

2001010506_19v3	통계 기반 데이터 분석
-----------------	--------------

진 단 영 역	진 단 문 항	전혀 그렇지 않다.	그렇지 않다는 편이다.	보통 이다.	다소 그런 편이다.	매우 그렇다.
가설 설정하 기	1. 나는 빅데이터 분석 프로젝트의 목적 달성을 위하여 비즈니스 도메인과 수집된 데이터를 파악할 수 있다.	①	②	③	④	⑤
	2. 나는 프로젝트의 목적에 적합한 모델 개발을 위하여 데이터의 사전탐색을 통하여 가설을 설정할 수 있다.	①	②	③	④	⑤
	3. 나는 분석 모델의 가설 평가를 위한 통계치의 기준을 설정하고 적용할 수 있다.	①	②	③	④	⑤
통계 모델 개발하기	1. 나는 설정된 가설을 검증하기 위하여 분석모형을 설계 할 수 있다.	①	②	③	④	⑤
	2. 나는 모델을 완성도를 평가하기 위한 성능평가 기준을 설정할 수 있다.	①	②	③	④	⑤
	3. 나는 모델에 적용하는 독립변수와 종속변수의 데이터 유형에 따라 다양한 분석기법을 적용할 수 있다.	①	②	③	④	⑤
	4. 나는 주어진 데이터 변수 추가하여 모델에서 필요로 되어지는 파생변수(Drived Variable)를 생성하여 모델링에 활용할 수 있다.	①	②	③	④	⑤
	5. 나는 단일 분석기법이 아닌 앙상블(Ensemble) 기법을 적용하여 모델의 품질목표를 제고할 수 있다	①	②	③	④	⑤
	6. 나는 데이터 분석 모델링의 원활한 수행과 결과보고를 위하여 데이터 시각화를 사용할 수 있다.	①	②	③	④	⑤
통계 모델 평가 검증하 기	1. 나는 구현된 모델은 모델 운영의 안정성과 적합성을 평가할 수 있다.	①	②	③	④	⑤
	2. 나는 설정되어 있는 모델의 평가 절차 및 기준에 따라 구현된 모델을 평가할 수 있다.	①	②	③	④	⑤
	3. 나는 모델 평가를 위한 데이터셋은 훈련 데이터셋과 구분되어 관리되어 평가의 객관성을 확보할 수 있다.	①	②	③	④	⑤
	4. 나는 모델 평가의 전과정을 절차에 따라 진행되었음을 증적으로 확보해야 하고 모델 평가결과를 리포팅할 수 있다.	①	②	③	④	⑤

[진단결과]

영역	문항 수	점 수	점수 ÷ 문항 수
가설 설정하기	3		
통계 모델 개발하기	6		
통계 모델 평가 검증하기	4		
합계	13		

※ 자신의 점수를 문항 수로 나눈 값이 ‘3점’ 이하에 해당하는 영역은 업무를 성공적으로 수행하는데 요구되는 능력이 부족한 것으로 교육훈련이나 학습을 통한 능력개발이 필요함

진 단 영 역	진 단 문 항	전혀 그렇지 않다.	그렇지 않다는 편이다.	보통 이다.	다소 그런 편이다.	매우 그렇다.
머신러닝 수 행방법 계획 하기	1. 나는 비즈니스 목표와 전략 및 방침에 따라 머신러닝 기반의 데이터 분석을 도입하고 전체 프로세스를 관리할 수 있다.	①	②	③	④	⑤
	2. 나는 통계학 기반의 데이터분석과 머신러닝 기반의 데이터분석의 차이와 활용목적을 이해하고 활용 목적에 맞추어 머신러닝 기법 적용 필요성을 판단할 수 있다.	①	②	③	④	⑤
	3. 나는 해결하고자 하는 이슈에 따라 데이터구조의 설명과 패턴화에 적합한 머신러닝 기법을 선정하고 적용 절차를 계획할 수 있다.	①	②	③	④	⑤
	4. 나는 수립된 계획과 절차에 맞추어 머신러닝 기반의 데이터분석이 효과적으로 이루어질 수 있도록 수행계획을 문서화하고, 구성원에게 공유할 수 있다.	①	②	③	④	⑤
데이터셋 분 할하기	1. 나는 과적합과 일반화의 의미 및 파급효과를 이해하고, 이의 해결을 위한 데이터셋 분할을 설계할 수 있다.	①	②	③	④	⑤
	2. 나는 분석하고자 하는 목적 및 데이터셋 특성에 따라 머신러닝 기법 적용을 위한 훈련 데이터 셋과 테스트 데이터 셋 분할 기준을 판단할 수 있다.	①	②	③	④	⑤
	3. 나는 해결하고자 하는 이슈와 적용할 기법에 따라 교차검증 필요성을 판단하여, 훈련 데이터셋과 검증 데이터 셋을 분할하고, 적합한 교차 검증 K값을 결정할 수 있다.	①	②	③	④	⑤
	4. 나는 데이터의 속성을 고려한 예측 및 분류 목적 변수의 분포를 고려하여 데이터 셋을 분할하고 샘플링할 수 있다.	①	②	③	④	⑤
	5. 나는 데이터셋 분할을 위한 다양한 샘플링 방법의 차이를 분석하여 샘플링 방법을 적용할 수 있다.	①	②	③	④	⑤
지도학습 모 델 적용하기	1. 나는 수립된 분석 계획에 따라 실제로 정확한 분류나 예측모형화를 위해 적합한 머신러닝 기법을 적용할 수 있다.	①	②	③	④	⑤
	2. 나는 명목형 및 이산형 목적변수에 주어진 경우, 이의 문제 해결 위해 다양한 분류 모델을 비교해보고 최적의 분류모델을 선정하여 적용할 수 있다.	①	②	③	④	⑤
	3. 나는 연속형 목적변수가 주어진 경우, 이의 문제 해결 위해 다양한 수치 예측 모델을 비교해 보고 최적의 수치예측모델을 선정하여 적용할 수 있다.	①	②	③	④	⑤

