

분류번호 :	2001070309_23v1
능력단위 명칭 :	인공지능 데이터 품질 검증
능력단위 정의 :	인공지능 데이터 품질 검증이란 데이터 품질 기준을 정의, 품질 검증 프로세스 수립을 통해 인공지능 품질을 개선하는 능력이다.

능 력 단 위 요 소	수 행 준 거
2001070309_23v1.1 데이터 품질 기준 정의하기	<p>1.1 검증된 모델의 요구사항, 도메인별 표준, 전처리된 데이터, 특징을 검토하여 프로젝트에 가장 관련성 높은 데이터의 품질 특성을 도출할 수 있다.</p> <p>1.2 데이터 품질 평가를 위한 품질 특성별 허용 수준과 목표를 설정할 수 있다.</p> <p>1.3 품질특성을 측정하기 위한 방법을 설계하고 문서화할 수 있다.</p> <p>【지식】</p> <ul style="list-style-type: none"> <li>• 데이터 품질 특성</li> <li>• 특정 분야에 대한 데이터 품질 특성에 대한 산업 표준</li> <li>• 데이터 품질을 평가할 수 있는 정량적, 정성적 방법론</li> <li>• 인공지능 윤리 가이드북</li> <li>• 인공지능 FATE 가이드북</li> </ul> <p>【기술】</p> <ul style="list-style-type: none"> <li>• 프로젝트 목표에 따라 적절한 데이터 품질 기준을 제작하는 능력</li> <li>• 정의된 데이터 품질 특성, 기준, 평가 방법을 명확하게 문서화하는 능력</li> <li>• 예상되는 데이터 품질의 문제를 식별할 수 있는 능력</li> </ul> <p>【태도】</p> <ul style="list-style-type: none"> <li>• 데이터 품질 특성 기준에 대해 모델의 목표를 반영하려는 적극적인 자세</li> <li>• 변화하는 목표에 따라 품질특성과 기준을 개선하고자 하는 노력</li> <li>• 인공지능 윤리기준을 준수하려는 노력</li> </ul>
2001070309_23v1.2 데이터 품질 검증 프로세스 구축하기	<p>2.1 인공지능 데이터 수집 단계별 검증할 데이터 품질 특성을 확정할 수 있다.</p> <p>2.2 확정된 데이터 품질을 검증하기 위한 도구, 기술, 방법론을 선택할 수 있다.</p> <p>2.3 데이터 품질 검증을 위한 프로세스와 계획을 수립할 수 있다.</p> <p>【지식】</p> <ul style="list-style-type: none"> <li>• 데이터 품질 검증 표준 프로세스</li> <li>• 인공지능 모델링 프로세스</li> </ul> <p>【기술】</p> <ul style="list-style-type: none"> <li>• 모델링 단계별 데이터 품질 문제의 잠재적 영향을 예측하는 능력</li> <li>• 도구, 일정, 조직의 역할을 고려하여 검증 계획을 수립하는 능력</li> <li>• 변화하는 프로젝트 요구사항 변화에 대응하는 능력</li> </ul> <p>【태도】</p> <ul style="list-style-type: none"> <li>• 검증 프로세스를 최적화하여 데이터 품질문제를 사전에 예방하려는 능동적 자세</li> <li>• 검증 프로세스를 지속적으로 개선하고자 하는 노력</li> </ul>
2001070309_23v1.3 데이터 품질 검증 실행하기	<p>3.1 데이터 품질 점검 프로세스와 계획에 따라 데이터 품질 평가를 수행할 수 있다.</p> <p>3.2 수행 결과에 따른 인공지능 모델의 공정성과 신뢰성의 영향도를 파악할 수 있다.</p> <p>3.3 파악된 데이터의 품질 평가 결과에 대해 문서화할 수 있다.</p>

<p>2001070309_23v13</p> <p>데이터 품질 검증 실행하기</p>	<p>【지식】</p> <ul style="list-style-type: none"> <li>• 데이터 프로파일링 도구</li> <li>• 데이터 유효성 검사 도구</li> <li>• 데이터 정리 도구</li> <li>• 데이터 통합 도구</li> </ul> <p>【기술】</p> <ul style="list-style-type: none"> <li>• 선택한 데이터품질 검증 도구를 활용할 수 있는 능력</li> <li>• 데이터 품질 평가 결과를 문서화하는 기술</li> </ul> <p>【태도】</p> <ul style="list-style-type: none"> <li>• 평가 결과를 숨김없이 이해관계자에게 전달하려는 의지</li> <li>• 이해관계자와 검증된 결과에 대해 수궁하고 함께 해결하고자 하는 노력</li> </ul>
<p>2001070309_23v14</p> <p>데이터 품질 개선하기</p>	<p>4.1 문서화된 데이터의 검증 결과를 분석하여 원인을 도출할 수 있다.</p> <p>4.2 도출된 원인에 대한 해결책을 도출할 수 있다.</p> <p>4.3 데이터 품질 문제에 대한 원인과 해결책을 문서화할 수 있다.</p> <p>4.4 데이터 품질 문제가 발생하지 않도록 품질 검증 결과를 개선할 수 있다.</p> <p>【지식】</p> <ul style="list-style-type: none"> <li>• 데이터 품질 문제 솔루션</li> <li>• 데이터 품질 문제의 원인 도출 방법론</li> </ul> <p>【기술】</p> <ul style="list-style-type: none"> <li>• 통계적 데이터 분석 기술</li> <li>• 데이터 분석 결과에 대한 시각화 기술</li> <li>• 데이터 품질 문제에 대한 원인을 검증하는 기술</li> </ul> <p>【태도】</p> <ul style="list-style-type: none"> <li>• 데이터 품질 문제에 대한 원인을 정확히 파악하고자 하는 노력</li> <li>• 데이터 품질을 지속적으로 향상시키고자 하는 태도</li> </ul>

