

## □ 필답평가

\* 5문제 이상인 경우, 문제 난이도를 3개의 기준으로 나누어 주세요. 상(1), 중(2), 하(2)

능력단위요소		문항	
난이도	하	[문항1] 데이터분석의 직접적인 대상이 되는 8가지 객체 중에서 고객과 제품을 서로 연결해 주는 것은?	
도메인 이슈 도출하기		1. 채널 2. 시공간 3. 설비 4. 법제도	
난이도	하	[문항2] 빅데이터 분석문제를 발생시키는데 관련성을 가지며, 문제해결에도 영향을 미치는 모든 관련자들을 지칭하는 용어는?	
도메인 이슈 도출하기		1. 프로젝트관리자 2. 이해관계자 3. 목표고객 4. 퍼실리테이터	
난이도	상	[문항3] 빅데이터 분석목표 중에서 기획중인 여러 아이템 중에서 어떤 것이 상대적인 효과를 더 많이 발생시키는지 분석하는 기법은?	
분석목표 수립하기		1. CR테스트 2. GN테스트 3. 딥러닝테스트 4. AB테스트	
난이도	중	[문항4] 빅데이터 분석목표 중에서 기존 이벤트 실시결과 누가 반응을 보이고, 누가 보이지 않는지, 관련규칙을 도출하는 기법은?	
분석목표 수립하기		1. 연관예측 2. 수치예측 3. 분류예측 4. 군집예측	

난이도	하	[문항5] 빅데이터 분석방법론 중에서 해결해야하는 분석문제가 있고, 여기에 필요한 데이터나 분석기법을 매칭해서 분석하는 방법은?	
프로젝트 계획하기		1. 수요기반 빅데이터 분석 2. 데이터주도 빅데이터 분석 3. 샘플기반 빅데이터 분석 4. 실시간 빅데이터 분석	
난이도	중	[문항6] 일반적인 빅데이터 분석절차에서 문제해결에 필요한 직간접 지식을 수집하기 위해 각종 문헌조사, 인터뷰, 관찰, 벤치마킹을 통해 조사하는 단계는?	
프로젝트 계획하기		1. 문제인식 2. 자료리뷰 3. 데이터모델링 4. 분석평가해석	
난이도	하	[문항7] 빅데이터가 가진 3V특성에 대한 설명가 거리가 먼 것은?	
보유데이터 자산확인하기		1. Velocity: 실시간 분석이 가능해 졌다. 2. Volume: 대규모 데이터로부터 다양한 분석이 가능해 졌다. 3. Vision: 새로운 비전을 제시하는 분석이 가능해 졌다. 4. Variety: 다양한 데이터 유형으로부터 크로스체킹이 가능해 졌다.	
난이도	상	[문항8] 다음 중 성격이 다른 데이터 유형은?	
보유데이터 자산확인하기		1. 카드결제 데이터 2. 금융입출금 데이터 3. 거래처리 데이터 4. 소셜데이터	

난이도	중	[문항9] 가설에 대한 설명 중 알맞은 것은?	
가설 설정하기		1. 보통 가설은 귀무가설과 대립가설로 구분된다. 2. 대립가설은 비교대상 간에 차이가 없다는 것을 전제로 하는 가설이다. 3. 귀무가설은 대립가설이 기각이 되었을 때 채택하는 가설이다. 4. 대립가설과 귀무가설 모두 결론이 나지 않을 때에는 중립가설을 사용한다.	
난이도	상	[문항10] 가설검정에 대한 설명 중 거리가 먼 것은?	
가설 설정하기		1. 가설검정에 따라 1종오류, 2종오류, 3종오류가 있다. 2. 가설검정에 사용하는 유의수준( $\alpha$ ) 값은 0.05를 항상 고수해야 한다. 3. 2종 오류보다는 1종 오류가 더 중요하다. 4. 가설도 하나의 미래에 발생가능한 상황에 대한 예측분석이다.	
난이도	하	[문항11] 문제해결에 필요한 분석대상 객체에서 변수를 선택하고, 이들 간의 관계를 설정하는 작업을 무엇이라고 하는가?	
빅데이터 모델개발하기		1. 가설수립 2. 모델링 3. 과제기획 4. 액션시나리오수립	
난이도	하	[문항12] 주어진 데이터에서 추가적인 변수를 만들어 내는 작업과 거리가 먼 것은?	
빅데이터 모델개발하기		1. 성과변수 2. 파생변수 3. 요약변수 4. 변수리코딩	