지도학습 모델 적용하기	4. 나는 주어진 상황과 데이터 특성에 따라 필요시, 여러 알고리즘이나 기법을 융합한 앙상블 모형을 적용하거나 새로운 방법론을 개발할 수 있다.	①	②	③	④	⑤
자율학습 모델 적용하기	1. 나는 수립된 분석 계획에 따라 필요한 자율학습 머신러닝 기법을 적용할 수 있다.	①	②	③	④	⑤
	2. 나는 개체수 축약이나 대상자 세분화 등의 분석 목적이 주어진 경우, 이의 문제 해결 위해 다양한 군집화 기법을 적용해보고 최적의 군집화 기법을 선정하여 적용할 수 있다.	①	②	③	④	⑤
	3. 나는 변수 축약이나 차원 축소가 필요한 경우, 이의 문제 해결 위해 주성분 분석 혹은 SVD 등의 차원 축소 기법을 선정하여 적용할 수 있다.	①	②	③	④	⑤
	4. 나는 대량의 트랜잭션 데이터에서의 규칙성이나 패턴화 도출의 목적이 주어진 경우, 이의 문제 해결 위해 다양한 연관 규칙 알고리즘을 비교하고 적용할 수 있다.	①	②	③	④	⑤
모델 성능 평가하기	1. 나는 분석하고자 하는 목적 및 데이터셋 특성에 따라 필요한 모델 성능 평가 방법을 선정하고 비교할 수 있다.	①	②	③	④	⑤
	2. 나는 머신러닝 결과 모델의 평가를 위해 필요한 평가 메트릭을 데이터로부터 실제로 계산해 내고 그 의미를 해석할 수 있다.	①	②	③	④	⑤
	3. 나는 모델 성능 평가 지표 도출 결과에 따라, 최적 모델 도출 위해 주어진 모델의 성능 향상을 시행할 수 있다.	①	②	③	④	⑤
학습결과 적용하기	1. 나는 머신러닝 기반 데이터 분석결과의 실제 비즈니스 목표와 전략과의 부합성을 판단하고 평가할 수 있다.	①	②	③	④	⑤
	2. 나는 조직 목표를 달성하기 위해 데이터 처리 절차와 모델 구축 결과 및 평가결과를 문서화하고 조직내 여러 계층에게 커뮤니케이션 할 수 있다.	①	②	③	④	⑤
	3. 나는 예측 및 새로운 데이터 평가 결과를 실제 현업 업무나 상용 서비스에 구현하기 위해 프로세스화 작업을 수행할 수 있다.	①	②	③	④	⑤
	4. 나는 현업 적용 결과의 개선과 새로운 고객 요구에 대응하기 위해 반복되는 Feedback Cycle을 시행하고 관리할 수 있다.	①	②	③	④	⑤

[진단결과]

영역	문항 수	점 수	점수 ÷ 문항 수
머신러닝 수행방법 계획하기	4		
데이터셋 분할하기	5		
지도학습 모델 적용하기	4		
자율학습 모델 적용하기	4		
모델 성능 평가하기	3		
학습결과 적용하기	4		
합계	24		

