

# AWS Certified Machine Learning – Specialty (MLS-C01) 시험 안내서

## 소개

AWS Certified Machine Learning – Specialty (MLS-C01) 시험은 인공지능/기계 학습(AI/ML) 개발 또는 데이터 과학 분야에서 역할을 수행하는 개인을 대상으로 합니다. 이 시험은 응시자가 AWS 클라우드를 사용하여 특정 비즈니스 문제에 대한 기계 학습 솔루션을 설계, 구축, 배포, 최적화, 훈련, 튜닝 및 유지 관리하는 능력을 검증합니다.

또한 이 시험에서는 응시자가 다음 태스크를 완료할 수 있는지도 확인합니다.

- 특정 비즈니스 문제에 대한 적절한 기계 학습 접근 방식 선택 및 정당성 증명
- 기계 학습 솔루션 구현에 적합한 AWS 서비스 식별
- 확장 가능하고, 비용 최적화되고, 신뢰할 수 있는 안전한 기계 학습 솔루션 설계 및 구현

## 대상 응시자 설명

대상 응시자는 AWS 클라우드에서 기계 학습 또는 딥 러닝 워크로드를 개발, 아키텍팅 및 실행한 실무 경험이 2년 이상 있어야 합니다.

## AWS 지식 권장 사항

대상 응시자는 다음과 같은 지식이 있어야 합니다.

- 기본 기계 학습 알고리즘을 직관적으로 표현할 수 있는 능력
- 기본적인 하이퍼파라미터 최적화 수행 경험
- 기계 학습 및 딥 러닝 프레임워크 경험
- 모델 훈련 모범 사례에 따라 수행할 수 있는 능력
- 배포 모범 사례에 따라 수행할 수 있는 능력
- 운영 모범 사례에 따라 수행할 수 있는 능력

### 대상 응시자가 갖추지 않아도 되는 것은 무엇입니까?

다음은 대상 응시자가 수행하지 않아도 될 것으로 예상되는 관련 작업 태스크의 목록입니다(전체 목록은 아님). 다음 항목은 시험 범위에 포함되지 않는 것으로 간주됩니다.

- 대규모 또는 복잡한 알고리즘 개발
- 대규모의 하이퍼파라미터 최적화
- 복잡한 수학적 증명 및 계산
- 고급 네트워킹 및 네트워크 설계
- 고급 데이터베이스, 보안 및 DevOps 개념
- Amazon EMR 을 위한 DevOps 관련 태스크

시험에서 다룰 수 있는 특정 도구 및 기술에 대한 자세한 목록과 범위 포함 여부가 표시된 AWS 서비스 목록은 부록을 참조하십시오.

## 시험 콘텐츠

### 응답 유형

이 시험의 문항은 두 가지 유형으로 제공됩니다.

- **선다형:** 정답 1 개와 오답 3 개(정답 이외의 답)가 있습니다.
- **다답형:** 5 개 이상의 응답 항목 중에 2 개 이상의 정답이 있습니다.

문장을 가장 잘 완성하거나 질문에 대한 답으로 가장 적합한 응답을 하나 이상 선택합니다. 정답 이외의 답 또는 오답은 지식이나 기술이 부족한 응시자가 선택할 가능성이 큰 응답 항목입니다. 정답 이외의 답은 일반적으로 콘텐츠 영역에 부합하여 맞아 보이는 응답입니다.

답을 하지 않은 문항은 오답으로 처리됩니다. 추측에 따른 불이익은 없습니다. 시험에는 점수에 반영되는 50 개의 문항이 있습니다.

### 채점 대상이 아닌 콘텐츠

시험에는 점수에 반영되지 않아 채점 대상이 아닌 15 개의 문항이 포함되어 있습니다. AWS 는 채점 대상이 아닌 질문에 대한 응시자 성과 정보를 수집하여 추후 채점 대상 질문으로 사용할 수 있도록 이러한 질문을 평가합니다. 이러한 채점 대상이 아닌 질문은 시험에서 식별되지 않습니다.

### 시험 결과

AWS Certified Machine Learning – Specialty (MLS-C01) 시험은 합격 또는 불합격이 결정되는 시험입니다. AWS 전문가가 자격증 분야 모범 사례 및 지침에 따라 설정한 최소 표준을 기준으로 시험 점수를 매깁니다.

시험 결과는 100~1,000 기준의 스케일링된 점수로 채점됩니다. 합격 최소 점수는 750 점입니다. 응시자의 점수는 전반적인 시험 성적과 합격 여부를 보여줍니다. 스케일링된 점수 모델은 난이도가 조금씩 다를 수 있는 여러 시험 형식에 걸쳐 점수를 균등하게 조정하는 데 도움이 됩니다.

점수 보고서에는 섹션 레벨별로 성적 분류표가 포함될 수 있습니다. 이 정보는 시험 성적에 대한 일반적인 피드백을 제공하기 위한 것입니다. 시험은 보상 점수 모델을 사용하므로 각 섹션에서 합격 점수를 얻을 필요는 없습니다. 전체 시험에만 합격하면 됩니다.

