

# ITPassLeader



Pass Your Next Certification Exam Fast!

Select a vendor... Select an test... Your email address [Free Download Demo](#)



Instant Download



365 Days Free Updates



Money Back Guarantee



Security & Privacy

Choose the version that fits your needs

PDF Version

Desktop Test Engine

Online Test Engine

Latest and Up-to-Date exam dumps with real exam questions answers.



Get 12-Months free updates without any extra charges.



Experience same exam environment before appearing in the certification exam.



100% exam passing guarantee in the first attempt.



20% discount on more than one license and 30% discount on 5+ license purchases.



100% secure purchase on SSL.



Completely private purchase without sharing your personal info with anyone.



<http://www.itpassleader.com>

High-praise Exam Dumps Questions grant you success by high pass rate - ITPassLeader

**Exam** : **AWS-Certified-Machine-Learning-Specialty-KR**

**Title** : **AWS Certified Machine Learning - Specialty (MLS-C01 Korean Version)**

**Vendor** : **Amazon**

**Version** : **DEMO**

**QUESTION NO: 1**

기계 학습 전문가가 회사에서 Amazon SageMaker를 교육에 활용할 수 있도록 사용자 지정 ResNet 모델을 Docker 컨테이너에 패키징하고 있습니다. 전문가는 Amazon EC2 P3 인스턴스를 사용하여 모델을 교육하고 있으며 NVIDIA GPU를 활용하도록 Docker 컨테이너를 적절하게 구성해야 합니다. 전문가가 해야 할 일?

- A. NVIDIA 드라이버를 Docker 이미지와 번들링
- B. NVIDIA-Docker와 호환되도록 Docker 컨테이너 빌드
- C. GPU 인스턴스에서 실행할 Docker 컨테이너의 파일 구조를 구성합니다.
- D. Amazon SageMaker Create TrainingJob 요청 본문에서 GPU 플래그 설정

**Answer: B**

Explanation:

To leverage the NVIDIA GPUs on Amazon EC2 P3 instances, the Machine Learning Specialist needs to build the Docker container to be NVIDIA-Docker compatible. NVIDIA-Docker is a tool that enables GPU- accelerated containers to run on Docker. It automatically configures the container to access the NVIDIA drivers and libraries on the host system. The Specialist does not need to bundle the NVIDIA drivers with the Docker image, as they are already installed on the EC2 P3 instances. The Specialist does not need to organize the Docker container's file structure to execute on GPU instances, as this is not relevant for GPU compatibility. The Specialist does not need to set the GPU flag in the Amazon SageMaker Create TrainingJob request body, as this is only required for using Elastic Inference accelerators, not EC2 P3 instances. References: NVIDIA-Docker, Using GPU-Accelerated Containers, Using Elastic Inference in Amazon SageMaker

**QUESTION NO: 2**

한 자동차 회사가 여러 도시에 대리점을 두고 있습니다. 이 회사는 머신러닝(ML) 추천 시스템을 사용하여 고객에게 자동차를 판매합니다.

ML 엔지니어는 각 자동차에 대한 여러 속성을 포함하는 데이터 세트를 기반으로 ML 추천 모델을 학습시켰습니다. 데이터 세트에는 자동차 브랜드, 차종, 연비, 가격 등의 속성이 포함되어 있습니다.

ML 엔지니어는 Amazon SageMaker Data Wrangler를 사용하여 데이터를 분석하고 시각화합니다. ML 엔지니어는 특정 유형의 자동차에 대한 가격 분포를 파악해야 합니다. 이러한 요구 사항을 충족하기 위해 ML 엔지니어는 어떤 유형의 시각화를 사용해야 합니까?

- A. SageMaker Data Wrangler 산점도 시각화를 사용하여 자동차 가격과 자동차 유형 간의 관계를 조사합니다.
- B. SageMaker Data Wrangler의 빠른 모델 시각화를 사용하여 데이터를 빠르게 평가하고 자동차 가격과 자동차 유형에 대한 중요도 점수를 생성합니다.
- C. SageMaker Data Wrangler 이상 탐지 시각화를 사용하여 특정 기능에 대한 이상치를 식별합니다.
- D. SageMaker Data Wrangler 히스토그램 시각화를 사용하여 특정 기능의 값 범위를 검사합니다.

**Answer: D**

**QUESTION NO: 3**

기계 학습 전문가는 이전에 로컬 머신에서 scikit-learn을 사용하여 로지스틱 회귀 모델을

훈련했으며 이제 전문가는 추론을 위해 프로덕션에 배포하려고 합니다.

Amazon SageMaker가 로컬에서 훈련된 모델을 호스팅할 수 있도록 하려면 어떤 단계를 수행해야 하나요?

- A. 추론 코드로 Docker 이미지를 빌드합니다. 레지스트리 호스트 이름으로 Docker 이미지에 태그를 지정하고 Amazon ECR에 업로드합니다.
- B. 훈련된 모델을 직렬화하여 배포를 위해 형식을 압축합니다. 레지스트리 호스트 이름으로 Docker 이미지에 태그를 지정하고 Amazon S3에 업로드합니다.
- C. 훈련된 모델을 직렬화하여 배포를 위해 형식을 압축합니다. 이미지를 빌드하고 Docker Hub에 업로드합니다.
- D. 추론 코드로 Docker 이미지를 빌드합니다. Docker Hub를 구성하고 이미지를 Amazon ECR에 업로드합니다.

**Answer: A**

Explanation:

To deploy a model that was trained locally to Amazon SageMaker, the steps are:

Build the Docker image with the inference code. The inference code should include the model loading, data preprocessing, prediction, and postprocessing logic. The Docker image should also include the dependencies and libraries required by the inference code and the model.

Tag the Docker image with the registry hostname and upload it to Amazon ECR. Amazon ECR is a fully managed container registry that makes it easy to store, manage, and deploy container images. The registry hostname is the Amazon ECR registry URI for your account and Region. You can use the AWS CLI or the Amazon ECR console to tag and push the Docker image to Amazon ECR.

Create a SageMaker model entity that points to the Docker image in Amazon ECR and the model artifacts in Amazon S3. The model entity is a logical representation of the model that contains the information needed to deploy the model for inference. The model artifacts are the files generated by the model training process, such as the model parameters and weights. You can use the AWS CLI, the SageMaker Python SDK, or the SageMaker console to create the model entity.

Create an endpoint configuration that specifies the instance type and number of instances to use for hosting the model. The endpoint configuration also defines the production variants, which are the different versions of the model that you want to deploy. You can use the AWS CLI, the SageMaker Python SDK, or the SageMaker console to create the endpoint configuration.

Create an endpoint that uses the endpoint configuration to deploy the model. The endpoint is a web service that exposes an HTTP API for inference requests. You can use the AWS CLI, the SageMaker Python SDK, or the SageMaker console to create the endpoint.

AWS Machine Learning Specialty Exam Guide

AWS Machine Learning Training - Deploy a Model on Amazon SageMaker

AWS Machine Learning Training - Use Your Own Inference Code with Amazon SageMaker Hosting Services

#### QUESTION NO: 4

회사에 TensorFlow에 제품 추천 엔진이 내장된 전자상거래 웹사이트가 있습니다. 추천 엔진 엔드포인트는 Amazon SageMaker에서 호스팅합니다. 3개의 컴퓨팅 최적화 인스턴스가 웹

사이트의 예상 최대 로드를 지원합니다.

매월 초 상품 추천 페이지의 응답 시간이 증가하고 있습니다. 일부 사용자에게 오류가 발생했습니다. 웹 사이트는 단일 시간대에서 평일 오전 8시에서 오후 6시 사이에 대부분의 트래픽을 수신합니다.

다음 중 비용을 최소화하면서 문제를 해결하는 데 가장 효과적인 옵션은 무엇입니까? (2개를 선택하세요.)

- A. Amazon Elastic Inference(EI) 액셀러레이터를 사용하도록 엔드포인트를 구성합니다.
- B. 두 가지 프로덕션 변형으로 새 끝점 구성을 만듭니다.
- C. InvocationsPerInstance 지표를 사용하여 자동으로 확장되도록 엔드포인트를 구성합니다.
- D. 두 번째 인스턴스 풀을 배포하여 모델의 블루/그린 배포를 지원합니다.
- E. 버스트 가능한 인스턴스를 사용하도록 엔드포인트를 재구성합니다.

**Answer:** A C

Explanation:

The solution A and C are the most effective in solving the issue while keeping costs to a minimum. The solution A and C involve the following steps:

Configure the endpoint to use Amazon Elastic Inference (EI) accelerators. This will enable the company to reduce the cost and latency of running TensorFlow inference on SageMaker. Amazon EI provides GPU- powered acceleration for deep learning models without requiring the use of GPU instances. Amazon EI can attach to any SageMaker instance type and provide the right amount of acceleration based on the workload<sup>1</sup>.

Configure the endpoint to automatically scale with the Invocations Per Instance metric. This will enable the company to adjust the number of instances based on the demand and traffic patterns of the website. The Invocations Per Instance metric measures the average number of requests that each instance processes over a period of time. By using this metric, the company can scale out the endpoint when the load increases and scale in when the load decreases. This can improve the response time and availability of the product recommendation engine<sup>2</sup>.

The other options are not suitable because:

Option B: Creating a new endpoint configuration with two production variants will not solve the issue of increasing response time and errors. Production variants are used to split the traffic between different models or versions of the same model. They can be useful for testing, updating, or A/B testing models. However, they do not provide any scaling or acceleration benefits for the inference workload<sup>3</sup>.

Option D: Deploying a second instance pool to support a blue/green deployment of models will not solve the issue of increasing response time and errors. Blue/green deployment is a technique for updating models without downtime or disruption. It involves creating a new endpoint configuration with a different instance pool and model version, and then shifting the traffic from the old endpoint to the new endpoint gradually. However, this technique does not provide any scaling or acceleration benefits for the inference workload<sup>4</sup>.

Option E: Reconfiguring the endpoint to use burstable instances will not solve the issue of increasing response time and errors. Burstable instances are instances that provide a baseline level of CPU performance with the ability to burst above the baseline when needed. They can be useful for workloads that have moderate CPU utilization and occasional spikes. However, they are not suitable for workloads that have high and consistent CPU utilization, such as the product recommendation engine. Moreover, burstable instances may incur

additional charges when they exceed their CPU credits<sup>5</sup>.

1: Amazon Elastic Inference

2: How to Scale Amazon SageMaker Endpoints

3: Deploying Models to Amazon SageMaker Hosting Services

4: Updating Models in Amazon SageMaker Hosting Services

5: Burstable Performance Instances

### QUESTION NO: 5

한 회사는 공유된 특성을 기반으로 대규모 고객 그룹을 하위 그룹으로 분류하려고 합니다. 회사의 데이터 과학자는 이 작업을 위해 Amazon SageMaker 내장 k-평균 클러스터링 알고리즘을 사용할 계획입니다. 데이터 과학자는 사용할 최적의 하위 그룹 수(k)를 결정해야 합니다.

k의 최적 값을 가장 정확하게 결정하는 데이터 시각화 접근법은 무엇입니까?

