



AWS Certified Machine Learning - Specialty

취득 가이드

박경규

AWS Certified Machine Learning – Specialty 자격증(이하 MLS)



인공지능/기계학습(AI/ML) 솔루션을 설계, 구축, 배포, 최적화, 교육, 튜닝 및 유지 관리하는 능력을 검증합니다.

현재 비즈니스 요구 사항과 향후 예상되는 요구 사항을 충족하도록 AWS 서비스를 통합하는 솔루션 설계

- 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

시험 개요

- 자격 요건: 없음
- 응시 방법: 온라인 (Pearson VUE, PSI) / 오프라인 테스트 센터
- 시험 형식: 65문항 / 130분
- 응시 비용: 300 USD
- 합격 조건: 1000 점 중 750점 이상
- 유효 기간: 취득 일자 기준 3년

취득 혜택

- 취득시 사내 지원 내용
 - 응시료 100% 지원
 - 자기계발 격려금 50만원
- AWS
 - 타 AWS 시험 50% 할인 쿠폰

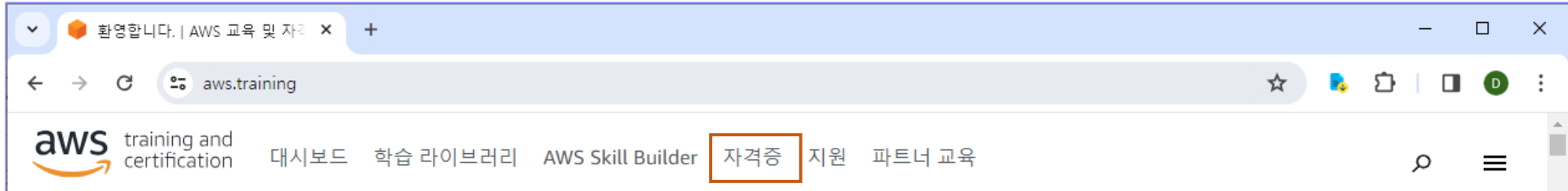
도메인	시험 비율
데이터 엔지니어링	20%
탐색적 데이터 분석	24%
모델링	36%
기계 학습 구현 및 운영	20%

TIP

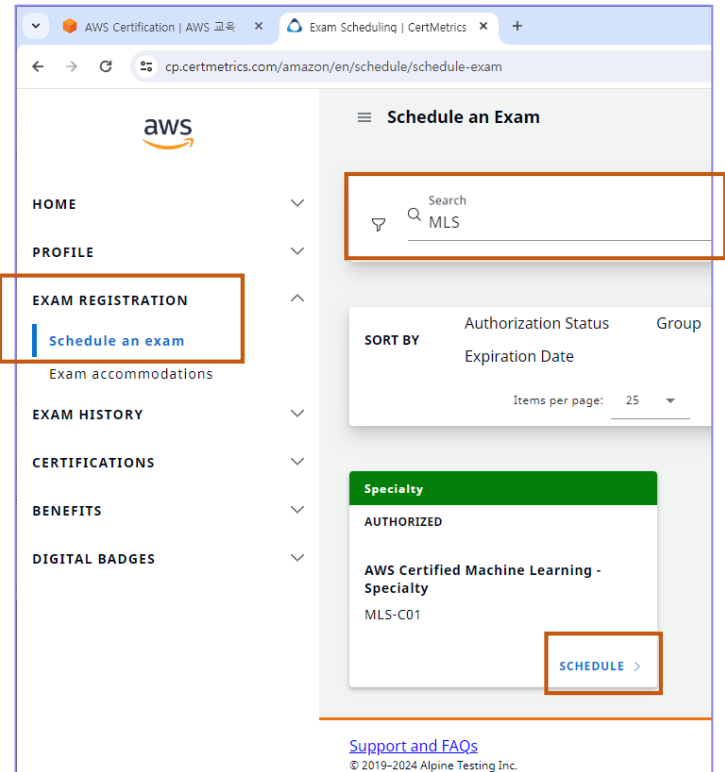
- 온라인으로 시험을 응시하면, 감독관과 의사소통, 주변 환경 점검, **여권** 준비 등 조건이 복잡합니다.
(오프라인의 경우 주민등록증/운전면허증 + 본인 명의 신용카드로 확인)
따라서, **오프라인**으로 시험을 보는 것을 추천하며, **Pearson VUE**로 응시하는 것을 추천합니다.
- 언어 선택 시 한국어를 선택하면 영문 문제도 제공되기 때문에 한국어를 선택하는 것을 추천합니다.