※ 자신의 점수를 문항 수로 나눈 값이 ‘3점’ 이하에 해당하는 영역은 업무를 성공적으로 수행하는데 요구되는 능력이 부족한 것으로 교육훈련이나 학습을 통한 능력개발이 필요함

진 단 영 역	진 단 문 항	전혀 그렇지 않다.	그렇지 않다는 편이다.	보통 이다.	다소 그런 편이다.	매우 그렇다.
텍스트 데이터 변환하기	1. 나는 주어진 텍스트 데이터를 단어와 문법적 특성에 맞추어 명사, 동사, 꾸밈어, 조사 등의 형태소로 분리할 수 있다.	①	②	③	④	⑤
	2. 나는 분석하고자 하는 목적 및 데이터 특성에 따라 단어들에 대한 가용어 및 불용어 처리를 통해 필요 키워드를 추출할 수 있다.	①	②	③	④	⑤
	3. 나는 텍스트 데이터의 정제, 변환을 통해 말뭉치를 작성할 수 있다.	①	②	③	④	⑤
	4. 나는 정제된 말뭉치로부터 단어와 문서의 관계를 표현하기 위해 문서-단어 행렬 혹은 단어-문서 행렬을 작성할 수 있다.	①	②	③	④	⑤
	5. 나는 문서-단어 행렬, 혹은 단어-문서 행렬로부터 단어의 빈도와 연관성 등의 기본 집계를 수행할 수 있다.	①	②	③	④	⑤
텍스트 데이터 분석 수행방법 계획하기	1. 나는 조직 목표와 전략 및 방침에 따라 텍스트 분석 필요성을 판단하여 텍스트 분석 전체 프로세스를 기획하고 관리할 수 있다.	①	②	③	④	⑤
	2. 나는 분석 대상 텍스트 데이터의 특성, 문맥 및 목적에 적합한 텍스트 분석 기법과 절차를 판단할 수 있다.	①	②	③	④	⑤
	3. 나는 텍스트 데이터 분석을 통해 도출하고자 하는 최종 결과물에 적합한 분석 방법론과 산출물을 사전 계획할 수 있다.	①	②	③	④	⑤
	4. 나는 수립된 계획과 절차에 맞추어 텍스트 데이터 분석이 효과적으로 이루어질 수 있도록 수행계획을 문서화하고 구성원에게 공유할 수 있다.	①	②	③	④	⑤
텍스트 데이터 분류 결과 분석하기	1. 나는 수립된 분석 계획에 따라 필요한 텍스트 데이터 분류 방법을 적용할 수 있다.	①	②	③	④	⑤
	2. 나는 분류된 항목에 따라 집계, 판별, 예측 등의 텍스트 데이터 분류분석 작업을 수행할 수 있다.	①	②	③	④	⑤
	3. 나는 분석 결과에서 분석의 목적 및 해결하고자 하는 이슈에 적합한 의미나 인사이트를 도출할 수 있다.	①	②	③	④	⑤
	4. 나는 분석 결과를 커뮤니케이션하기 위해 시각화하고 문서화할 수 있다.	①	②	③	④	⑤
정형 데이터 결합 분석 수행하기	1. 나는 조직의 이슈와 당면과제에 따라 필요한 정형-비정형 데이터 분석 모델을 기획할 수 있다.	①	②	③	④	⑤

정형 데이터 결합 분석 수행하기	2. 나는 분석 목표 및 계획에 따라 필요한 정형 데이터를 추출하고 비정형 텍스트 데이터와 결합할 수 있다.	①	②	③	④	⑤
	3. 나는 분석 목표 달성을 위해 필요한 분석기법과 기술을 적용하여 결과를 도출할 수 있다.	①	②	③	④	⑤
	4. 나는 조직 목표를 달성하기 위해 데이터 처리 절차와 모델 구축 결과 및 평가결과를 문서화하고 조직내 여러 계층에게 커뮤니케이션할 수 있다.	①	②	③	④	⑤
	5. 나는 예측 및 새로운 데이터 평가 결과를 실제 현업 업무나 상용 서비스에 구현하기 위해 프로세스화 작업을 수행할 수 있다.	①	②	③	④	⑤

[진단결과]

영역	문항 수	점 수	점수 ÷ 문항 수
텍스트 데이터 변환하기	5		
텍스트 데이터 분석 수행방법 계획하기	4		
텍스트 데이터 분류 결과 분석하기	4		
정형 데이터 결합 분석 수행하기	5		
합계	18		