**A.** 주성분 분석(PCA) 성분을 계산합니다. 처음 두 개의 PCA 구성 요소만 사용하여 k 범위에 대해 k-평균 클러스터링 알고리즘을 실행합니다. 각 k 값에 대해 각 클러스터에 대해 서로 다른 색상을 사용하여 산점도를 만듭니다. k의 최적 값은 클러스터가 합리적으로 분리되어 보이기 시작하는 값입니다.

**B.** 주성분 분석(PCA) 성분을 계산합니다. 설명된 분산에 대한 성분 수의 선 그래프를 만듭니다. k의 최적 값은 곡선이 선형 방식으로 감소하기 시작하는 PCA 구성 요소의 수입니다.

**C.** Perplexity 값 범위에 대한 t-분포 확률적 이웃 임베딩(t-SNE) 플롯을 만듭니다. k의 최적 값은 클러스터가 합리적으로 분리된 것처럼 보이기 시작하는 perplexity의 값입니다.

**D.** k 범위에 대해 k-평균 클러스터링 알고리즘을 실행합니다. 각 k 값에 대해 SSE(제곱 오류 합계)를 계산합니다. 각 k 값에 대한 SSE의 선형 차트를 그립니다. k의 최적 값은 곡선이 선형 방식으로 감소하기 시작하는 지점입니다.

**Answer:** D

Explanation:

The solution D is the best data visualization approach to determine the optimal value of k for the k-means clustering algorithm. The solution D involves the following steps:

Run the k-means clustering algorithm for a range of k. For each value of k, calculate the sum of squared errors (SSE). The SSE is a measure of how well the clusters fit the data. It is calculated by summing the squared distances of each data point to its closest cluster center.

A lower SSE indicates a better fit, but it will always decrease as the number of clusters increases. Therefore, the goal is to find the smallest value of k that still has a low SSE<sup>1</sup>.

Plot a line chart of the SSE for each value of k. The line chart will show how the SSE changes as the value of k increases. Typically, the line chart will have a shape of an elbow, where the SSE drops rapidly at first and then levels off. The optimal value of k is the point after which the curve starts decreasing in a linear fashion. This point is also known as the elbow point, and it represents the balance between the number of clusters and the SSE<sup>1</sup>.

The other options are not suitable because:

Option A: Calculating the principal component analysis (PCA) components, running the k-means clustering algorithm for a range of k by using only the first two PCA components, and creating a scatter plot with a different color for each cluster will not accurately determine the optimal value of k. PCA is a technique that reduces the dimensionality of the data by transforming it into a new set of features that capture the most variance in the data. However, PCA may not preserve the original structure and distances of the data, and it may lose some

information in the process. Therefore, running the k-means clustering algorithm on the PCA components may not reflect the true clusters in the data. Moreover, using only the first two PCA components may not capture enough variance to represent the data well. Furthermore, creating a scatter plot may not be reliable, as it depends on the subjective judgment of the data scientist to decide when the clusters look reasonably separated<sup>2</sup>.

Option B: Calculating the PCA components and creating a line plot of the number of components against the explained variance will not determine the optimal value of k. This approach is used to determine the optimal number of PCA components to use for dimensionality reduction, not for clustering. The explained variance is the ratio of the variance of each PCA component to the total variance of the data. The optimal number of PCA components is the point where adding more components does not significantly increase the explained variance. However, this number may not correspond to the optimal number of clusters, as PCA and k-means clustering have different objectives and assumptions<sup>2</sup>.

Option C: Creating a t-distributed stochastic neighbor embedding (t-SNE) plot for a range of perplexity values will not determine the optimal value of k. t-SNE is a technique that reduces the dimensionality of the data by embedding it into a lower-dimensional space, such as a two-dimensional plane. t-SNE preserves the local structure and distances of the data, and it can reveal clusters and patterns in the data. However, t-SNE does not assign labels or centroids to the clusters, and it does not provide a measure of how well the clusters fit the data. Therefore, t-SNE cannot determine the optimal number of clusters, as it only visualizes the data.

Moreover, t-SNE depends on the perplexity parameter, which is a measure of how many neighbors each point considers. The perplexity parameter can affect the shape and size of the clusters, and there is no optimal value for it. Therefore, creating a t-SNE plot for a range of perplexity values may not be consistent or reliable<sup>3</sup>.

1: How to Determine the Optimal K for K-Means?

2: Principal Component Analysis

3: t-Distributed Stochastic Neighbor Embedding

## QUESTION NO: 6

한 회사에 지하철 역의 비디오 피드와 이미지가 있습니다. 회사는 역에 기차가 없을 때 승객이 노란색 안전선을 넘으면 역 관리자에게 경고하는 딥 러닝 모델을 만들고자 합니다. 알림은 비디오 피드를 기반으로 합니다. 회사는 모델이 노란색 선, 노란색 선을 건너는 승객 및 비디오 피드에서 기차를 감지하기를 원합니다. 이 작업에는 레이블이 필요합니다. 비디오 데이터는 기밀로 유지되어야 합니다.

데이터 과학자는 경계 상자를 만들어 샘플 데이터에 레이블을 지정하고 객체 감지 모델을 사용합니다. 그러나 물체 감지 모델은 노란색 선, 노란색 선을 건너는 승객 및 기차를 명확하게 구분할 수 없습니다.

회사에서 이 모델을 개선하는 데 도움이 되는 라벨링 접근 방식은 무엇입니까?

**A.** Amazon Rekognition Custom Labels를 사용하여 데이터 세트에 레이블을 지정하고 사용자 지정 Amazon Rekognition 객체 감지 모델을 생성합니다. 개인 인력을 만듭니다. Amazon Augmented AI(Amazon A2I)를 사용하여 신뢰도가 낮은 예측을 검토하고 사용자 지정 Amazon Rekognition 모델을 재교육하십시오.

**B.** Amazon SageMaker Ground Truth 객체 감지 레이블 지정 작업을 사용합니다. Amazon Mechanical Turk를 라벨링 인력으로 사용하십시오.

C. Amazon Rekognition Custom Labels를 사용하여 데이터 세트에 레이블을 지정하고 사용자 지정 Amazon Rekognition 객체 감지 모델을 생성합니다. 타사 AWS Marketplace 공급업체와 함께 인력을 생성합니다. Amazon Augmented AI(Amazon A2I)를 사용하여 신뢰도가 낮은 예측을 검토하고 사용자 지정 Amazon Rekognition 모델을 재교육하십시오.

D. Amazon SageMaker Ground Truth 의미론적 세분화 레이블 지정 작업을 사용합니다. 라벨링 인력으로 개인 인력을 사용합니다.

**Answer: D**

### QUESTION NO: 7

한 회사가 수천 명의 사람들에게 온라인 객관식 설문 조사를 배포합니다. 설문 조사 응답자는 각 질문에 대해 여러 옵션을 선택할 수 있습니다.

머신 러닝(ML) 엔지니어는 데이터 세트에 있는 모든 응답자의 모든 응답을 종합적으로 표현해야 합니다. ML 엔지니어는 데이터 세트를 사용하여 로지스틱 회귀 모델을 학습합니다. 어떤 솔루션이 이러한 요구 사항을 충족시킬까요?

A. 설문조사의 각 질문에 대한 모든 가능한 옵션에 대해 원핫 인코딩을 수행합니다.

B. 각 질문에 대해 각 응답자가 선택한 모든 답변에 대해 구분을 수행합니다.

C. Amazon Mechanical Turk를 사용하여 가능한 각 응답 세트에 대한 범주형 레이블을 만듭니다.

D. Amazon Textract를 사용하여 가능한 각 응답 세트에 대한 숫자형 특성을 생성합니다.

**Answer: A**

Explanation:

In cases where survey questions allow multiple choices per question, one-hot encoding is an effective way to represent responses as binary features. Each possible option for each question is transformed into a separate binary column (1 if selected, 0 if not), providing a comprehensive and machine-readable format that logistic regression models can interpret effectively.

This approach ensures that each respondent's selections are accurately captured in a format suitable for training, offering a straightforward representation for multi-choice responses.

### QUESTION NO: 8

기계 학습 전문가는 과일 가공 회사에서 일하며 사과를 세 가지 유형으로 분류하는 시스템을 구축해야 합니다. 전문가는 사과의 각 유형에 대한 150개의 이미지가 포함된 데이터 세트를 수집하고 이 데이터 세트로 ImageNet에서 사전 훈련된 신경망에 전이 학습을 적용했습니다. 회사는 모델을 사용하기 위해 최소 85%의 정확도를 요구합니다.

철저한 그리드 검색 후 최적의 하이퍼파라미터는 다음을 생성했습니다.

훈련 세트에서 68% 정확도

검증 세트의 67% 정확도

기계 학습 전문가는 시스템의 정확도를 향상시키기 위해 무엇을 할 수 있습니까?

A. 모델을 Amazon SageMaker 노트북 인스턴스에 업로드하고 Amazon SageMaker HPO 기능을 사용하여 모델의 하이퍼파라미터를 최적화합니다.

B. 훈련 세트에 더 많은 데이터를 추가하고 편향을 줄이기 위해 전이 학습을 사용하여 모델을 다시 훈련합니다.

C. ImageNet에서 사전 훈련된 더 많은 레이어가 있는 신경망 모델을 사용하고 분산을 증가시키기 위해 전이 학습을 적용합니다.

**D.** 현재 신경망 아키텍처를 사용하여 새 모델을 학습합니다.

**Answer: B**

Explanation:

The problem described in the question is a case of underfitting, where the neural network model performs poorly on both the training and validation sets. This means that the model has not learned the features of the data well enough and has high bias. To solve this issue, the machine learning specialist should consider the following change:

Add more data to the training set and retrain the model using transfer learning to reduce the bias: Adding more data to the training set can help the model learn more patterns and variations in the data and improve its performance. Transfer learning can also help the model leverage the knowledge from the pre-trained network and adapt it to the new data. This can reduce the bias and increase the accuracy of the model.

Transfer learning for TensorFlow image classification models in Amazon SageMaker  
Transfer learning for custom labels using a TensorFlow container and "bring your own algorithm" in Amazon SageMaker Machine Learning Concepts - AWS Training and Certification

**QUESTION NO: 9**

데이터 과학자는 고속의 실시간 스트리밍 데이터를 위한 서버리스 수집 및 분석 솔루션을 만들어야 합니다.

수집 프로세스는 데이터 손실 없이 JSON에서 들어오는 레코드를 버퍼링하고 쿼리에 최적화된 열 형식으로 변환해야 합니다. 출력 데이터 저장소는 고가용성이어야 하며 분석가는 데이터에 대해 SQL 쿼리를 실행하고 기존 비즈니스 인텔리전스 대시보드에 연결할 수 있어야 합니다.

데이터 과학자는 요구 사항을 충족하기 위해 어떤 솔루션을 구축해야 합니까?

**A.** 수신 데이터 형식의 AWS Glue 데이터 카탈로그에 스키마를 생성합니다. Amazon Kinesis Data Firehose 전송 스트림을 사용하여 데이터를 스트리밍하고 Amazon S3로 전송하기 전에 AWS Glue 데이터 카탈로그를 사용하여 데이터를 Apache Parquet 또는 ORC 형식으로 변환합니다. 분석가가 Amazon Athena를 사용하여 Amazon S3에서 직접 데이터를 쿼리하고 Athena JDBC(Java Database Connectivity) 커넥터를 사용하여 BI 도구에 연결하도록 합니다.