AWS MLS 시험 접수 방법

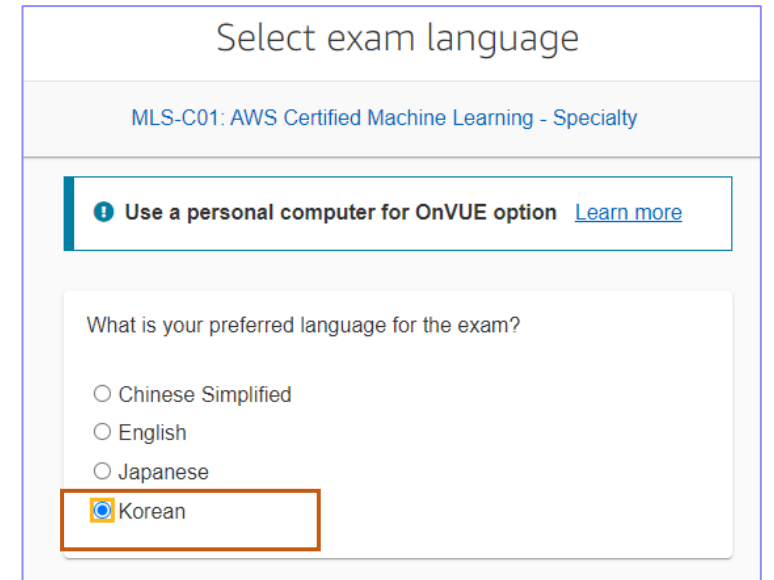
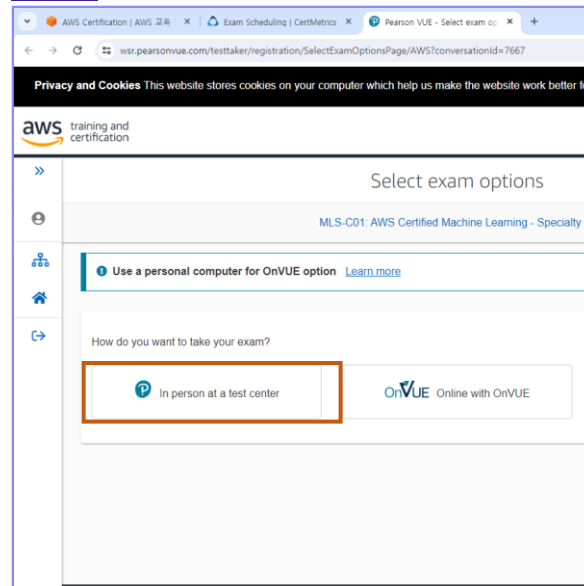
1 <https://www.aws.training/> 접속, 개인메일로 가입, 자격증 메뉴 클릭



2 MLS를 검색하여, Pearson VUE로 예약



3 시험 옵션 선택



4 시험 장소, 일자 선택 및 결제

시험 등록이 완료되면 Pearson VUE로 부터 확인 메일이 옵니다.

AWS MLS 권장 지식

권장 지식

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

문제 유형

- 선다형: 정답 1개와 오답 3개(정답 이외의 답)가 있습니다.
- 복수 응답형: 5개 이상의 응답 항목 중에 2개 이상의 정답이 있습니다.
- 총 65문제로 시험에는 점수에 반영되지 않아 채점되지 않는 15개의 문항이 포함되어 있습니다.