난이도	상	[문항13] 주어진 데이터셋을 한 번에 모델에 넣어 분석하지 않고, 나누어 분석하는 것과 관련한 설명 중 성격이 다른 것은?	
	빅데이터 모델평가 검증하기	<ol style="list-style-type: none"> <li>1. 데이터분할</li> <li>2. 교차검증</li> <li>3. 훈련데이터</li> <li>4. 반복측정</li> </ol>	
난이도	하	[문항14] 분류예측에서 데이터셋을 3개로 나누었을 때의 명칭과 거리가 먼 것은?	
	빅데이터 모델평가 검증하기	<ol style="list-style-type: none"> <li>1. verification data</li> <li>2. test data</li> <li>3. training data</li> <li>4. target data</li> </ol>	
난이도	하	[문항15] 데이터분석 모델 개발과 분석, 운영에 적합한 방법론이 아닌 것은?	
	빅데이터모델 운영방안 마련하기	<ol style="list-style-type: none"> <li>1. agile 방법론</li> <li>2. water-fall 방법론</li> <li>3. spiral 방법론</li> <li>4. prototype 방법론</li> </ol>	
난이도	중	[문항16] 분류예측 모델의 성능을 평가할 때 여러 평가기준을 대표할 수 있는 기준으로 사용할 수 있는 것은?	
	빅데이터모델 운영방안 마련하기	<ol style="list-style-type: none"> <li>1. 진양성률</li> <li>2. 민감도</li> <li>3. ROC 커브</li> <li>4. 정밀도</li> </ol>	

난이도	하	[문항17] 주어진 데이터셋을 요약해서 개괄적인 데이터의 특성을 파악하는 분석과 관련성이 적은 항목은?	
기본통계 확인하기		1. 중심성 2. 분석성 3. 변동성 4. 정규성	
난이도	하	[문항18] 다음 중 성격이 다른 데이터 기본통계 확인기법은?	
기본통계 확인하기		1. 탐색적 데이터 분석 2. 기술통계 분석 3. 데이터 전처리 분석 4. 추론통계 분석	
난이도	중	[문항19] 독립변수와 종속변수 간의 관계를 조건에 따라서 다른 패턴이 나오도록 영향을 미치는 변수와 거리가 먼 것은?	
데이터분포 분석하기		1. 조절변수 2. 상황변수 3. 매개변수 4. 교란변수	
난이도	상	[문항20] 연속변수의 분포모양이 정규분포와 유사한지를 파악하는 기법 중의 하나는?	
데이터분포 분석하기		1. t.test() 2. lm() 3. shapiro.test() 4. anova()	

난이도	상	[문항21] 상관분석과 인과분석에 대한 설명중 거리가 먼 것은?	
변수간 관계 확인하기		1. 빅데이터시대는 인과분석보다는 상관분석을 많이 활용한다. 2. 인과분석은 원인과 결과변수간의 관계를 판단한다. 3. 상관성의 정도는 0~1사이의 값을 기준으로 판단한다. 4. 상관분석으로 업무에 먼저 활용하고, 차후에 인과분석까지 분석을 수행한다.	
난이도	하	[문항22] 다음 변수에 대한 명칭 중 성격이 다른 것은?	
변수간 관계 확인하기		1. 예측변수 2. 설명변수 3. 투입변수 4. 반응변수	
난이도	하	[문항23] 다음 중 데이터 정제작업시 직접적인 대상으로 보기 어려운 것은?	
데이터 정제하기		1. 이상치 2. 결측치 3. 잡음 4. 시그널	
난이도	중	[문항24] 예측모델에 입력하는 변수들의 단위나 척도, 범위, 규모 등이 달라 이를 서로 비슷한 수준으로 맞추어 주는 작업은?	
데이터 정제하기		1. 스케줄링(scheduling) 2. 스케일링(scaling) 3. 스무딩(smoothing) 4. 스캐닝(scanning)	

