

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

### 서론

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

또한 이 시험에서는 응시자가 다음 태스크를 완료할 수 있는지도 확인합니다.

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

### 대상 응시자 설명

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

#### 권장되는 AWS 지식

대상 응시자는 다음과 같은 AWS 관련 지식이 있어야 합니다.

- 기본 ML 알고리즘을 직관적으로 표현할 수 있는 능력
- 기본적인 하이퍼파라미터 최적화 수행 경험
- ML 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 지식

다음 목록에는 대상 응시자가 갖추지 않아도 되는 지식이 포함되어 있습니다. 이 목록에 모든 사항이 포함된 것은 아닙니다. 다음 영역에 대한 지식은 시험 범위가 아닙니다.

- 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 EMR을 위한 DevOps 관련 태스크

부록을 참고하여 시험에 출제될 수 있는 기술 및 개념 목록, 시험 범위에 해당하는 AWS 서비스 및 기능 목록, 시험 범위가 아닌 AWS 서비스 및 기능 목록을 확인하시기 바랍니다.

## 시험 콘텐츠

### 답안 유형

이 시험의 문항은 두 가지 유형으로 제공됩니다.

- **선다형:** 정답 1개와 오답 3개(정답 이외의 답)가 있습니다.
- **복수 응답형:** 5개 이상의 응답 항목 중에 2개 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 50개의 문항이 포함되어 있습니다.

## 채점되지 않는 콘텐츠

시험에는 점수에 반영되지 않아 채점되지 않는 15개의 문항이 포함되어 있습니다. AWS는 채점되지 않는 문항에 대한 응시자 성적 정보를 수집하여 추후 채점 대상 문항으로 사용할 수 있도록 이러한 문항을 평가합니다. 이러한 채점되지 않는 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 100~1,000점의 변환 점수로 보고됩니다. 합격 최소 점수는 750점입니다. 응시자의 점수는 전반적인 시험 성적과 합격 여부를 보여줍니다. 변환 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. 시험은 보상 점수 모델을 사용하므로 각 섹션에서 합격 점수를 얻을 필요는 없으며, 전체 시험에만 합격하면 됩니다.

시험의 섹션마다 특정 가중치가 적용되므로 일부 섹션은 다른 섹션보다 문항 수가 많습니다. 분류표에는 응시자의 장단점을 강조하여 보여주는 일반 정보가 포함되어 있습니다. 섹션별 피드백을 파악할 때 주의하시기 바랍니다.

## 내용 개요

이 시험 안내서에서는 시험의 가중치, 콘텐츠 도메인 및 태스크 설명 자료를 제공합니다. 이 안내서에서 시험 내용의 전체 목록을 제공하지는 않습니다. 그러나 각 태스크 설명에 관한 추가 맥락 정보를 사용하여 시험을 준비하는 데 참고할 수 있습니다.

험의 콘텐츠 도메인과 가중치는 다음과 같습니다.

- 도메인 1: 데이터 엔지니어링(채점되는 콘텐츠의 20%)
- 도메인 2: 탐색적 데이터 분석(채점되는 콘텐츠의 24%)
- 도메인 3: 모델링(채점되는 콘텐츠의 36%)
- 도메인 4: 기계 학습 구현 및 운영(채점되는 콘텐츠의 20%)

## 도메인 1: 데이터 엔지니어링

태스크 설명 1.1: ML용 데이터 리포지토리 생성

- 데이터 소스 식별(예: 콘텐츠 및 위치, 사용자 데이터와 같은 주요 소스)
- 스토리지 매체 결정(예: 데이터베이스, Amazon S3, Amazon Elastic File System[Amazon EFS], Amazon Elastic Block Store[Amazon EBS])

태스크 설명 1.2: 데이터 수집 솔루션 식별 및 구현

- 데이터 작업 스타일 및 작업 유형 식별(예: 배치 로드, 스트리밍)
- 데이터 수집 파이프라인 오케스트레이션(배치 기반 ML 워크로드 및 스트리밍 기반 ML 워크로드)
  - Amazon Kinesis
  - Amazon Data Firehose
  - Amazon EMR
  - AWS Glue
  - Amazon Managed Service for Apache Flink
- 작업 예약