시험의 섹션마다 특정 가중치가 적용되므로 일부 섹션은 다른 섹션보다 문항 수가 많습니다. 표에는 응시자의 장단점을 보여주는 일반 정보가 포함되어 있습니다. 섹션 레벨 피드백을 검토할 때 주의하시기 바랍니다.

## 콘텐츠 개요

이 시험 가이드는 시험의 가중치, 테스트 영역 및 목표를 제공하며, 이 시험에 대한 종합적인 콘텐츠 목록은 아닙니다. 하지만 시험을 준비하는 데 도움이 되는 각 목표에 대한 추가 배경 정보가 있습니다. 다음 표에는 주요 콘텐츠 영역과 가중치가 나열되어 있습니다. 이 표는 추가 배경 정보가 포함되어 있는 전체 시험 콘텐츠 개요 앞에 나옵니다. 각 영역의 백분율은 채점되는 콘텐츠만 나타냅니다.

영역	시험 비율(%)
영역 1: 데이터 엔지니어링	20%
영역 2: 탐색적 데이터 분석	24%
영역 3: 모델링	36%
영역 4: 기계 학습 구현 및 운영	20%
합계	100%

## 영역 1: 데이터 엔지니어링

- 1.1 기계 학습용 데이터 리포지토리를 생성합니다.
  - 데이터 원본 식별(예: 콘텐츠 및 위치, 사용자 데이터와 같은 주요 소스)
  - 스토리지 미디어 결정(예: DB, 데이터 레이크, S3, EFS, EBS)
- 1.2 데이터 수집 솔루션을 식별하고 구현합니다.
  - 데이터 작업 스타일/유형(배치 부하, 스트리밍)

- 데이터 수집 파이프라인(배치 기반 기계 학습 워크로드 및 스트리밍 기반 기계 학습 워크로드)
  - Kinesis
  - Kinesis Analytics
  - Kinesis Firehose
  - EMR
  - Glue
- 작업 일정 관리

### 1.3 데이터 변환 솔루션을 식별하고 구현합니다.

- 데이터 전송 변환(ETL: Glue, EMR, AWS Batch)
- MapReduce(Hadoop, Spark, Hive)를 사용하여 기계 학습 관련 데이터 처리

## 영역 2: 탐색적 데이터 분석

### 2.1 모델링용 데이터를 정제하고 준비합니다.

- 누락 데이터, 손상 데이터, 중단 단어 등을 식별하고 처리
- 데이터 서식 지정, 정규화, 증대 및 크기 조정
- 레이블 지정 데이터(레이블이 지정된 데이터가 충분한 시기를 인식하고 완화 전략을 파악[데이터 레이블 지정 도구(Mechanical Turk, 수동 작업)])

### 2.2 특성 추출을 수행합니다.

- 텍스트, 음성, 이미지, 공개 데이터 집합 등과 같은 데이터 원본을 포함한 데이터 집합에서 특성을 파악하고 추출합니다.
- 특성 추출 개념 분석/평가(무기, 토큰화, 이상값, 종합 특성, 원-핫 인코딩, 데이터 차원 축소)

### 2.3 기계 학습용 데이터를 분석하고 시각화합니다.

- 그래프 작성(산점도, 시계열, 히스토그램, 박스 플롯)
- 기술된 통계 해석(상관 관계, 요약 통계, p 값)
- 클러스터링(계층적, 진단, 엘보우 플롯, 클러스터 크기)

## 영역 3: 모델링

### 3.1 비즈니스 문제를 기계 학습 문제로 설정합니다.

- 기계 학습을 사용해야 하는 경우/사용하지 않아야 하는 경우 확인
- 지도 학습과 비지도 학습의 차이점 이해
- 분류, 회귀, 예측, 클러스터링, 권장 사항 등 중에서 선택

### 3.2 특정 기계 학습 문제에 적합한 모델을 선택합니다.

- Xgboost, 로지스틱 회귀, K-평균, 선형 회귀, 결정 트리, 랜덤 포레스트, RNN, CNN, 앙상블, 전이 학습
- 모델 배경을 직관적으로 표현

### 3.3 기계 학습 모델을 훈련시킵니다.

- 검증 테스트 분할, 교차 검증 훈련
- Optimizer, 경사 하강법, 손실 함수, 로컬 미니마(local minima), 수렴, 배치, 확률 등
- 컴퓨팅 선택(GPU 와 CPU, 분산과 비분산, 플랫폼[Spark 와 비-Spark])
- 모델 업데이트 및 재훈련
  - 배치와 실시간/온라인

### 3.4 하이퍼파라미터 최적화를 수행합니다.

- 정규화
  - 드롭 아웃
  - L1/L2
- 교차 검증
- 모델 초기화
- 신경망 아키텍처(계층/노드), 학습률, 활성화 함수
- 트리 기반 모델(나무 수, 단계 수)
- 선형 모델(학습률)

### 3.5 기계 학습 모델을 평가합니다.

- 과적합/과소 적합 방지(편향 및 분산을 감지하고 처리)
- 지표(AUC-ROC, 정확도, 정밀도, 리콜, RMSE, F1 점수)
- 혼동 행렬
- 오프라인 및 온라인 모델 평가, A/B 테스트
- 지표를 사용하여 모델 비교(모델 훈련 시간, 모델 품질, 엔지니어링 비용)
- 교차 검증