**B.** 각 JSON 레코드를 Amazon S3의 스테이징 위치에 씁니다. S3 Put 이벤트를 사용하여 데이터를 Apache Parquet 또는 ORC 형식으로 변환하고 Amazon S3의 처리된 데이터 위치에 데이터를 쓰는 AWS Lambda 함수를 트리거합니다. 분석가가 Amazon Athena를 사용하여 Amazon S3에서 직접 데이터를 쿼리하고 Athena JDBC(Java Database Connectivity) 커넥터를 사용하여 BI 도구에 연결하도록 합니다.

**C.** 각 JSON 레코드를 Amazon S3의 스테이징 위치에 씁니다. S3 Put 이벤트를 사용하여 데이터를 Apache Parquet 또는 ORC 형식으로 변환하고 Amazon RDS PostgreSQL 데이터베이스에 삽입하는 AWS Lambda 함수를 트리거합니다. 분석가가 RDS 데이터베이스에서 대시보드를 쿼리하고 실행하도록 합니다.

**D.** Amazon Kinesis Data Analytics를 사용하여 스트리밍 데이터를 수집하고 실시간 SQL 쿼리를 수행하여 Amazon S3로 전송하기 전에 레코드를 Apache Parquet으로 변환합니다. 분석가가 Amazon Athena를 사용하여 Amazon S3에서 직접 데이터를 쿼리하고 Athena JDBC(Java Database Connectivity) 커넥터를 사용하여 BI 도구에 연결하도록 합니다.

**Answer: A**

Explanation:

To create a serverless ingestion and analytics solution for high-velocity, real-time streaming data, the Data Scientist should use the following AWS services:

**AWS Glue Data Catalog:** This is a managed service that acts as a central metadata repository for data assets across AWS and on-premises data sources. The Data Scientist can use AWS Glue Data Catalog to create a schema of the incoming data format, which defines the structure, format, and data types of the JSON records. The schema can be used by other AWS services to understand and process the data<sup>1</sup>.

**Amazon Kinesis Data Firehose:** This is a fully managed service that delivers real-time streaming data to destinations such as Amazon S3, Amazon Redshift, Amazon Elasticsearch Service, and Splunk. The Data Scientist can use Amazon Kinesis Data Firehose to stream the data from the source and transform the data to a query-optimized, columnar format such as Apache Parquet or ORC using the AWS Glue Data Catalog before delivering to Amazon S3. This enables efficient compression, partitioning, and fast analytics on the data<sup>2</sup>.

**Amazon S3:** This is an object storage service that offers high durability, availability, and scalability. The Data Scientist can use Amazon S3 as the output datastore for the transformed data, which can be organized into buckets and prefixes according to the desired partitioning scheme. Amazon S3 also integrates with other AWS services such as Amazon Athena, Amazon EMR, and Amazon Redshift Spectrum for analytics<sup>3</sup>.

**Amazon Athena:** This is a serverless interactive query service that allows users to analyze data in Amazon S3 using standard SQL. The Data Scientist can use Amazon Athena to run SQL queries against the data in Amazon S3 and connect to existing business intelligence dashboards using the Athena Java Database Connectivity (JDBC) connector. Amazon Athena leverages the AWS Glue Data Catalog to access the schema information and supports formats such as Parquet and ORC for fast and cost-effective queries<sup>4</sup>.

1: What Is the AWS Glue Data Catalog? - AWS Glue

2: What Is Amazon Kinesis Data Firehose? - Amazon Kinesis Data Firehose

3: What Is Amazon S3? - Amazon Simple Storage Service

4: What Is Amazon Athena? - Amazon Athena

### QUESTION NO: 10

한 대형 이동통신망 운영 회사가 서비스 해지 가능성이 높은 고객을 예측하는 머신러닝 모델을 구축하고 있습니다. 이 회사는 이러한 고객에게 인센티브를 제공할 계획입니다. 이 탈 비용이 인센티브 비용보다 훨씬 크기 때문입니다.

이 모델은 100명의 고객으로 구성된 테스트 데이터 세트를 평가한 후 다음과 같은 혼동 행렬을 생성합니다.

모델 평가 결과를 바탕으로 볼 때, 이 모델이 생산에 적합한 이유는 무엇입니까?

**A.** 모델의 정확도는 86%이고, 거짓 부정으로 인해 회사가 부담하는 비용은 거짓 긍정으로 인해 발생하는 비용보다 적습니다.

**B.** 모델의 정밀도는 86%로 모델의 정확도보다 낮습니다.

**C.** 모델의 정확도는 86%이고, 거짓 양성으로 인해 회사가 부담하는 비용은 거짓 음성으로 인해 발생하는 비용보다 적습니다.

**D.** 모델의 정밀도는 86%로 모델의 정확도보다 높습니다.

**Answer: C**

Explanation:

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

model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives. The accuracy of the model is the proportion of correct predictions out of the total predictions, which can be calculated by adding the true positives and true negatives and dividing by the total number of observations. In this case, the accuracy of the model is  $(10 + 76) / 100 = 0.86$ , which means that the model correctly predicted

86% of the customers' churn status. The cost incurred by the company as a result of false positives and false negatives is the loss or damage that the company suffers when the model makes incorrect predictions. A false positive is when the model predicts that a customer will churn, but the customer actually does not churn. A false negative is when the model predicts that a customer will not churn, but the customer actually churns. In this case, the cost of a false positive is the incentive that the company offers to the customer who is predicted to churn, which is a relatively low cost. The cost of a false negative is the revenue that the company loses when the customer churns, which is a relatively high cost. Therefore, the cost of a false positive is less than the cost of a false negative, and the company would prefer to have more false positives than false negatives.

The model has 10 false positives and 4 false negatives, which means that the company's cost is lower than if the model had more false negatives and fewer false positives.

#### QUESTION NO: 11

소매 업체에서 기계 학습을 사용하여 신제품을 분류하려고 합니다. 현재 제품의 레이블이 지정된 데이터 세트가 데이터 과학 팀에 제공되었습니다. 데이터 세트에 1,200 개의 제품이 포함되어 있습니다.

제목 크기, 무게 및 가격과 같은 각 제품에 대한 15 가지 기능 각 제품에는 서적, 게임, 전자 제품 및 영화와 같은 6 가지 범주 중 하나에 속하는 것으로 표시되어 있습니다.

교육을 위해 제공된 데이터 세트를 사용하여 신제품을 분류하기 위해 어떤 모델을 사용해야 합니까?

- A. objective 매개 변수가 multi로 설정된 XGBoost 모델 : softmax
- B. 마지막 계층에 대해 softmax 활성화 기능이 있는 심층 컨볼 루션 신경망 (CNN)
- C. 트리 수가 제품 범주 수와 동일하게 설정된 회귀 포리스트
- D. RNN (Recurrent Neural Network) 기반의 DeepAR 예측 모델

**Answer: A**

Explanation:

XGBoost is a machine learning framework that can be used for classification, regression, ranking, and other tasks. It is based on the gradient boosting algorithm, which builds an ensemble of weak learners (usually decision trees) to produce a strong learner. XGBoost has several advantages over other algorithms, such as scalability, parallelization, regularization, and sparsity handling. For categorizing new products using the provided dataset, an XGBoost model would be a suitable choice, because it can handle multiple features and multiple classes efficiently and accurately. To train an XGBoost model for multi-class classification, the objective parameter should be set to multi: softmax, which means that the model will output a probability distribution over the classes and predict the class with the highest probability. Alternatively, the objective parameter can be set to multi: softprob, which means that the model will output the raw probability of each class instead of the predicted class label. This can be useful for evaluating the model performance or for post-processing

the predictions. References:

XGBoost: A tutorial on how to use XGBoost with Amazon SageMaker.

XGBoost Parameters: A reference guide for the parameters of XGBoost.

### QUESTION NO: 12

한 제약 회사는 중요한 발견 사항을 신속하게 해결하기 위해 임상 시험 현장에 대한 정기적인 감사를 수행합니다.

회사는 감사 문서를 텍스트 형식으로 저장합니다. 감사자들은 문서를 신속하게 분석하기 위해 데이터 과학팀에 도움을 요청했습니다. 감사자는 문서 내에서 10가지 주요 주제를 찾아 감사팀 구성원 간에 검토 작업의 우선순위를 정하고 배포해야 합니다. 부작용을 설명하는 문서는 가장 높은 우선순위를 받아야 합니다.

데이터 과학자는 통계 모델링을 사용하여 추상적인 주제를 발견하고 각 범주에 대한 상위 단어 목록을 제공하여 감사자가 주제의 관련성을 평가하는 데 도움을 줍니다.

이 시나리오에 가장 적합한 알고리즘은 무엇입니까? (2개를 선택하세요.)

- A. 잠재 디리클레 할당(LDA)
- B. 랜덤 포레스트 분류기
- C. 신경 주제 모델링(NTM)
- D. 선형 서포트 벡터 머신
- E. 선형 회귀

**Answer:** A C

Explanation:

The algorithms that are best suited to this scenario are latent Dirichlet allocation (LDA) and neural topic modeling (NTM), as they are both unsupervised learning methods that can discover abstract topics from a collection of text documents. LDA and NTM can provide a list of the top words for each topic, as well as the topic distribution for each document, which can help the auditors assess the relevance and priority of the topic<sup>12</sup>.

The other options are not suitable because:

Option B: A random forest classifier is a supervised learning method that can perform classification or regression tasks by using an ensemble of decision trees. A random forest classifier is not suitable for discovering abstract topics from text documents, as it requires labeled data and predefined classes<sup>3</sup>.

Option D: A linear support vector machine is a supervised learning method that can perform classification or regression tasks by using a linear function that separates the data into different classes. A linear support vector machine is not suitable for discovering abstract topics from text documents, as it requires labeled data and predefined classes<sup>4</sup>.

Option E: A linear regression is a supervised learning method that can perform regression tasks by using a linear function that models the relationship between a dependent variable and one or more independent variables. A linear regression is not suitable for discovering abstract topics from text documents, as it requires labeled data and a continuous output variable<sup>5</sup>.

- 1: Latent Dirichlet Allocation
- 2: Neural Topic Modeling
- 3: Random Forest Classifier
- 4: Linear Support Vector Machine
- 5: Linear Regression

**QUESTION NO: 13**

데이터 과학자가 Amazon SageMaker에 내장된 BlazingText 알고리즘을 사용하여 텍스트 분류 모델을 학습하고 있습니다. 데이터 세트에는 5개의 클래스가 있으며, A 범주에는 300개의 샘플, B 범주에는 292개의 샘플이 있습니다.

카테고리 C의 경우 샘플 240개, 카테고리 D의 경우 샘플 258개, 카테고리 E의 경우 샘플 310개.

데이터 과학자는 데이터를 섞고 10%를 테스트용으로 나눕니다. 모델을 학습시킨 후, 데이터 과학자는 학습 세트와 테스트 세트에 대한 혼동 행렬을 생성합니다.

데이터 과학자는 이러한 결과를 통해 어떤 결론을 내릴 수 있을까?