※ 자신의 점수를 문항 수로 나눈 값이 ‘3점’ 이하에 해당하는 영역은 업무를 성공적으로 수행하는데 요구되는 능력이 부족한 것으로 교육훈련이나 학습을 통한 능력개발이 필요함

진 단 영 역	진 단 문 항	전혀 그렇지 않다.	그렇지 않다는 편이다.	보통 이다.	다소 그런 편이다.	매우 그렇다.
분석 결과 스토리텔링 하기	1. 나는 시각화 대상이 되는 다양한 형태의 데이터와 분석 결과를 이용하여 시각화를 위한 표현 규칙과 패턴을 도출하고 정의할 수 있다.	①	②	③	④	⑤
	2. 나는 정의된 규칙과 패턴에 기반하여 시각화 활용과 사용자 가치 관점에서의 시각화 요건을 정의할 수 있다.	①	②	③	④	⑤
	3. 나는 시각화 요건에 따른 사용자 시나리오를 작성하고, 분석 결과를 명확히 전달하기 위한 핵심 요소 및 그래픽 요소를 발굴하여 스토리보드를 기획할 수 있다.	①	②	③	④	⑤
분석 정보 시각화하기	1. 나는 계획된 스토리 보드에 따라 시각화를 구현하기 위한 시각화 도구와 기술을 선택하고, 정보 시각화를 위한 접근 방법을 결정할 수 있다.	①	②	③	④	⑤
	2. 나는 시각화 도구에서 제공하는 시각화 표현 요소를 이용하여 분석의 핵심 내용을 반영할 수 있도록 분석 정보를 구조적으로 시각화할 수 있다.	①	②	③	④	⑤
	3. 나는 단순하고 명료한 메시지 전달을 위해 시각적 형태와 모양이 갖추어 지도록 시각화 과정을 반복적으로 수행할 수 있다.	①	②	③	④	⑤
정보 시각화 구현하기	1. 나는 추가적 정보를 이용하여 그래프를 보정하고 시각 표현을 극대화하도록 실험을 통하여 시각화 구현을 완성할 수 있다.	①	②	③	④	⑤
	2. 나는 정보 표현을 위한 그래픽 요소를 반영하여 전달 요소를 강조할 수 있다.	①	②	③	④	⑤
	3. 나는 시각화 결과물의 품질 향상을 위해 타이포그래피, 그리드, 색상, 아이소타이프 등 그래픽 디자인 원리를 적용할 수 있다.	①	②	③	④	⑤
	4. 나는 정보 사용자의 행동이나 조작에 따른 반응, 감각의 확장, 정보 시각화의 변화 등으로 정보와 정보 사용자간의 관계 확장을 위해 사용자 참여를 유도하는 인터렉션 기능을 구현할 수 있다.	①	②	③	④	⑤
	5. 나는 최종 디자인된 시각화 결과물이 원래 의도와 목적에 맞게 데이터와 비주얼 스토리텔링이 적절하게 구현 되었는지 확인할 수 있다.	①	②	③	④	⑤

[진단결과]

영역	문항 수	점 수	점수 ÷ 문항 수
분석 결과 스토리텔링하기	3		
분석 정보 시각화하기	3		
정보 시각화 구현하기	5		
합계	11		

※ 자신의 점수를 문항 수로 나눈 값이 ‘3점’ 이하에 해당하는 영역은 업무를 성공적으로 수행하는데 요구되는 능력이 부족한 것으로 교육훈련이나 학습을 통한 능력개발이 필요함

