP} = \frac{100}{300+100} = 0.25$ 이다.
38. 계층적 군집모형은 군집의 개수가 정해지지 않았을 때 사용하며, 군집의 개수를 모를 때 사용하기 때문에 몇 개의 군집으로 나누어야 하는지 결정하기 위해 사용하기도 한다.
39. 연관규칙분석은 상품의 구매, 서비스 등 일련의 거래 또는 사건들 간의 규칙을 발견하기 위해 적용하며, 조건과 반응의 형태(if-then)로 이루어져 있다.
40. 정상성을 만족하는지 판단하기 위해서는 평균과 분산이 일정한지 등에 대해 판단해야 한다.