- A. 클래스 C와 D는 너무 비슷합니다.
- B. 데이터 세트가 너무 작아서 홀드아웃 교차 검증에 사용할 수 없습니다.
- C. 데이터 분포가 왜곡되어 있습니다.
- D. 이 모델은 B와 E 클래스에 대해 과적합됩니다.

**Answer: D**

Explanation:

A confusion matrix is a matrix that summarizes the performance of a machine learning model on a set of test data. It displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) produced by the model on the test data<sup>1</sup>. For multi-class classification, the matrix shape will be equal to the number of classes i.e for n classes it will be  $n \times n$ . The diagonal values represent the number of correct predictions for each class, and the off-diagonal values represent the number of incorrect predictions for each class<sup>1</sup>.

The BlazingText algorithm is a proprietary machine learning algorithm for forecasting time series using causal convolutional neural networks (CNNs). BlazingText works best with large datasets containing hundreds of time series. It accepts item metadata, and is the only Forecast algorithm that accepts related time series data without future values<sup>2</sup>.

From the confusion matrices for the training and test sets, we can observe the following:

The model has a high accuracy on the training set, as most of the diagonal values are high and the off-diagonal values are low. This means that the model is able to learn the patterns and features of the training data well.

However, the model has a lower accuracy on the test set, as some of the diagonal values are lower and some of the off-diagonal values are higher. This means that the model is not able to generalize well to the unseen data and makes more errors.

The model has a particularly high error rate for classes B and E on the test set, as the values of  $M_{22}$  and  $M_{55}$  are much lower than the values of  $M_{12}$ ,  $M_{21}$ ,  $M_{15}$ ,  $M_{25}$ ,  $M_{51}$ , and  $M_{52}$ . This means that the model is confusing classes B and E with other classes more often than it should.

The model has a relatively low error rate for classes A, C, and D on the test set, as the values of  $M_{11}$ ,  $M_{33}$ , and  $M_{44}$  are high and the values of  $M_{13}$ ,  $M_{14}$ ,  $M_{23}$ ,  $M_{24}$ ,  $M_{31}$ ,  $M_{32}$ ,  $M_{34}$ ,  $M_{41}$ ,  $M_{42}$ , and  $M_{43}$  are low. This means that the model is able to distinguish classes A, C, and D from other classes well.

These results indicate that the model is overfitting for classes B and E, meaning that it is memorizing the specific features of these classes in the training data, but failing to capture the general features that are applicable to the test data. Overfitting is a common problem in machine learning, where the model performs well on the training data, but poorly on the test

data3. Some possible causes of overfitting are:

The model is too complex or has too many parameters for the given data. This makes the model flexible enough to fit the noise and outliers in the training data, but reduces its ability to generalize to new data.

The data is too small or not representative of the population. This makes the model learn from a limited or biased sample of data, but fails to capture the variability and diversity of the population.

The data is imbalanced or skewed. This makes the model learn from a disproportionate or uneven distribution of data, but fails to account for the minority or rare classes.

Some possible solutions to prevent or reduce overfitting are:

Simplify the model or use regularization techniques. This reduces the complexity or the number of parameters of the model, and prevents it from fitting the noise and outliers in the data. Regularization techniques, such as L1 or L2 regularization, add a penalty term to the loss function of the model, which shrinks the weights of the model and reduces overfitting3. Increase the size or diversity of the data. This provides more information and examples for the model to learn from, and increases its ability to generalize to new data. Data augmentation techniques, such as rotation, flipping, cropping, or noise addition, can generate new data from the existing data by applying some transformations3.

Balance or resample the data. This adjusts the distribution or the frequency of the data, and ensures that the model learns from all classes equally. Resampling techniques, such as oversampling or undersampling, can create a balanced dataset by increasing or decreasing the number of samples for each class3.

Confusion Matrix in Machine Learning - GeeksforGeeks

BlazingText algorithm - Amazon SageMaker

Overfitting and Underfitting in Machine Learning - GeeksforGeeks

#### QUESTION NO: 14

머신러닝(ML) 전문가가 한 회사를 위한 모델을 개발하고 있습니다. 이 모델은 비디오에 표시되는 객체의 시퀀스를 분류하고 예측합니다. ML 전문가는 합성곱 신경망(CNN)과 분류기 3층 순환 신경망(RNN)으로 구성된 하이브리드 아키텍처를 사용하기로 결정했습니다. 이 회사는 이전에도 비슷한 모델을 개발했지만, 다른 객체 세트를 분류하도록 모델을 훈련시켰습니다.

ML 전문가는 이전에 훈련된 모델을 활용하고 해당 모델을 현재의 사용 사례와 객체 세트에 맞게 조정하여 시간을 절약하고자 합니다.

어떤 단계 조합이 가장 적은 노력으로 이 목표를 달성할 수 있을까요? (두 가지를 선택하세요.)

- A. CNN 전체의 가중치를 다시 초기화합니다. 새로운 객체 세트를 사용하여 CNN을 분류 작업으로 다시 학습시킵니다.
- B. 전체 네트워크의 가중치를 다시 초기화합니다. 새로운 객체 세트를 사용하여 전체 네트워크를 예측 작업으로 다시 학습합니다.
- C. 전체 RNN의 가중치를 다시 초기화합니다. 새로운 객체 세트를 사용하여 전체 모델을 예측 작업으로 다시 학습합니다.
- D. CNN의 마지막 완전 연결 계층의 가중치를 다시 초기화합니다. 새로운 객체 세트를 사용하여 CNN을 분류 작업으로 다시 학습시킵니다.
- E. RNN의 마지막 계층의 가중치를 다시 초기화합니다. 새로운 객체 세트를 사용하여 전체 모델을 예측 작업으로 다시 학습합니다.

**Answer:** D E

Explanation:

To adapt a previously trained model to a new but related task efficiently, the best practice is to leverage transfer learning. This involves retaining the learned features from the earlier model and only retraining the final layers to accommodate the new classification categories. In the context of a hybrid architecture combining a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN):

**CNN Component:** The CNN is responsible for extracting spatial features from video frames. Since the early layers of a CNN capture generic features like edges and textures, they are often transferable across tasks.

Therefore, only the last fully connected layer, which maps these features to specific object classes, needs to be reinitialized and retrained for the new set of objects.

**RNN Component:** The RNN handles the temporal dynamics of the sequence data. Similar to the CNN, the earlier layers of the RNN capture general sequence patterns. Thus, reinitializing and retraining only the last layer of the RNN allows the model to adapt to the new prediction task without the need to retrain the entire network.

This approach minimizes training time and computational resources while effectively adapting the model to new tasks.

#### QUESTION NO: 15

기계 학습 전문가는 엔드포인트 자동 조정 구성에 대해 적절한 SageMakerVariant Invocations Per Instance 설정을 결정하려고 합니다. Specialist는 단일 인스턴스에 대한 부하 테스트를 수행했으며 서비스 저하가 없는 최대 RPS(초당 요청 수)는 약 20RPS인 것으로 확인되었습니다. 이것이 첫 번째 배포이므로 Specialist는 호출 안전 계수를 0.5로 설정하려고 합니다. 명시된 매개변수와 인스턴스당 호출 설정이 분 단위로 측정되는 경우 전문가는 sageMakervariantinvocationsPerinstance 설정으로 무엇을 설정해야 합니까?

- A. 10
- B. 30
- C. 600
- D. 2,400

**Answer:** C

Explanation:

The SageMaker Variant Invocations Per Instance setting is the target value for the average number of invocations per instance per minute for the model variant. It is used by the automatic scaling policy to add or remove instances to keep the metric close to the specified value. To determine this value, the following equation can be used in combination with load testing:

$$\text{SageMakerVariantInvocationsPerInstance} = (\text{MAX\_RPS} * \text{SAFETY\_FACTOR}) * 60$$
 Where MAX\_RPS is the maximum requests per second that the model variant can handle without service degradation, SAFETY\_FACTOR is a factor that ensures that the clients do not exceed the maximum RPS, and 60 is the conversion factor from seconds to minutes. In this case, the given parameters are:

$$\text{MAX\_RPS} = 20 \quad \text{SAFETY\_FACTOR} = 0.5$$

Plugging these values into the equation, we get:

$$\text{SageMakerVariantInvocationsPerInstance} = (20 * 0.5) * 60$$

SageMakerVariantInvocationsPerInstance = 600 Therefore, the Specialist should set the SageMaker Variant Invocations Per Instance setting to 600.

Load testing your auto scaling configuration - Amazon SageMaker

Configure model auto scaling with the console - Amazon SageMaker

### QUESTION NO: 16

한 회사에서 기계 학습(ML) 시스템에 대한 감사를 강화하려고 합니다. 감사 시스템은 ML 모델이 사용하는 기능에 대한 메타데이터 분석을 수행할 수 있어야 합니다. 감사 솔루션은 메타데이터를 분석하는 보고서를 생성해야 합니다. 또한 솔루션은 데이터 민감도와 기능 작성자를 설정할 수 있어야 합니다.

최소한의 개발 노력으로 이러한 요구 사항을 충족하는 솔루션은 무엇입니까?

- A.** Amazon SageMaker Feature Store를 사용하여 기능을 선택합니다. 기능 수준 메타데이터 분석을 수행하는 데이터 흐름을 만듭니다. 기능 수준 메타데이터를 저장할 Amazon DynamoDB 테이블을 생성합니다. Amazon QuickSight를 사용하여 메타데이터를 분석합니다.
- B.** Amazon SageMaker Feature Store를 사용하여 ML 모델이 사용하는 현재 기능에 대한 기능 그룹을 설정합니다. 각 기능에 필요한 메타데이터를 할당합니다. SageMaker Studio를 사용하여 메타데이터를 분석합니다.
- C.** Amazon SageMaker Feature Store를 사용하여 사용자 지정 알고리즘을 적용하여 회사에 필요한 기능 수준 메타데이터를 분석합니다. 기능 수준 메타데이터를 저장할 Amazon DynamoDB 테이블을 생성합니다. Amazon QuickSight를 사용하여 메타데이터를 분석합니다.
- D.** Amazon SageMaker Feature Store를 사용하여 ML 모델이 사용하는 현재 기능에 대한 기능 그룹을 설정합니다. 각 기능에 필요한 메타데이터를 할당합니다. Amazon QuickSight를 사용하여 메타데이터를 분석합니다.

**Answer:** D

Explanation:

The solution that will meet the requirements with the least development effort is to use Amazon SageMaker Feature Store to set feature groups for the current features that the ML models use, assign the required metadata for each feature, and use Amazon QuickSight to analyze the metadata. This solution can leverage the existing AWS services and features to perform feature-level metadata analysis and reporting.