태스크 설명 1.3: 데이터 변환 솔루션 식별 및 구현

- 전송 중인 데이터 변환(ETL, AWS Glue, Amazon EMR, AWS Batch)
- MapReduce를 사용하여 ML별로 데이터 처리(예: Apache Hadoop, Apache Spark, Apache Hive)

## 도메인 2: 탐색적 데이터 분석

### 태스크 설명 2.1: 모델링용 데이터 정제 및 준비

- 누락 데이터, 손상 데이터, 중단 단어 식별 및 처리
- 데이터 포맷, 정규화, 강화, 크기 조정
- 레이블이 지정된 데이터가 충분한지 확인
  - 완화 전략 파악
  - 데이터 라벨링 도구 활용(예: Amazon Mechanical Turk)

### 태스크 설명 2.2: 특성 추출 수행

- 텍스트, 음성, 이미지, 공개 데이터세트 등의 데이터 소스를 비롯해 데이터세트에서 특성 파악 및 추출
- 특성 추출 개념 분석 및 평가(예: 묶기, 토큰화, 이상값, 종합 기능, 원-핫 인코딩, 데이터 차원 축소)

### 태스크 설명 2.3: ML용 데이터 분석 및 데이터 시각화

- 그래프 생성(예: 산점도, 시계열, 히스토그램, 박스 플롯)
- 기술적 통계 해석(예: 상관관계, 요약 통계, p 값)
- 클러스터 분석 수행(예: 계층적 분석, 진단, 엘보우 플롯, 클러스터 크기)

## 도메인 3: 모델링

### 태스크 설명 3.1: 비즈니스 문제를 ML 문제로 규정

- ML을 사용해야 하는 경우와 그렇지 않은 경우를 결정
- 지도 학습과 비지도 학습의 차이점 이해
- 분류, 회귀, 예측, 클러스터링, 권장 사항, 파운데이션 모델 중에서 선택

### 태스크 설명 3.2: 특정 ML 문제에 적합한 모델 선택

- XGBoost, 로지스틱 회귀, K-평균, 선형 회귀, 의사결정 트리, 랜덤 포레스트, RNN, CNN, 앙상블, 전이 학습, 대규모 언어 모델(LLM)
- 모델 배경을 직관적으로 표현

### 태스크 설명 3.3: ML 모델 훈련

- 훈련과 검증 간의 데이터 분할(예: 크로스 검증)
- ML 훈련에 최적화된 기법 파악(예: 경사 하강법, 손실 함수, 수렴)
- 적절한 컴퓨팅 리소스 선택(예: GPU 또는 CPU, 분산 또는 비분산)
  - 적절한 컴퓨팅 플랫폼 선택(Spark 또는 Spark가 아닌 플랫폼)
- 모델 업데이트 및 재훈련
  - 배치 또는 실시간/온라인

### 태스크 설명 3.4: 하이퍼파라미터 최적화 수행

- 정규화 수행
  - 드롭아웃
  - L1/L2
- 크로스 검증 수행
- 모델 초기화
- 신경망 아키텍처(계층 및 노드), 학습률, 활성화 함수 이해
- 트리 기반 모델 이해(트리 수, 레벨 수)
- 선형 모델 이해(학습률)

### 태스크 설명 3.5: ML 모델 평가

- 과대 적합 또는 과소 적합 방지
  - 편향과 분산 탐지 및 처리
- 지표 평가(예: 곡선형 아래 영역[AUC] - 수신기 작동 특성[ROC], 정확도, 정밀도, 재현율, 루트 평균 제곱 오차[RMSE], F1 점수)
- 혼동 행렬 해석
- 오프라인 및 온라인 모델 평가 수행(A/B 테스트)
- 지표를 사용하여 모델 비교(예: 모델 훈련 시간, 모델 품질, 엔지니어링 비용)
- 크로스 검증 수행

## 도메인 4: Machine Learning 구현 및 운영

태스크 설명 4.1: 성능, 가용성, 확장성, 복원성 및 내결함성을 고려한 ML 솔루션 구축

- AWS 환경 로깅 및 모니터링
  - AWS CloudTrail 및 Amazon CloudWatch
  - 오류 모니터링 솔루션 구축
- 여러 AWS 리전과 여러 가용 영역에 배포
- AMI와 골든 이미지 생성
- Docker 컨테이너 생성
- Auto Scaling 그룹 배포
- 리소스 크기를 적절하게 조정(예: 인스턴스, 프로비저닝된 IOPS, 볼륨)
- 로드 밸런싱 수행
- AWS 모범 사례 따르기