AWS MLS 시험 문항

도메인 1: 데이터 엔지니어링

1.1 ML용 데이터 리포지토리 생성

- 데이터 원본 식별 : 콘텐츠 및 위치, 사용자 데이터와 같은 주요 소스
- 스토리지 매체 결정 : 데이터베이스, Amazon S3, Amazon EFS, Amazon EBS

1.2 데이터 수집 솔루션 식별 및 구현

- 데이터 작업 유형 식별 : 배치 로드, 스트리밍
- 데이터 수집 파이프라인 오케스트레이션 : Kinesis, Kinesis Data Firehose, Amazon EMR, AWS Glue, Amazon Managed Service for Apache Flink
- 작업 예약

1.3. 데이터 변환 솔루션 식별 및 구현

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

도메인 2: 탐색적 데이터 분석

2.1 모델링용 데이터 정제 및 준비

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

2.2 특성 추출 수행

- 텍스트, 음성, 이미지, 공개 데이터 집합과 같은 데이터 원본 등의 데이터 집합에서 기능을 파악하고 추출
- 특성 추출 개념 분석 및 평가 : 묶기, 토큰화, 이상값, 종합 기능, 원-핫 인코딩, 데이터 차원 축소

2.3 ML용 데이터 분석 및 데이터 시각화

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

AWS MLS 시험 문항

도메인 3: 모델링

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 모델 중에서 선택

3.2 특정 ML 문제에 적합한 모델 선택

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

3.3 ML 모델 훈련

- 훈련과 검증 간의 데이터 분할 : 교차 검증
- ML 훈련에 최적화된 기법 파악 : 경사 하강법, 손실 함수, 수렴
- 적절한 컴퓨팅 리소스 선택 : GPU 또는 CPU, 분산 또는 비분산
- 모델 업데이트 및 재훈련 : 배치 또는 실시간/온라인

3.4: 하이퍼파라미터 최적화 수행

- 정규화 수행 : 드롭 아웃, L1/L2
- 교차 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)

3.5: ML 모델 평가

- 과대 적합 또는 과소 적합 방지
- 지표 평가 : , 정확도, 정밀도, 재현율, RMSE, F1 점수, ROC, AUC, 혼동 행렬 해석
- 오프라인 및 온라인 모델 평가 수행 : A/B 테스트
- 지표를 사용하여 모델 비교 : 모델 훈련 시간, 모델 품질, 엔지니어링 비용
- 교차 검증 수행

AWS MLS 시험 문항

도메인 4: Machine Learning 구현 및 운영

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

- AWS 환경 로깅 및 모니터링: AWS CloudTrail, Amazon CloudWatch
- 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

4.2 특정 문제에 적합한 ML 서비스 및 기능 추천 및 구현

- ML on AWS(애플리케이션 서비스) : Amazon Polly, Lex, Transcribe
- 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 딥러닝 모델 훈련에 스팟 인스턴스 사용

4.3 ML 솔루션에 기본 AWS 보안 사례 적용

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

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

기출 문제

<https://www.examtopycs.com/exams/amazon/aws-certified-machine-learning-specialty/>

AWS Certified Machine Learning

examtopics.com/exams/amazon/aws-certified-machine-learning-specialty/view/1/

HOMEUNLIMITED ACCESSPOPULAR EXAMSVIEW ALL EXAMSDOWNLOAD FREECOURSESCONTACTFORUM

Question #1

A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.

The model produces the following confusion matrix after evaluating on a test dataset of 100 customers:

n= 100	PREDICTED CHURN	PREDICTED CHURN
	Yes	No
ACTUAL Churn Yes	10	4
Actual No	10	76

Based on the model evaluation results, why is this a viable model for production?