Amazon SageMaker Feature Store is a fully managed, purpose-built repository to store, update, search, and share machine learning (ML) features. The service provides feature management capabilities such as enabling easy feature reuse, low latency serving, time travel, and ensuring consistency between features used in training and inference workflows. A feature group is a logical grouping of ML features whose organization and structure is defined by a feature group schema. A feature group schema consists of a list of feature definitions, each of which specifies the name, type, and metadata of a feature. The metadata can include information such as data sensitivity, authorship, description, and parameters. The metadata can help make features discoverable, understandable, and traceable. Amazon SageMaker Feature Store allows users to set feature groups for the current features that the ML models use, and assign the required metadata for each feature using the AWS SDK for Python (Boto3), AWS Command Line Interface (AWS CLI), or Amazon SageMaker Studio1. Amazon QuickSight is a fully managed, serverless business intelligence service that makes it easy to create and publish interactive dashboards that include ML insights. Amazon QuickSight can connect to various data sources, such as Amazon S3, Amazon Athena,

Amazon Redshift, and Amazon SageMaker Feature Store, and analyze the data using standard SQL or built-in ML-powered analytics. Amazon QuickSight can also create rich visualizations and reports that can be accessed from any device, and securely shared with anyone inside or outside an organization. Amazon QuickSight can be used to analyze the metadata of the features stored in Amazon SageMaker Feature Store, and generate a report that summarizes the metadata analysis<sup>2</sup>.

The other options are either more complex or less effective than the proposed solution. Using Amazon SageMaker Data Wrangler to select the features and create a data flow to perform feature-level metadata analysis would require additional steps and resources, and may not capture all the metadata attributes that the company requires. Creating an Amazon DynamoDB table to store feature-level metadata would introduce redundancy and inconsistency, as the metadata is already stored in Amazon SageMaker Feature Store. Using SageMaker Studio to analyze the metadata would not generate a report that can be easily shared and accessed by the company.

1: Amazon SageMaker Feature Store - Amazon Web Services

2: Amazon QuickSight - Business Intelligence Service - Amazon Web Services

#### QUESTION NO: 17

한 회사에서 많은 수의 구조화되지 않은 종이 영수증을 이미지로 변환하고 있습니다. 회사는 날짜, 위치, 메모와 같은 관련 엔터티와 영수증 번호와 같은 일부 사용자 지정 엔터티를 찾기 위해 자연어 처리(NLP)를 기반으로 하는 모델을 생성하려고 합니다.

이 회사는 광학 문자 인식(OCR)을 사용하여 데이터 라벨링을 위한 텍스트를 추출하고 있습니다. 그러나 문서의 구조와 형식이 다르기 때문에 회사는 문서 유형별로 수동 워크플로를 설정하는 데 어려움을 겪고 있습니다. 또한 회사는 작은 샘플 크기를 사용하여 사용자 지정 엔터티 감지를 위해 명명된 엔터티 인식(NER) 모델을 훈련했습니다. 이 모델은 신뢰도 점수가 매우 낮으며 대규모 데이터 세트로 재학습해야 합니다.

텍스트 추출 및 엔터티 감지를 위한 솔루션 중 가장 적은 노력이 필요한 솔루션은 무엇입니까?

**A.** Amazon Textract를 사용하여 영수증 이미지에서 텍스트를 추출합니다. Amazon SageMaker BlazingText 알고리즘을 사용하여 엔터티 및 사용자 지정 엔터티에 대한 텍스트를 학습합니다.

**B.** AWS Marketplace의 딥 러닝 OCR 모델을 사용하여 영수증 이미지에서 텍스트를 추출합니다. NER 딥 러닝 모델을 사용하여 엔터티를 추출합니다.

**C.** Amazon Textract를 사용하여 영수증 이미지에서 텍스트를 추출합니다. 엔터티 감지에는 Amazon Comprehend를 사용하고 사용자 지정 엔터티 감지에는 Amazon Comprehend 사용자 지정 엔터티 인식을 사용합니다.

**D.** AWS Marketplace의 딥러닝 OCR 모델을 사용하여 영수증 이미지에서 텍스트를 추출합니다. 엔터티 감지에는 Amazon Comprehend를 사용하고 사용자 지정 엔터티 감지에는 Amazon Comprehend 사용자 지정 엔터티 인식을 사용합니다.

**Answer: C**

Explanation:

The best solution for text extraction and entity detection with the least amount of effort is to use Amazon Textract and Amazon Comprehend. These services are:

Amazon Textract for text extraction from receipt images. Amazon Textract is a machine learning service that can automatically extract text and data from scanned documents. It can handle different structures and formats of documents, such as PDF, TIFF, PNG, and JPEG,

without any preprocessing steps. It can also extract key-value pairs and tables from documents. Amazon Comprehend for entity detection and custom entity detection. Amazon Comprehend is a natural language processing service that can identify entities, such as dates, locations, and notes, from unstructured text. It can also detect custom entities, such as receipt numbers, by using a custom entity recognizer that can be trained with a small amount of labeled data. The other options are not suitable because they either require more effort for text extraction, entity detection, or custom entity detection. For example:

Option A uses the Amazon SageMaker BlazingText algorithm to train on the text for entities and custom entities. BlazingText is a supervised learning algorithm that can perform text classification and word2vec. It requires users to provide a large amount of labeled data, preprocess the data into a specific format, and tune the hyperparameters of the model. Option B uses a deep learning OCR model from the AWS Marketplace and a NER deep learning model for text extraction and entity detection. These models are pre-trained and may not be suitable for the specific use case of receipt processing. They also require users to deploy and manage the models on Amazon SageMaker or Amazon EC2 instances. Option D uses a deep learning OCR model from the AWS Marketplace for text extraction. This model has the same drawbacks as option B. It also requires users to integrate the model output with Amazon Comprehend for entity detection and custom entity detection.

- 1: Amazon Textract - Extract text and data from documents
- 2: Amazon Comprehend - Natural Language Processing (NLP) and Machine Learning (ML)
- 3: BlazingText - Amazon SageMaker
- 4: AWS Marketplace: OCR

**QUESTION NO: 18**

한 회사는 부동산 가격을 예측하기 위해 회사 웹사이트에 기계 학습(ML) 모델을 배포했습니다.

배포 후 몇 달이 지난 후 ML 엔지니어는 모델의 정확도가 점차 감소하는 것을 발견했습니다. ML 엔지니어는 모델의 정확도를 향상해야 합니다. 엔지니어는 향후 성능 문제에 대한 알림도 받아야 합니다.

어떤 솔루션이 이러한 요구 사항을 충족합니까?

- A. 증분 학습을 수행하여 모델을 업데이트합니다. Amazon SageMaker 모델 모니터를 활성화하여 모델 성능 문제를 감지하고 알림을 보냅니다.
- B. Amazon SageMaker 모델 거버넌스를 사용합니다. 모델 하이퍼 매개변수를 자동으로 조정하도록 모델 거버넌스를 구성합니다. Amazon CloudWatch에서 성능 임계값 경보를 생성하여 알림을 보냅니다.
- C. 적절한 임계값과 함께 Amazon SageMaker 디버거를 사용합니다. 팀에 알리기 위해 Amazon CloudWatch 경보를 보내도록 디버거를 구성합니다. 이전 몇 달 동안의 데이터만 사용하여 모델을 재교육합니다.
- D. 이전 몇 달 동안의 데이터만 사용하여 증분 훈련을 수행하여 모델을 업데이트합니다. Amazon SageMaker 모델 모니터를 사용하여 모델 성능 문제를 감지하고 알림을 보냅니다.

**Answer: A**

Explanation:

The best solution to improve the accuracy of the model and receive notifications for any future performance issues is to perform incremental training to update the model and activate Amazon SageMaker Model Monitor to detect model performance issues and to send

notifications. Incremental training is a technique that allows you to update an existing model with new data without retraining the entire model from scratch. This can save time and resources, and help the model adapt to changing data patterns. Amazon SageMaker Model Monitor is a feature that continuously monitors the quality of machine learning models in production and notifies you when there are deviations in the model quality, such as data drift and anomalies. You can set up alerts that trigger actions, such as sending notifications to Amazon Simple Notification Service (Amazon SNS) topics, when certain conditions are met. Option B is incorrect because Amazon SageMaker Model Governance is a set of tools that help you implement ML responsibly by simplifying access control and enhancing transparency. It does not provide a mechanism to automatically adjust model hyperparameters or improve model accuracy.

Option C is incorrect because Amazon SageMaker Debugger is a feature that helps you debug and optimize your model training process by capturing relevant data and providing real-time analysis. However, using Debugger alone does not update the model or monitor its performance in production. Also, retraining the model by using only data from the previous several months may not capture the full range of data variability and may introduce bias or overfitting.

Option D is incorrect because using only data from the previous several months to perform incremental training may not be sufficient to improve the model accuracy, as explained above. Moreover, this option does not specify how to activate Amazon SageMaker Model Monitor or configure the alerts and notifications.

Incremental training

Amazon SageMaker Model Monitor

Amazon SageMaker Model Governance

Amazon SageMaker Debugger

### QUESTION NO: 19

네트워크 보안 공급업체는 전 세계에서 실행되는 수천 개의 엔드포인트에서 원격 측정 데이터를 수집해야 합니다. 데이터는 50개의 필드를 포함하는 레코드 형태로 30초마다 전송됩니다. 각 레코드의 크기는 최대 1KB입니다. 보안 공급업체는 Amazon Kinesis Data Streams를 사용하여 데이터를 수집합니다. 공급업체는 Kinesis Data Streams가 수집하는 레코드에 대한 시간별 요약은 요구합니다. 공급업체는 Amazon Athena를 사용하여 기록을 쿼리하고 요약을 생성합니다. Athena 쿼리는 사용 가능한 데이터 필드 중 7~12개를 대상으로 합니다.

수집된 데이터를 변환하고 저장하기 위해 최소한의 사용자 정의로 이러한 요구 사항을 충족하는 솔루션은 무엇입니까?

- A. AWS Lambda를 사용하여 매시간 데이터를 읽고 집계합니다. Amazon Kinesis Data Firehose를 사용하여 데이터를 변환하고 Amazon S3에 저장합니다.
- B. Amazon Kinesis Data Firehose를 사용하여 매시간 데이터를 읽고 집계합니다. 단기 Amazon EMR 클러스터를 사용하여 데이터를 변환하고 Amazon S3에 저장합니다.
- C. Amazon Kinesis Data Analytics를 사용하여 매시간 데이터를 읽고 집계합니다. Amazon Kinesis Data Firehose를 사용하여 데이터를 변환하고 Amazon S3에 저장합니다.
- D. Amazon Kinesis Data Firehose를 사용하여 매시간 데이터를 읽고 집계합니다. AWS Lambda를 사용하여 데이터를 변환하고 Amazon S3에 저장합니다.

**Answer: C**

**Explanation:**

The solution that will meet the requirements with the least amount of customization to transform and store the ingested data is to use Amazon Kinesis Data Analytics to read and aggregate the data hourly, transform the data and store it in Amazon S3 by using Amazon Kinesis Data Firehose. This solution leverages the built-in features of Kinesis Data Analytics to perform SQL queries on streaming data and generate hourly summaries.