## □ 적용범위 및 작업상황

### 고려사항

- 이 능력단위는 데이터 품질 기준을 정의하고, 품질 검증 프로세스 수립을 통해 인공지능 학습 데이터를 검증함으로써 모델의 공정성과 신뢰성을 확보하는 업무에 적용한다.
- 데이터 품질을 평가할 수 있는 정량적, 정성적 방법론에는, 데이터 품질 프레임워크 (Total Data Quality Management (TDQM) and Data Quality Assessment Framework (DQAF)), 데이터프로파일링, 데이터 오디팅, 데이터품질 스코어 카드 등이 있다.
- 데이터 품질 특성에는 모델 학습에 사용되는 데이터의 품질을 평가하는 데 도움이 되는 정확성, 완전성, 일관성, 적시성, 대표성 및 관련성 등이 포함된다.
- 데이터 품질 검증을 위한 거버넌스는 데이터 품질의 평가 및 개선을 관리하는 규칙, 정책 및 프로세스를 의미하며, 역할 및 책임, 정책, 프로세스, 도구, 보고체계 등이 포함된다.
- 데이터 품질 검증 실행 이후 이해관계자에게 전달하여 개선이 필요한 영역에 대한 피드백을 받아 조치의 우선순위를 결정한다.
- 데이터 프로파일링 도구는 데이터내의 구조를 분석하여 불일치, 오류, 누락된 값을 식별하는 도구로써 Talend, Trifacta, DataWrangler 등이 있다.
- 데이터 유효성 검사 도구는 정의된 규칙 또는 기준에 따라 데이터를 확인하여 정확성, 일관성 등을 파악하는 도구로써 Informatica Data Validation, Oracle Data Quality 등이 있다.
- 데이터 통합 도구는 통합된 데이터 셋의 일관성과 완전성, 정확성을 확인하는 도구로써 Apache NiFi, FME, Alteryx 등이 있다.
- 데이터 품질 검증 표준 프로세스로는 ISO 8000, ISO 25012, DAMA-DMBOK(Data Management Body of Knowledge)등이 있다.
- 데이터 품질 문제에 대한 해결방법으로는 데이터클린징(Data Cleansing) 데이터프로파일링(Data Profiling), ETL (Extract, Transform, Load) 방법 등이 있다. 데이터 품질 문제의 원인을 도출하는 방법론으로는 데이터 검사(Data Examination), 근본원인분석(Root Cause Analysis) 방법 등이 있다.
- 데이터 품질 검증 방법에는 데이터검사(Data Examination), 근본원인분석(Root Cause Analysis) 등이 있다.

### 자료 및 관련 서류

- 인공지능 데이터 품질 정의서
- 인공지능 데이터 품질 검증 프로세스 정의서
- 인공지능 데이터 품질 검증 결과 보고서

### 장비 및 도구

- 컴퓨터, 문서작성 프로그램
- 데이터 품질 검증 자동화 소프트웨어 또는 패키지
- 데이터 시각화 프로그램

### 재료

- 해당사항 없음

## □ 평가지침

### 권장평가방법

- 평가자는 능력단위 인공지능 데이터 품질 검증의 수행준거에 제시되어 있는 내용을 평가하기 위해 이론과 실기를 나누어 평가하거나 종합적인 결과물의 평가 등 다양한 평가 방법을 사용할 수 있다.
- 평가자는 다음 사항을 평가해야 한다.

권 장 평 가 방 법	평 가 유 형	
	과 정 평 가	결 과 평 가
A.포트폴리오		
B.문제해결 시나리오	V	
C.서술형시험	V	V
D.논술형시험		V
E.사례연구	V	
F.평가자 질문		V
G.평가자 체크리스트	V	V
H.피평가자 체크리스트		
I.일지/저널		
J.역할연기		
K.구두발표	V	V
L.작업장평가		
M.기타		

### 평가지 고려사항

- 수행준거에 제시되어 있는 내용을 성공적으로 수행할 수 있는지를 평가해야 한다.
  - 평가자는 다음 사항을 평가해야 한다.
- 인공지능 데이터 품질 기준 정의
  - 인공지능 데이터 품질 프로세스 구축 능력
  - 인공지능 데이터 품질 검증 실행 능력
  - 인공지능 데이터 품질 개선 능력

□ 관련기초능력

순번	관 련 기 초 능 력	
	주 요 영 역	하 위 영 역
1	의사소통능력	경청 능력, 기초외국어 능력, 문서이해 능력, 문서작성 능력, 의사표현 능력
2	수리능력	기초연산 능력, 기초통계 능력, 도표분석 능력, 도표작성 능력
3	문제해결능력	문제처리 능력, 사고력
4	정보능력	정보처리 능력, 컴퓨터활용 능력
5	기술능력	기술선택 능력, 기술이해 능력, 기술적용 능력

□ 개발·개선 이력

구 분		내 용
직무명칭(능력단위명)		인공지능모델링(인공지능 데이터 품질 검증)
분류번호	기존	2001070309_23v1
	현재	2001070309_23v1
개발·개선연도	현재	2023
버전번호		v1
개발·개선기관	현재	
향후 보완 연도(예정)		-