A. The model is 86% accurate and the cost incurred by the company as a result of false negatives is less than the false positives. **Most Voted**

B. The precision of the model is 86%, which is less than the accuracy of the model.

C. The model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives. **Most Voted**

D. The precision of the model is 86%, which is greater than the accuracy of the model.

Hide Solution

Discussion 76

Correct Answer: A

Community vote distribution

C (71%)

A (29%)

Exam AWS Certified Machine Learning - Specialty topic 1 question 1 discussion

Switch to a [voting comment](#) **New**

Type your comment...

Submit

befydorta **Highly Voted** 3 months, 1 week ago

Selected Answer: C

itexamstest.com

no disscusion c :)

upvoted 40 times

[Removed] **Highly Voted** 2 years, 5 months ago

Should it be C? Cost incurred by the company as a result of false positives (predicted churn, actual not churn) is less than the false negatives (predicted not churn, actual churn). Incentive cost < churn cost.

upvoted 39 times

daidaidai 10 months, 2 weeks ago

Fully Agree:

FN = Predict Not churn, actual churn, high cost

FP = Predict churn, actual not churn, pay incentive, low cost.

so $FP < FN$,

The answer is C.

upvoted 3 times

기출 문제 풀이

번역



↓ Windows용 다운로드

↓ Mac용 다운로드

단축키 : Ctrl+C+C

단축키 : ⌘+C+C

문제풀이 확인

문제출처 : <https://www.examtopy.com/exams/amazon/aws-certified-machine-learning-specialty/view/14/>

다음 문제를 풀어줘. 문제의 지시문을 준수하고 가장 적합한 답을 찾아. 설명은 한국어로 해줘:

Question #131

A financial services company wants to adopt Amazon SageMaker as its default data science environment. The company's data scientists run machine learning(ML) models on confidential financial data. The company is worried about data egress and wants an ML engineer to secure the environment.

Which mechanisms can the ML engineer use to control data egress from SageMaker? (Choose three.)

- A. Connect to SageMaker by using a VPC interface endpoint powered by AWS PrivateLink.
- C. Disable root access on the SageMaker notebook instances.
- D. Enable network isolation for training jobs and models.
- E. Restrict notebook presigned URLs to specific IPs used by the company.
- F. Protect data with encryption at rest and in transit. Use AWS Key Management Service (AWS KMS) to manage encryption keys.

문제 #131: 금융 서비스 회사의 SageMaker 데이터

정답은 다음 세 가지입니다. (Choose three)

문제 #131: 금융 서비스 기업의 SageMaker 데이터

정답은 다음 세 가지입니다. (Choose three)

- A. VPC 인터페이스 엔드포인트 사용 (Connect to SageMaker by using a VPC interface endpoint powered by AWS PrivateLink)
- D. 훈련 작업 및 모델의 네트워크 격리 사용 (Enable network isolation for training jobs and models)

<https://gemini.google.com/>

<https://claude.ai/>

<https://chat.openai.com/>

• 금융 서비스 회사에서 기밀 재무 데이터를 처리하는 데 Amazon SageMaker를 사용할

기출문제 풀이

10MLS-01-65.docx

10MLS-66-130.docx

10MLS-131-192.docx

10MLS-193-273.docx

Question #198

회사에서는 매플리케이션에서 생성된 문서의 분류를 예측하려고 합니다.

새 문서는 3초마다 Amazon S3 버킷에 저장됩니다.

이 회사는 문서 텍스트를 분류하기 위해 Amazon SageMaker 내에서 기계학습(ML) 모델의

세 가지 버전을 개발했습니다.

회사는 각 문서의 분류를 예측하기 위해 이 세 가지 버전을 배포하려고 합니다.

최소한의 운영 오버헤드로 이러한 요구사항을 충족하는 접근 방식은 무엇입니까?

A. 새 문서가 생성될 때 AWS Lambda 함수를 호출하는 S3 이벤트 알림을 구성합니다.

3개의 SageMaker 배치 변환 작업(각 문서의 각 모델에 대해 하나의 배치 변환 작업)을 생성하도록 Lambda 함수를 구성합니다.