Kinesis Data Analytics can also output the transformed data to Kinesis Data Firehose, which can then deliver the data to S3 in a specified format and partitioning scheme. This solution does not require any custom code or additional infrastructure to process the data. The other solutions either require more customization (such as using Lambda or EMR) or do not meet the requirement of aggregating the data hourly (such as using Lambda to read the data from Kinesis Data Streams). References:

- 1: Boosting Resiliency with an ML-based Telemetry Analytics Architecture | AWS Architecture Blog
- 2: AWS Cloud Data Ingestion Patterns and Practices
- 3: IoT ingestion and Machine Learning analytics pipeline with AWS IoT ...
- 4: AWS IoT Data Ingestion Simplified 101: The Complete Guide - Hevo Data

**QUESTION NO: 20**

기계 학습 전문가는 기계 학습 파이프라인을 통해 처리되는 모든 고객 방문에 대해 분석을 실행하려는 온라인 소매업체에서 일하고 있습니다. Amazon Kinesis Data Streams는 초당 최대 100개의 트랜잭션으로 데이터를 수집해야 하며 JSON 데이터 Blob의 크기는 100KB입니다.

전문가가 이 데이터를 성공적으로 수집하기 위해 사용해야 하는 Kinesis Data Streams의 최소 샤드 수는 얼마입니까?

- A. 1 샤드
- B. 파편 10개
- C. 100 샤드
- D. 1,000 샤드

**Answer: A**

**Explanation:**

According to the Amazon Kinesis Data Streams documentation, the maximum size of data blob (the data payload before Base64-encoding) per record is 1 MB. The maximum number of records that can be sent to a shard per second is 1,000. Therefore, the maximum throughput of a shard is 1 MB/sec for input and 2 MB/sec for output. In this case, the input throughput is 100 transactions per second \* 100 KB per transaction = 10 MB /sec. Therefore, the minimum number of shards required is 10 MB/sec / 1 MB/sec = 10 shards. However, the question asks for the minimum number of shards in Kinesis Data Streams, not the minimum number of shards per stream. A Kinesis Data Streams account can have multiple streams, each with its own number of shards.

Therefore, the minimum number of shards in Kinesis Data Streams is 1, which is the minimum number of shards per stream. References:

- Amazon Kinesis Data Streams Terminology and Concepts
- Amazon Kinesis Data Streams Limits

**QUESTION NO: 21**

회사는 매일 수백만 건의 주문을 처리합니다. 회사는 Amazon DynamoDB 테이블을 사용하여 주문 정보를 저장합니다. 고객이 새 주문을 제출하면 새 주문이 즉시 DynamoDB 테이블에 추가됩니다. 새로운 주문이 DynamoDB 테이블에 지속적으로 도착합니다.

데이터 과학자는 피크타임 예측 솔루션을 구축해야 합니다. 또한 데이터 과학자는 거의 실시간 주문 통찰력을 표시하기 위해 Amazon QuickSight 대시보드를 생성해야 합니다. 데이터 과학자는 새로운 주문 정보가 도착하자마자 QuickSight가 데이터에 액세스할 수 있는 솔루션을 구축해야 합니다.

새로운 주문이 처리되는 시점과 QuickSight가 새로운 주문 정보에 액세스할 수 있는 시점 사이의 지연 시간을 최소화하면서 이러한 요구 사항을 충족하는 솔루션은 무엇입니까?

**A.** AWS Glue를 사용하여 Amazon DynamoDB에서 Amazon S3로 데이터를 내보냅니다. Amazon S3의 데이터에 액세스하도록 QuickSight를 구성합니다.

**B.** Amazon Kinesis Data Streams를 사용하여 Amazon DynamoDB에서 Amazon S3로 데이터를 내보냅니다.

Amazon S3의 데이터에 액세스하도록 QuickSight를 구성합니다.

**C.** QuickSight의 API 호출을 사용하여 Amazon DynamoDB에 있는 데이터에 직접 액세스합니다.

**D.** Amazon Kinesis Data Firehose를 사용하여 Amazon DynamoDB에서 Amazon S3로 데이터를 내보냅니다. QuickSight를 구성하여 Amazon S3의 데이터에 액세스합니다.

**Answer: B**

Explanation:

The best solution for this scenario is to use Amazon Kinesis Data Streams to export the data from Amazon DynamoDB to Amazon S3, and then configure QuickSight to access the data in Amazon S3. This solution has the following advantages:

It allows near real-time data ingestion from DynamoDB to S3 using Kinesis Data Streams, which can capture and process data continuously and at scale<sup>1</sup>.

It enables QuickSight to access the data in S3 using the Athena connector, which supports federated queries to multiple data sources, including Kinesis Data Streams<sup>2</sup>.

It avoids the need to create and manage a Lambda function or a Glue crawler, which are required for the other solutions.

The other solutions have the following drawbacks:

Using AWS Glue to export the data from DynamoDB to S3 introduces additional latency and complexity, as Glue is a batch-oriented service that requires scheduling and configuration<sup>3</sup>.

Using an API call from QuickSight to access the data in DynamoDB directly is not possible, as QuickSight does not support direct querying of DynamoDB<sup>4</sup>.

Using Kinesis Data Firehose to export the data from DynamoDB to S3 is less efficient and flexible than using Kinesis Data Streams, as Firehose does not support custom data processing or transformation, and has a minimum buffer interval of 60 seconds<sup>5</sup>.

1: Amazon Kinesis Data Streams - Amazon Web Services

2: Visualize Amazon DynamoDB insights in Amazon QuickSight using the Amazon Athena DynamoDB connector and AWS Glue | AWS Big Data Blog

3: AWS Glue - Amazon Web Services

4: Visualising your Amazon DynamoDB data with Amazon QuickSight - DEV Community

5: Amazon Kinesis Data Firehose - Amazon Web Services

**QUESTION NO: 22**

기계 학습 전문가는 미디어 회사와 협력하여 회사 웹 사이트의 인기 기사에 대한 분류를 수행하고 있습니다. 회사는 무작위 포리스트를 사용하여 기사가 게시되기 전에 얼마나 인기가 있을지 분류하고 있습니다. 사용 중인 데이터의 샘플은 다음과 같습니다. 데이터 세트가 주어지면 Specialist는 Day-Of\_Week 열을 이진 값으로 변환하려고 합니다. 이 열을 이진 값으로 변환하려면 어떤 기술을 사용해야 합니다.

Article_Title	Author	Top_Keywords	Day_Of_Week	URL_of_Article	Page_Views
Building a Big Data Platform	Jane Doe	Big Data, Spark, Hadoop	Tuesday	http://examplecorp.com/data_platform.html	1300456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	http://examplecorp.com/started_deep_learning.html	1230661
MXNet ML Guide	Jane Doe	Machine Learning, MXNet, Logistic Regression	Thursday	http://examplecorp.com/mxnet_guide.html	937291
Intro to NoSQL Databases	Mary Major	NoSQL, Operations, Database	Monday	http://examplecorp.com/nosql_intro_guide.html	407812

- A. 이진화
- B. 원-핫 인코딩
- C. 토큰화
- D. 정규화 변환

**Answer: B**

Explanation:

One-hot encoding is a technique that can be used to convert a categorical variable, such as the Day-Of\_Week column, to binary values. One-hot encoding creates a new binary column for each unique value in the original column, and assigns a value of 1 to the column that corresponds to the value in the original column, and 0 to the rest. For example, if the original column has values Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday, one-hot encoding will create seven new columns, each representing one day of the week. If the value in the original column is Tuesday, then the column for Tuesday will have a value of 1, and the other columns will have a value of 0. One-hot encoding can help improve the performance of machine learning models, as it eliminates the ordinal relationship between the values and creates a more informative and sparse representation of the data.

One-Hot Encoding - Amazon SageMaker

One-Hot Encoding: A Simple Guide for Beginners | by Jana Schmidt ...

One-Hot Encoding in Machine Learning | by Nishant Malik | Towards ...

**QUESTION NO: 23**

관광 회사는 머신 러닝(ML) 모델을 사용하여 고객에게 추천을 합니다. 이 회사는 Amazon SageMaker 환경을 사용하고 하이퍼파라미터 튜닝 완료 기준을 MaxNumberOfTrainingJobs로 설정합니다.

ML 전문가는 하이퍼파라미터 튜닝 완료 기준을 변경하려고 합니다. ML 전문가는 내부 알고리즘이 튜닝 작업이 최상의 학습 작업의 목표 지표보다 1% 이상 개선될 가능성이 없다고

판단한 직후 튜닝을 중단하려고 합니다.

이 요구 사항을 충족하려면 어떤 완료 기준이 필요합니까?

- A. 최대 실행 시간 초
- B. 대상 목표 메트릭 값
- C. CompleteOnConvergence
- D. 최대 훈련 작업 수 개선되지 않음

**Answer: C**

Explanation:

In Amazon SageMaker, hyperparameter tuning jobs optimize model performance by adjusting hyperparameters. Amazon SageMaker's hyperparameter tuning supports completion criteria settings that enable efficient management of tuning resources. In this scenario, the ML specialist aims to set a completion criterion that will terminate the tuning job as soon as SageMaker detects that further improvements in the objective metric are unlikely to exceed 1%.

The CompleteOnConvergence setting is designed for such requirements. This criterion enables the tuning job to automatically stop when SageMaker determines that additional hyperparameter evaluations are unlikely to improve the objective metric beyond a certain threshold, allowing for efficient tuning completion. The convergence process relies on an internal optimization algorithm that continuously evaluates the objective metric during tuning and stops when performance stabilizes without further improvement.

This is supported by AWS documentation, which explains that CompleteOnConvergence is an efficient way to manage tuning by stopping unnecessary evaluations once the model performance stabilizes within the specified threshold.

#### QUESTION NO: 24

회사는 방대한 양의 데이터를 빠르게 이해하고 이를 통해 통찰력을 얻어야 합니다. 데이터는 다양한 형식이며, 스키마는 자주 변경되고, 새로운 데이터 소스가 정기적으로 추가됩니다. 회사는 AWS 서비스를 사용하여 여러 데이터 소스를 탐색하고, 스키마를 제안하고, 데이터를 풍부하게 하고 변환하고자 합니다. 솔루션은 데이터 흐름에 대한 최소한의 코딩 노력과 최소한의 인프라 관리가 필요합니다.

이러한 요구 사항을 충족하려면 어떤 AWS 서비스 조합이 필요할까요?