난이도	하	[문항25] 다음 중 사용자가 직접 분석용 데이터에 접근해 다양한 분석을 직접 수행하는 것과 거리가 먼 것은?	
	데이터변화 적재하기	<ol style="list-style-type: none"> <li>1. OLTP</li> <li>2. OLAP</li> <li>3. Data Warehouse</li> <li>4. Data Mart</li> </ol>	
난이도	상	[문항26] 모델링을 통해 도출된 분석모델에 맞추어 분석대상 데이터의 잡음을 제거하고 모델에 적합한 데이터를 준비해 모델에 투입하는 것을 무엇이라고 하는가?	
	데이터변화 적재하기	<ol style="list-style-type: none"> <li>1. 모델 training</li> <li>2. 모델 loading</li> <li>3. 모델 fitting</li> <li>4. 모델 slicing</li> </ol>	
난이도	중	[문항27] 데이터 무작위 추출시 언제든지 동일한 데이터셋이 추출되도록 사전에 설정해 주는 작업은?	
	분석용 데이터 검증하기	<ol style="list-style-type: none"> <li>1. sample()</li> <li>2. random()</li> <li>3. rnorm()</li> <li>4. set.seed()</li> </ol>	
난이도	하	[문항28] R에서 사용하는 분석용 데이터의 객체유형에 속하지 않은 것은?	
	분석용 데이터 검증하기	<ol style="list-style-type: none"> <li>1. 데이터프레임</li> <li>2. 딕셔너리</li> <li>3. 매트릭스</li> <li>4. 리스트</li> </ol>	

난이도	하	[문항29] 커피를 구매하는 고객이 비스킷을 같이 구매한다는 패턴을 찾는 분석기법은?	
머신러닝 수행방법 계획하기		1. 연관분석 2. 군집분석 3. 상관분석 4. 교차분석	
난이도	중	[문항30] 개체간 유사성을 기준으로 그룹핑해주는 유형화 분석과 거리가 먼 것은?	
머신러닝 수행방법 계획하기		1. 비계층적 군집 분석 2. 다차원척도 분석 3. 주성분 분석 4. 계층적군집 분석	
난이도	하	[문항31] 예측모델이 주어진 데이터셋의 변화패턴에 어느 정도 적합(fitting) 되었는지를 나타내는 개념과 거리가 먼 것은?	
데이터터셋 분할하기		1. 과대적합 2. 부분적합 3. 적정적합 4. 과소적합	
난이도	상	[문항32] 예측모델 규칙 도출시 훈련데이터를 다루는 방법유형과 거리가 먼 것은?	
데이터터셋 분할하기		1. 잭나이프 기법 2. 교차검증 기법 3. 부스팅 기법 4. 증화추출 기법	



난이도	하	[문항33] 다음 중 성격이 다른 지도학습 분석기법은?	
지도학습 모델적용하기		1. 의사결정나무분석 2. 로지스틱회귀분석 3. 서포트벡터머신 4. 회귀분석	
난이도	하	[문항34] 수치예측 분석에서 독립변수간의 상호관련성으로 인해서 종속변수의 예측시 심각한 오류를 초래할 수 있는 현상은?	
지도학습 모델적용하기		1. 교란효과 2. 정규성 3. 다중공선성 4. 선형성	
난이도	하	[문항35] 다음 중 성격이 다른 자율학습 분석기법은?	
자율학습 모델적용하기		1. 주성분분석 2. 연관분석 3. 군집분석 4. 다차원척도법	
난이도	중	[문항36] 연관분석시 사용하는 규칙도출 기준으로 거리가 먼 것은?	
자율학습 모델적용하기		1. 향상도 2. 추천도 3. 지지도 4. 신뢰도	

난이도	하	[문항37] 분류예측모형의 성능평가시 사용되는 평가메트릭에 대한 보기 중 성격이 다른 것은?	
모델성능 평가하기		1. 특이도 2. 정확도 3. 정밀도 4. 난이도	
난이도	중	[문항38] 수치예측모형의 성능평가시 사용되는 평가메트릭에 대한 보기 중 성격이 다른 것은?	
모델성능 평가하기		1. $R^2$ (설명력) 2. 평균절대오차(MAE) 3. 평균제곱오차(MSE) 4. 유의확률(p-value)	
난이도	상	[문항39] 도출된 예측모델을 업무에 적용했을 때 바람직한 조치는?	
학습결과 적용하기		1. 예측모델도 ABT를 통해서 실제 효과발생 가능성에 대해서 검정한다. 2. 예측모델의 성능이 높으면 1년 정도 사용해 효과를 극대화 한다. 3. 한번 사용한 예측모델은 폐기하고 항상 새로운 모델을 개발해야 한다. 4. 각 예측모델은 고유문제해결을 위한 것이므로 타기법과 결합하면 안 된다.	
난이도	중	[문항40] PPT로 분석결과 자료를 작성할 때 전체내용을 대표하는 핵심 메시지를 2줄 정도 상단에 작성해 배치하는 것은?	
학습결과 적용하기		1. 바디메시지 2. 거버닝메시지 3. 코어메시지 4. 전략메시지	