## 영역 4: 기계 학습 구현 및 운영

### 4.1 성능, 가용성, 확장성, 복원성 및 내결함성을 고려한 기계 학습 솔루션을 구축합니다.

- AWS 환경 로깅 및 모니터링
  - CloudTrail 및 CloudWatch
  - 구축 오류 모니터링
- 다중 리전, 다중 AZ
- AMI/골든 이미지
- Docker 컨테이너

- Auto Scaling 그룹
- 라이트사이징(Rightsizing)
  - 인스턴스
  - 프로비저닝된 IOPS
  - 볼륨
- 로드 밸런싱
- AWS 모범 사례

4.2 특정 문제에 대한 적절한 기계 학습 서비스 및 기능을 추천하고 구현합니다.

- ML on AWS(애플리케이션 서비스)
  - Poly
  - Lex
  - Transcribe
- AWS 서비스 한도
- 자체 모델 구축과 SageMaker 기본 제공 알고리즘
- 인프라: (스팟, 인스턴스 유형), 비용 고려 사항
  - AWS Batch 를 사용한 딥 러닝 모델 교육에 스팟 인스턴스 사용

4.3 기계 학습 솔루션에 기본 AWS 보안 사례를 적용합니다.

- IAM
- S3 버킷 정책
- 보안 그룹
- VPC
- 암호화/익명화

4.4 기계 학습 솔루션을 배포하고 운영 가능하도록 전환합니다.

- 엔드포인트 노출 및 엔드포인트와 상호 작용
- 기계 학습 모델 버전 관리
- A/B 테스트
- 파이프라인 재훈련
- 기계 학습 디버깅/문제 해결
  - 성능 저하 감지 및 완화
  - 모델 성능 모니터링

## 부록

### 시험에서 다룰 수 있는 주요 도구, 기술 및 개념은 무엇입니까?

다음은 시험에서 다룰 수 있는 도구 및 기술 목록입니다(전체 목록은 아님). 이 목록은 변경될 수 있으며 시험에서 다루는 서비스, 기능 또는 기술의 일반적인 범위를 이해하는 데 도움이 됩니다. 이 목록에서 일반 도구 및 기술은 특별한 순서 없이 표시됩니다. AWS 서비스는 기본 기능에 따라 그룹화됩니다. 이러한 기술 중 일부는 시험에서 다른 기술보다 더 많이 다룰 수 있지만, 이 목록에서 순서 및 배치는 상대적인 비중이나 중요성을 나타내지 않습니다.

- 수집/모음
- 처리/ETL
- 데이터 분석/시각화
- 모델 훈련
- 모델 배포/추론
- 운영
- AWS 기계 학습 애플리케이션 서비스
- 기계 학습과 관련된 언어(예: Python, Java, Scala, R, SQL)
- 노트북 및 통합 개발 환경(IDE)

### AWS 서비스 및 기능

분석:

- Amazon Athena
- Amazon EMR
- Amazon Kinesis Data Analytics
- Amazon Kinesis Data Firehose
- Amazon Kinesis Data Streams
- Amazon QuickSight

컴퓨팅:

- AWS Batch
- Amazon EC2

컨테이너:

- Amazon Elastic Container Registry(Amazon ECR)
- Amazon Elastic Container Service(Amazon ECS)

- Amazon Elastic Kubernetes Service(Amazon EKS)

데이터베이스:

- AWS Glue
- Amazon Redshift

사물 인터넷(IoT)

- AWS IoT Greengrass

기계 학습:

- Amazon Comprehend
- AWS Deep Learning AMIs(DLAMI)
- AWS DeepLens
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Lex
- Amazon Polly
- Amazon Rekognition
- Amazon SageMaker
- Amazon Textract
- Amazon Transcribe
- Amazon Translate

관리 및 거버넌스:

- AWS CloudTrail
- Amazon CloudWatch

네트워킹 및 콘텐츠 전송:

- Amazon VPC

보안, 자격 증명 및 규정 준수:

- AWS Identity and Access Management(IAM)

서버리스:

- AWS Fargate
- AWS Lambda

스토리지:

- Amazon Elastic File System(Amazon EFS)



- Amazon FSx
- Amazon S3

## 대상 범위가 아닌 AWS 서비스 및 기능

다음은 시험에서 다루지 않는 AWS 서비스 및 기능의 목록입니다(전체 목록은 아님). 여기에 나열된 서비스와 기능이 시험 콘텐츠에서 제외되는 모든 AWS 서비스 및 기능을 나타내지는 않습니다. 시험의 대상 직무 역할과 전혀 관련이 없는 서비스 또는 기능은 관련성이 없는 것으로 간주되므로 이 목록에서 제외됩니다.

대상 범위가 아닌 AWS 서비스 및 기능에는 다음이 포함됩니다.

- AWS Data Pipeline
- AWS DeepRacer
- Amazon Machine Learning(Amazon ML)