- A. 데이터 검색, 강화 및 변환을 위한 Amazon EMR  
표준 SQL을 사용하여 Amazon S3에서 결과를 쿼리하고 분석하기 위한 Amazon Athena, 보고 및 통찰력 얻기 위한 Amazon QuickSight
- B. 데이터 수집을 위한 Amazon Kinesis Data Analytics  
데이터 검색, 강화 및 변환을 위한 Amazon EMR  
Amazon S3에서 결과를 쿼리하고 분석하기 위한 Amazon Redshift
- C. 데이터 검색, 강화 및 변환을 위한 AWS Glue  
표준 SQL을 사용하여 Amazon S3에서 결과를 쿼리하고 분석하기 위한 Amazon Athena, 보고 및 통찰력 얻기 위한 Amazon QuickSight
- D. 데이터 전송을 위한 AWS 데이터 파이프라인, 데이터 검색, 강화 및 변환을 위한 AWS Lambda 작업을 조정하기 위한 AWS Step Functions, 표준 SQL을 사용하여 Amazon S3에서 결과를 쿼리하고 분석하기 위한 Amazon Athena, 보고 및 통찰력 얻기 위한 Amazon QuickSight

**Answer: C**

Explanation:

The best combination of AWS services to meet the requirements of data discovery, enrichment, transformation, querying, analysis, and reporting with the least coding and infrastructure management is AWS Glue, Amazon Athena, and Amazon QuickSight. These services are:

AWS Glue for data discovery, enrichment, and transformation. AWS Glue is a serverless data integration service that automatically crawls, catalogs, and prepares data from various sources and formats. It also provides a visual interface called AWS Glue DataBrew that allows users to apply over 250 transformations to clean, normalize, and enrich data without writing code<sup>1</sup> Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL. Amazon Athena is a serverless interactive query service that allows users to analyze data in Amazon S3 using standard SQL. It supports a variety of data formats, such as CSV, JSON, ORC, Parquet, and Avro. It also integrates with AWS Glue Data Catalog to provide a unified view of the data sources and schemas<sup>2</sup> Amazon QuickSight for reporting and getting insights. Amazon QuickSight is a serverless business intelligence service that allows users to create and share interactive dashboards and reports. It also provides ML-powered features, such as anomaly detection, forecasting, and natural language queries, to help users discover hidden insights from their data<sup>3</sup> The other options are not suitable because they either require more coding effort, more infrastructure management, or do not support the desired use cases. For example:

Option A uses Amazon EMR for data discovery, enrichment, and transformation. Amazon EMR is a managed cluster platform that runs Apache Spark, Apache Hive, and other open-source frameworks for big data processing. It requires users to write code in languages such as Python, Scala, or SQL to perform data integration tasks. It also requires users to provision, configure, and scale the clusters according to their needs<sup>4</sup> Option B uses Amazon Kinesis Data Analytics for data ingestion. Amazon Kinesis Data Analytics is a service that allows users to process streaming data in real time using SQL or Apache Flink. It is not suitable for data discovery, enrichment, and transformation, which are typically batch-oriented tasks. It also requires users to write code to define the data processing logic and the output destination<sup>5</sup> Option D uses AWS Data Pipeline for data transfer and AWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformation. AWS Data Pipeline is a service that helps users move data between AWS services and on-premises data sources. AWS Step Functions is a service that helps users coordinate multiple AWS services into workflows. AWS Lambda is a service that lets users run code without provisioning or managing servers. These services require users to write code to define the data sources, destinations, transformations, and workflows. They also require users to manage the scalability, performance, and reliability of the data pipelines.

1: AWS Glue - Data Integration Service - Amazon Web Services

2: Amazon Athena - Interactive SQL Query Service - AWS

3: Amazon QuickSight - Business Intelligence Service - AWS

4: Amazon EMR - Amazon Web Services

5: Amazon Kinesis Data Analytics - Amazon Web Services

AWS Data Pipeline - Amazon Web Services

AWS Step Functions - Amazon Web Services

## AWS Lambda - Amazon Web Services

### QUESTION NO: 25

기계 학습 팀은 Amazon SageMaker에서 자체 교육 알고리즘을 실행합니다. 학습 알고리즘에는 외부 자산이 필요합니다. 팀은 자체 알고리즘 코드와 알고리즘별 파라미터를 모두 Amazon SageMaker에 제출해야 합니다.

팀은 Amazon SageMaker에서 사용자 지정 알고리즘을 구축하기 위해 어떤 서비스 조합을 사용해야 합니까?

(2개를 선택하세요.)

- A. AWS Secrets Manager
- B. AWS 코드스타
- C. 아마존 ECR
- D. 아마존 ECS
- E. 아마존 S3

**Answer:** C E

Explanation:

The Machine Learning team wants to use its own training algorithm on Amazon SageMaker, and submit both its own algorithm code and algorithm-specific parameters. The best combination of services to build a custom algorithm in Amazon SageMaker are Amazon ECR and Amazon S3.

Amazon ECR is a fully managed container registry service that allows you to store, manage, and deploy Docker container images. You can use Amazon ECR to create a Docker image that contains your training algorithm code and any dependencies or libraries that it requires. You can also use Amazon ECR to push, pull, and manage your Docker images securely and reliably.

Amazon S3 is a durable, scalable, and secure object storage service that can store any amount and type of data. You can use Amazon S3 to store your training data, model artifacts, and algorithm-specific parameters.

You can also use Amazon S3 to access your data and parameters from your training algorithm code, and to write your model output to a specified location.

Therefore, the Machine Learning team can use the following steps to build a custom algorithm in Amazon SageMaker:

Write the training algorithm code in Python, using the Amazon SageMaker Python SDK or the Amazon SageMaker Containers library to interact with the Amazon SageMaker service. The code should be able to read the input data and parameters from Amazon S3, and write the model output to Amazon S3.

Create a Dockerfile that defines the base image, the dependencies, the environment variables, and the commands to run the training algorithm code. The Dockerfile should also expose the ports that Amazon SageMaker uses to communicate with the container.

Build the Docker image using the Dockerfile, and tag it with a meaningful name and version. Push the Docker image to Amazon ECR, and note the registry path of the image.

Upload the training data, model artifacts, and algorithm-specific parameters to Amazon S3, and note the S3 URIs of the objects.

Create an Amazon SageMaker training job, using the Amazon SageMaker Python SDK or the AWS CLI.

Specify the registry path of the Docker image, the S3 URIs of the input and output data, the algorithm-specific parameters, and other configuration options, such as the instance type, the number of instances, the IAM role, and the hyperparameters.

Monitor the status and logs of the training job, and retrieve the model output from Amazon S3.

Use Your Own Training Algorithms

Amazon ECR - Amazon Web Services

Amazon S3 - Amazon Web Services

### QUESTION NO: 26

기계 학습(ML) 전문가는 Amazon SageMaker DeepAR 예측 알고리즘을 사용하여 CPU 기반 Amazon EC2 온디맨드 인스턴스에서 모델을 교육하고 있습니다. 현재 모델을 훈련하는 데 여러 시간이 걸립니다.

ML 전문가는 모델의 학습 시간을 줄이고 싶어합니다.

이 요구 사항을 충족하는 접근 방식?

- A. 온디맨드 인스턴스를 스팟 인스턴스로 교체
- B. 모델 자동 확장을 동적으로 구성하여 인스턴스 수를 자동으로 조정합니다.
- C. CPU 기반 EC2 인스턴스를 GPU 기반 EC2 인스턴스로 교체합니다.
- D. 여러 훈련 인스턴스를 사용합니다.
- E. 사전 훈련된 모델 버전을 사용합니다. 증분 훈련을 실행합니다.

**Answer:** C D

Explanation:

The best approaches to decrease the training time of the model are C and D, because they can improve the computational efficiency and parallelization of the training process. These approaches have the following benefits:

C: Replacing CPU-based EC2 instances with GPU-based EC2 instances can speed up the training of the DeepAR algorithm, as it can leverage the parallel processing power of GPUs to perform matrix operations and gradient computations faster than CPUs<sup>12</sup>. The DeepAR algorithm supports GPU-based EC2 instances such as ml.p2 and ml.p3<sup>3</sup>.

D: Using multiple training instances can also reduce the training time of the DeepAR algorithm, as it can distribute the workload across multiple nodes and perform data parallelism<sup>4</sup>. The DeepAR algorithm supports distributed training with multiple CPU-based or GPU-based EC2 instances<sup>3</sup>.

The other options are not effective or relevant, because they have the following drawbacks:

A: Replacing On-Demand Instances with Spot Instances can reduce the cost of the training, but not necessarily the time, as Spot Instances are subject to interruption and availability<sup>5</sup>.

Moreover, the DeepAR algorithm does not support checkpointing, which means that the training cannot resume from the last saved state if the Spot Instance is terminated<sup>3</sup>.

B: Configuring model auto scaling dynamically to adjust the number of instances automatically is not applicable, as this feature is only available for inference endpoints, not for training jobs<sup>6</sup>.

E: Using a pre-trained version of the model and running incremental training is not possible, as the DeepAR algorithm does not support incremental training or transfer learning<sup>3</sup>. The DeepAR algorithm requires a full retraining of the model whenever new data is added or the hyperparameters are changed<sup>7</sup>.

- 1: GPU vs CPU: What Matters Most for Machine Learning? | by Louis (What's AI) Bouchard | Towards Data Science
- 2: How GPUs Accelerate Machine Learning Training | NVIDIA Developer Blog
- 3: DeepAR Forecasting Algorithm - Amazon SageMaker
- 4: Distributed Training - Amazon SageMaker
- 5: Managed Spot Training - Amazon SageMaker
- 6: Automatic Scaling - Amazon SageMaker
- 7: How the DeepAR Algorithm Works - Amazon SageMaker

### QUESTION NO: 27

데이터 과학자가 선형 회귀 모델을 구축하고 있습니다. 과학자는 데이터세트를 조사한 후 분포 최빈값이 중앙값보다 낮고 중앙값이 평균보다 낮다는 것을 확인했습니다. 데이터 과학자에게 선형 회귀 모델을 적용할 수 있는 능력을 제공하는 데이터 변환은 무엇입니까?

- A. 지수변환
- B. 로그 변환
- C. 다항식 변환
- D. 정현파 변환

**Answer: B**

Explanation:

A logarithmic transformation is a suitable data transformation for a linear regression model when the data has a skewed distribution, such as when the mode is lower than the median and the median is lower than the mean. A logarithmic transformation can reduce the skewness and make the data more symmetric and normally distributed, which are desirable properties for linear regression. A logarithmic transformation can also reduce the effect of outliers and heteroscedasticity (unequal variance) in the data. An exponential transformation would have the opposite effect of increasing the skewness and making the data more asymmetric. A polynomial transformation may not be able to capture the nonlinearity in the data and may introduce multicollinearity among the transformed variables. A sinusoidal transformation is not appropriate for data that does not have a periodic pattern.

Data Transformation - Scaler Topics

Linear Regression - GeeksforGeeks

Linear Regression - Scribbr

### QUESTION NO: 28

한 은행에서 저금리 신용 프로모션을 시작하려고 합니다. 그 은행은 최근 경제적 어려움을 겪은 마을에 위치해 있다. 은행 고객 중 일부만이 위기의 영향을 받았으므로 은행 신용팀은 프로모션 대상 고객을 식별해야 합니다. 그러나 신용팀은 결정을 내릴 때 단골 고객의 전체 신용 기록이 고려되기를 원합니다.

은행의 데이터 과학 팀은 계좌 거래를 분류하고 신용 자격을 이해하는 모델을 개발했습니다.

데이터 과학 팀은 XGBoost 알고리즘을 사용하여 모델을 훈련했습니다. 팀은 며칠에 걸쳐 훈련 및 초매개변수 조정을 위해 7년간의 은행 거래 내역 데이터를 사용했습니다.

모델의 정확도는 충분하지만 신용팀은 모델이 일부 고객에 대한 신용을 거부하는 이유를 정확하게 설명하기 위해 고군분투하고 있습니다. 신용팀은 데이터 과학에 대한 기술이 거의 없습니다.

가장 운영상 효율적인 방식으로 이 문제를 해결하기 위해 데이터 과학 팀은 무엇을 해야 합니까?