2001010510_19v3	분석용 데이터 구축
-----------------	------------

진 단 영 역	진 단 문 항	전혀 그렇지 않다.	그렇지 않다는 편이다.	보통 이다.	다소 그런 편이다.	매우 그렇다.
데이터 정제 하기	1. 나는 빅데이터 분석에 필요한 데이터를 생성 주체 및 활용 유형 등에 따라 식별하여 분류할 수 있다.	①	②	③	④	⑤
	2. 나는 효율적인 분석을 위하여 식별된 데이터를 분석에 적합한 품질의 데이터로 변환할 수 있다.	①	②	③	④	⑤
	3. 나는 분석용 데이터셋을 이용하여 데이터의 오류, 결측치 등을 파악하고 분석요구사항에 맞도록 정제할 수 있다.	①	②	③	④	⑤
데이터 변환 적재하기	1. 나는 분석의 효율성을 위하여 정제된 데이터셋을 분석 도구에 맞는 데이터 형태로 변경할 수 있다.	①	②	③	④	⑤
	2. 나는 데이터웨어하우스(Data Warehouse)에 저장된 데이터를 분석용 데이터 마트(Data Mart)나 샌드박스(Sandbox)에 적재 할 수 있다.	①	②	③	④	⑤
	3. 나는 데이터의 용량과 데이터 처리의 실시간성을 판단하여 적재가 가능한 상태로 편성할 수 있다.	①	②	③	④	⑤
	4. 나는 빅데이터 분석에 용이한 형태의 데이터셋을 편성 및 적재하고 적재 전후의 데이터 완전성을 확보할 수 있다.	①	②	③	④	⑤
데이터 검증 하기	1. 나는 정제된 분석용 데이터가 올바르게 적재되었는지 확인할 수 있다.	①	②	③	④	⑤
	2. 나는 빅데이터의 데이터 품질에 대하여 정확성 및 충실성에 맞게 분석용 데이터셋으로 적재할 수 있다.	①	②	③	④	⑤
	3. 나는 적재된 데이터의 신뢰성이 확보되어 있는지 파악할 수 있다.	①	②	③	④	⑤
	4. 나는 데이터 사전 처리가 개인정보보호 등의 규정에 위반되지 않게 처리되어 적재되었는지 확인할 수 있다.	①	②	③	④	⑤
	5. 나는 정제되어 적재된 데이터의 정합성에 문제가 있을 경우 데이터 정합성 개선개화를 수립할 수 있다.	①	②	③	④	⑤
	6. 나는 분석 모델링을 위한 훈련 데이터, 테스트 데이터 분할 계획을 수립하고 실 데이터 검증을 위한 데이터 셋 준비상태를 확인할 수 있다.	①	②	③	④	⑤

[진단결과]

영역	문항 수	점 수	점수 ÷ 문항 수
데이터 정제하기	3		
데이터 변환 적재하기	4		
데이터 검증하기	6		
합계	13		

※ 자신의 점수를 문항 수로 나눈 값이 ‘3점’ 이하에 해당하는 영역은 업무를 성공적으로 수행하는데 요구되는 능력이 부족한 것으로 교육훈련이나 학습을 통한 능력개발이 필요함

2001010511_19v3	탐색적 데이터 분석
-----------------	------------

진 단 영 역	진 단 문 항	전혀 그렇지 않다.	그렇지 않다는 편이다.	보통 이다.	다소 그런 편이다.	매우 그렇다.
기본통계 확 인하기	1. 나는 탐색적 데이터 분석을 위하여 데이터 표본을 추출하는 기법을 선정할 수 있다.	①	②	③	④	⑤
	2. 나는 표본조사 및 실험을 통하여 추출된 원소나 실험 단위로부터 주어진 목적에 적합하도록 자료를 측정할 수 있다.	①	②	③	④	⑤
	3. 나는 수집된 자료를 정리, 요약하기 위하여 기초적인 통계 기법을 사용할 수 있다.	①	②	③	④	⑤
데이터 분포 분석하기	1. 나는 분석 목적과 요건, 데이터 특성을 기반으로 적합한 데이터 분석 기법을 선정할 수 있다.	①	②	③	④	⑤
	2. 나는 선정된 데이터 분석기법을 기준으로 분석 모형을 설계할 수 있다.	①	②	③	④	⑤
	3. 나는 설계한 분석 모형을 기준으로 유의성을 분석해 높은 유의성을 보유한 변수들을 식별할 수 있다.	①	②	③	④	⑤
	4. 나는 높은 유의성을 보유한 최소한의 변수들로 분석 모형들을 구축할 수 있다.	①	②	③	④	⑤
변수간 관계 확인하기	1. 나는 탐색적 분석을 통하여 준비된 데이터의 가설 적합성과 충분성을 사전 검증할 수 있다.	①	②	③	④	⑤
	2. 나는 변수의 유의성 검증 후 유의성이 높은 최소한의 변수들로 분석 모형을 검증할 수 있다.	①	②	③	④	⑤
	3. 나는 시뮬레이션을 통해 기 수립된 분석 모형의 타당성과 적합성을 판단해 반복적으로 보정할 수 있다.	①	②	③	④	⑤

[진단결과]

영 역	문항 수	점 수	점수 ÷ 문항 수
기본통계 확인하기	3		
데이터 분포 분석하기	4		
변수간 관계 확인하기	3		
합계	10		

※ 자신의 점수를 문항 수로 나눈 값이 ‘3점’ 이하에 해당하는 영역은 업무를 성공적으로 수행하는데 요구되는 능력이 부족한 것으로 교육훈련이나 학습을 통한 능력개발이 필요함