태스크 설명 4.2: 특정 문제에 적합한 ML 서비스 및 기능 권장 사항 및 구현

- AWS 기반 ML(애플리케이션 서비스), 예:
  - Amazon Polly
  - Amazon Lex
  - Amazon Transcribe
  - Amazon Q
- AWS Service Quotas 이해
- 사용자 지정 모델 구축 시기 및 Amazon SageMaker 기본 제공 알고리즘의 사용 시기 결정
- AWS 인프라(예: 인스턴스 유형)와 비용 고려 사항 파악
  - AWS Batch를 사용한 딥 러닝 모델 훈련에 스팟 인스턴스 사용

#### 태스크 설명 4.3: ML 솔루션에 기본 AWS 보안 사례 적용

- AWS Identity and Access Management(IAM)
- S3 버킷 정책
- 보안 그룹
- VPC
- 암호화 및 익명화

#### 태스크 설명 4.4: ML 솔루션 배포 및 운영

- 엔드포인트 노출 및 엔드포인트와 상호작용
- ML 모델 이해
- A/B 테스트 수행
- 파이프라인 재훈련
- ML 모델 디버깅 및 문제 해결
  - 성능 저하 감지 및 완화
  - 모델 성능 모니터링



## 부록

### 시험에 출제될 수 있는 기술 및 개념

다음 목록에는 시험에 출제될 수 있는 기술 및 개념이 포함되어 있습니다. 이 목록에 모든 사항이 포함된 것은 아니며 변경될 수 있습니다. 이 목록에 나와 있는 다음 항목의 배치와 순서가 시험에서의 상대적 가중치 또는 중요도를 의미하지는 않습니다.

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

### 시험 범위에 포함되는 AWS 서비스 및 기능

다음 목록에는 시험 범위에 해당하는 AWS 서비스 및 기능이 나와 있습니다. 이 목록에 모든 사항이 포함된 것은 아니며 변경될 수 있습니다. AWS 제품 및 서비스는 주요 기능에 따라 다음과 같은 카테고리로 분류됩니다.

분석:

- Amazon Athena
- Amazon Data Firehose
- Amazon EMR
- AWS Glue
- Amazon Kinesis
- Amazon Kinesis Data Streams
- AWS Lake Formation
- Amazon Managed Service for Apache Flink
- Amazon OpenSearch Service
- Amazon QuickSight

#### 컴퓨팅:

- AWS Batch
- Amazon EC2
- AWS Lambda

#### 컨테이너:

- Amazon Elastic Container Registry(Amazon ECR)
- Amazon Elastic Container Service(Amazon ECS)
- Amazon Elastic Kubernetes Service(Amazon EKS)
- AWS Fargate

#### 데이터베이스:

- Amazon Redshift

#### 사물 인터넷(IoT):

- AWS IoT Greengrass

#### 기계 학습:

- Amazon Bedrock
- Amazon Comprehend
- AWS Deep Learning AMIs(DLAMI)
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Lex
- Amazon Kendra
- Amazon Mechanical Turk
- Amazon Polly
- Amazon Q
- 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)

스토리지:

- Amazon Elastic Block Store(Amazon EBS)
- Amazon Elastic File System(Amazon EFS)
- Amazon FSx
- Amazon S3

## 시험 범위가 아닌 AWS 서비스 및 기능

다음 목록에는 시험 범위가 아닌 AWS 서비스 및 기능이 나와 있습니다. 이 목록에 모든 사항이 포함된 것은 아니며 변경될 수 있습니다. 시험의 대상 작업 역할과 전혀 관련이 없는 AWS 제품 및 서비스는 다음 목록에서 제외됩니다.

분석:

- AWS Data Pipeline

기계 학습:

- AWS DeepRacer
- Amazon Machine Learning(Amazon ML)

## 설문 조사

이 시험 안내서가 도움이 되었나요? [설문 조사에 참여](#)하여 의견을 공유해 주시기 바랍니다.