B. 모든 모델을 단일 SageMaker 엔드포인트에 배포합니다.

각 모델을 production variant으로 취급합니다.

새 문서가 생성될 때 AWS Lambda 함수를 호출하는 S3 이벤트 알림을 구성합니다.

각 production variant을 호출하고 각 모델의 결과를 반환하도록 Lambda 함수를 구성합니다.

C. 각 모델을 자체 SageMaker 엔드포인트에 배포합니다.

새 문서가 생성될 때 AWS Lambda 함수를 호출하는 S3 이벤트 알림을 구성합니다.

각 엔드포인트를 호출하고 각 모델의 결과를 반환하도록 Lambda 함수를 구성합니다.

D. 각 모델을 자체 SageMaker 엔드포인트에 배포합니다.

3개의 AWS Lambda 함수를 생성합니다.

서로 다른 엔드포인트를 호출하고 결과를 반환하도록 각 Lambda 함수를 구성합니다.

새 문서가 생성될 때 Lambda 함수를 호출하도록 세 가지 S3 이벤트 알림을 구성합니다.

Question #198

A company wants to predict the classification of documents that are created from an application.

New documents are saved to an Amazon S3 bucket every 3 seconds.

The company has developed three versions of a machine learning (ML) model within Amazon SageMaker to classify document text.

The company wants to deploy these three versions to predict the classification of each document.

Which approach will meet these requirements with the LEAST operational overhead?

A. Configure an S3 event notification that invokes an AWS Lambda function when new documents are created.

Configure the Lambda function to create three SageMaker batch transform jobs, one batch transform job for each model for each document.

B. Deploy all the models to a single SageMaker endpoint.

Treat each model as a production variant.

Configure an S3 event notification that invokes an AWS Lambda function when new documents are created.

Configure the Lambda function to call each production variant and return the results of each model.

C. Deploy each model to its own SageMaker endpoint. Configure an S3 event notification that invokes an AWS Lambda function when new documents are created.

Configure the Lambda function to call each endpoint and return the results of each model.

D. Deploy each model to its own SageMaker endpoint.

Create three AWS Lambda functions.

Configure each Lambda function to call a different endpoint and return the results.

Configure three S3 event notifications to invoke the Lambda functions when new documents are created.

Correct Answer: B

최적의 운영 부담을 고려할 때 회사는 B 옵션을 선택하는 것이 가장 좋습니다.

이유는 다음과 같습니다.

운영 부담:

- A 옵션: 가장 복잡합니다. 여러 SageMaker 배치 변환 작업을 만들고 관리해야 하며, Lambda 함수는 각 문서에 대해 세 개의 작업을 트리거해야 합니다.
- B 옵션: 가장 간단합니다. 모든 모델을 단일 엔드포인트에 배포하고 Lambda 함수를 사용하여 각 변형을 호출하면 됩니다.
- C 옵션: A 옵션보다 간단하지만 각 모델에 별도의 엔드포인트를 관리해야 합니다.
- D 옵션: 세 개의 Lambda 함수와 세 개의 S3 이벤트 알림을 관리해야 하므로 가장 복잡합니다.

성능:

- 모든 옵션은 새로운 문서가 들어오면 3 초 이내에 분류 결과를 반환할 수 있습니다.

비용:

- A 옵션: 세 개의 배치 변환 작업을 실행하므로 가장 비용이 많이 듭니다.
- B 옵션: 한 번의 엔드포인트 호출로 결과를 얻으므로 가장 비용이 적습니다.
- C 옵션: 각 모델에 대해 별도의 엔드포인트를 유지보수해야 하므로 B 옵션보다 비용이 많이 듭니다.
- D 옵션: 세 개의 Lambda 함수와 세 개의 S3 이벤트 알림을 사용하므로 C 옵션보다 비용이 많이 듭니다.



Thank you 😊