- A.** Amazon SageMaker Studio를 사용하여 모델을 재구축합니다. XGBoost 교육 컨테이너를 사용하여 모델 교육을 수행하는 노트북을 만듭니다. 엔드포인트에 모델을 배포합니다. 추론을 저장하려면 Amazon SageMaker 모델 모니터를 활성화하세요. 추론을 사용하여 모델 동작을 설명하는 데 도움이 되는 Shapley 값을 생성합니다. 기능과 SHAP(SHapley Additive exPlanations) 값을 표시하는 차트를 만들어 기능이 모델 결과에 어떻게 영향을 미치는지 신용팀에 설명하세요.
- B.** Amazon SageMaker Studio를 사용하여 모델을 재구축합니다. XGBoost 교육 컨테이너를 사용하여 모델 교육을 수행하는 노트북을 만듭니다. Amazon SageMaker 디버거를 활성화하고 Shapley 값을 계산하고 수집하도록 구성합니다. 기능과 SHAP(SHapley Additive exPlanations) 값을 표시하는 차트를 만들어 기능이 모델 결과에 어떻게 영향을 미치는지 신용팀에 설명하세요.
- C.** Amazon SageMaker 노트북 인스턴스를 생성합니다. 노트북 인스턴스와 XGBoost 라이브러리를 사용하여 모델을 로컬에서 재교육합니다. Python XGBoost 인터페이스의plot\_importance() 메서드를 사용하여 기능 중요도 차트를 만듭니다. 해당 차트를 사용하여 기능이 모델 결과에 어떤 영향을 미치는지 신용 팀에 설명하십시오.
- D.** Amazon SageMaker Studio를 사용하여 모델을 재구축합니다. XGBoost 교육 컨테이너를 사용하여 모델 교육을 수행하는 노트북을 만듭니다. 엔드포인트에 모델을 배포합니다. Amazon SageMaker 처리를 사용하여 모델을 사후 분석하고 신용 팀을 위한 기능 중요도 설명 차트를 자동으로 생성합니다.

**Answer: A**

Explanation:

The best option is to use Amazon SageMaker Studio to rebuild the model and deploy it at an endpoint. Then, use Amazon SageMaker Model Monitor to store inferences and use the inferences to create Shapley values that help explain model behavior. Shapley values are a way of attributing the contribution of each feature to the model output. They can help the credit team understand why the model makes certain decisions and how the features affect the model outcomes. A chart that shows features and SHapley Additive exPlanations (SHAP) values can be created using the SHAP library in Python. This option is the most operationally efficient because it leverages the existing XGBoost training container and the built-in capabilities of Amazon SageMaker Model Monitor and SHAP library. References:

Amazon SageMaker Studio

Amazon SageMaker Model Monitor

SHAP library

**QUESTION NO: 29**

회사는 사용 가능한 과거 판매 데이터를 기반으로 주택의 판매 가격을 예측하려고 합니다. 회사 데이터 세트의 대상 변수는 판매 가격입니다. 기능에는 부지 크기, 거실 면적 측정, 비거실 면적 측정, 침실 수, 욕실 수, 건축 연도 및 우편 번호와 같은 매개변수가 포함됩니다. 이 회사는 다변수 선형 회귀를 사용하여 주택 판매 가격을 예측하려고 합니다.

기계 학습 전문가가 분석과 관련이 없는 기능을 제거하고 모델의 복잡성을 줄이기 위해 어떤 단계를 수행해야 합니까?

- A.** 기능의 히스토그램을 플롯하고 표준 편차를 계산합니다. 분산이 높은 기능을 제거합니다.
- B.** 기능의 히스토그램을 플롯하고 표준 편차를 계산합니다. 분산이 낮은 기능을 제거합니다.

C. 자체에 대한 데이터 세트의 상관 관계를 보여주는 히트맵을 작성합니다. 상호 상관 점수가 낮은 기능을 제거합니다.

D. 대상 변수에 대한 모든 기능의 상관 관계 검사를 실행합니다. 목표 변수 상관 점수가 낮은 기능을 제거합니다.

**Answer: D**

Explanation:

Feature selection is the process of reducing the number of input variables to those that are most relevant for predicting the target variable. One way to do this is to run a correlation check of all features against the target variable and remove features with low target variable correlation scores. This means that these features have little or no linear relationship with the target variable and are not useful for the prediction. This can reduce the model's complexity and improve its performance. References:

Feature engineering - Machine Learning Lens

Feature Selection For Machine Learning in Python

### QUESTION NO: 30

한 회사는 Amazon SageMaker를 사용하여 고객 통화 기록을 기반으로 고객 이탈을 예측하는 기계 학습(ML) 모델을 구축하고 있습니다. 고객 통화의 오디오 파일은 페타바이트 단위의 통화가 녹음된 온프레미스 VoIP 시스템에 있습니다. 온프레미스 인프라는 고속 네트워킹을 갖추고 있으며 100Mbps 연결을 통한 VPN 연결을 통해 회사의 AWS 인프라에 연결됩니다. 이 회사는 추론을 위해 GPU가 필요한 고객 통화를 기록하는 알고리즘을 보유하고 있습니다. 회사는 모델 개발을 위해 AWS 클라우드의 Amazon S3 버킷에 이러한 트랜스크립션을 저장하려고 합니다.

ML 전문가는 가능한 한 빨리 S3 버킷에 기록을 전달하기 위해 어떤 솔루션을 사용해야 합니까?

A. NVIDIA Tesla 모듈과 함께 AWS Snowball Edge Compute Optimized 디바이스를 주문하고 사용하여 전사 알고리즘을 실행합니다. AWS DataSync를 사용하여 결과 트랜스크립션을 트랜스크립션 S3 버킷으로 보냅니다.

B. Amazon EC2 Inf1 인스턴스와 함께 AWS Snowcone 디바이스를 주문하고 사용하여 전사 알고리즘을 실행합니다. AWS DataSync를 사용하여 결과 전사를 전사 S3 버킷으로 보냅니다.

C. AWS Outposts를 주문하고 사용하여 GPU 기반 Amazon EC2 인스턴스에서 전사 알고리즘을 실행합니다.

결과 전사를 전사 S3 버킷에 저장합니다.

D. AWS DataSync를 사용하여 오디오 파일을 Amazon S3에 수집합니다. Amazon S3에 업로드될 때 오디오 파일에 대해 전사 알고리즘을 실행하는 AWS Lambda 함수를 생성합니다. 결과 트랜스크립션을 트랜스크립션 S3 버킷에 쓰도록 함수를 구성합니다.

**Answer: A**

Explanation:

The company needs to transcribe petabytes of audio files from an on-premises VoIP system to an S3 bucket in the AWS Cloud. The transcription algorithm requires GPUs for inference, which are not available on the on-premises system. The VPN connection over a 100 Mbps connection is not sufficient to transfer the large amount of data quickly. Therefore, the company should use an AWS Snowball Edge Compute Optimized device with an NVIDIA Tesla module to run the transcription algorithm locally and leverage the GPU power. The device can store up to 42 TB of data and can be shipped back to AWS for data

ingestion. The company can use AWS DataSync to send the resulting transcriptions to the transcription S3 bucket in the AWS Cloud.

This solution minimizes the network bandwidth and latency issues and enables faster data processing and transfer.

Option B is incorrect because AWS Snowcone is a small, portable, rugged, and secure edge computing and data transfer device that can store up to 8 TB of data. It is not suitable for processing petabytes of data and does not support GPU-based instances.

Option C is incorrect because AWS Outposts is a service that extends AWS infrastructure, services, APIs, and tools to virtually any data center, co-location space, or on-premises facility. It is not designed for data transfer and ingestion, and it would require additional infrastructure and maintenance costs.

Option D is incorrect because AWS DataSync is a service that makes it easy to move large amounts of data to and from AWS over the internet or AWS Direct Connect. However, using DataSync to ingest the audio files to S3 would still be limited by the network bandwidth and latency. Moreover, running the transcription algorithm on AWS Lambda would incur additional costs and complexity, and it would not leverage the GPU power that the algorithm requires.

AWS Snowball Edge Compute Optimized

AWS DataSync

AWS Snowcone

AWS Outposts

AWS Lambda

### QUESTION NO: 31

머신 러닝(ML) 전문가는 마케팅 데이터 세트에 대한 이진 분류 문제를 해결해야 합니다.

ML 전문가는 XGBoost 알고리즘을 학습하여 알고리즘의 ROC 곡선 아래 면적(AUC)을 최대화해야 합니다. ML 전문가는 가장 정확한 모델을 생성하는 eta, alpha, min\_child\_weight, max\_depth 하이퍼파라미터 값을 찾아야 합니다.

어떤 접근 방식이 운영 비용을 최소화하면서 이러한 요구 사항을 충족할 수 있을까요?

**A.** 부트스트랩 스크립트를 사용하여 Amazon EMR 클러스터에 scikit-learn을 설치합니다.

EMR 클러스터를 배포합니다.

알고리즘에 k-겹 교차 검증 방법을 적용합니다.

**B.** scikit-learn이 설치된 Amazon SageMaker의 사전 빌드된 Docker 이미지를 배포합니다.

알고리즘에 k-겹 교차 검증 방법을 적용합니다.

**C.** Amazon SageMaker 자동 모델 튜닝(AMT)을 사용합니다. 각 하이퍼파라미터에 대해 값 범위를 지정합니다.

**D.** AWS Marketplace에 있는 AUC 알고리즘을 구독합니다. 각 하이퍼파라미터에 대한 값 범위를 지정합니다.

**Answer: C**

Explanation:

SageMaker Automatic Model Tuning (AMT) is a fully managed hyperparameter optimization feature that finds the best values for model parameters like eta, alpha, min\_child\_weight, and max\_depth.

"Use SageMaker automatic model tuning to find the best version of a model by running many training jobs on your dataset using the ranges of hyperparameters that you specify." It

supports built-in algorithms like XGBoost, and can optimize for evaluation metrics like AUC, making it the least operational overhead solution for this task.

**QUESTION NO: 32**

한 자동차 회사는 이미지에 자동차가 있는지 여부를 감지하는 기계 학습 솔루션을 개발하고 있습니다. 이미지 데이터세트는 백만 개의 이미지로 구성됩니다. 데이터세트의 각 이미지는 높이 200픽셀, 너비 200픽셀입니다. 각 이미지는 자동차가 있거나 없는 것으로 표시되어 있습니다.

이미지에 자동차가 있는지 여부를 가장 높은 정확도로 감지하는 모델을 생성할 가능성이 가장 높은 아키텍처는 무엇입니까?

- A. 이미지를 입력으로 사용하는 심층 합성곱 신경망(CNN) 분류기를 사용합니다. 이미지에 자동차가 포함될 확률을 출력하는 선형 출력 레이어를 포함합니다.
- B. 이미지를 입력으로 사용하는 심층 합성곱 신경망(CNN) 분류기를 사용합니다. 이미지에 자동차가 포함될 확률을 출력하는 소프트맥스 출력 레이어를 포함합니다.
- C. 이미지를 입력으로 사용하는 심층 다층 퍼셉트론(MLP) 분류기를 사용합니다. 이미지에 자동차가 포함될 확률을 출력하는 선형 출력 레이어를 포함합니다.
- D. 이미지를 입력으로 사용하는 심층 다층 퍼셉트론(MLP) 분류기를 사용합니다. 이미지에 자동차가 포함될 확률을 출력하는 소프트맥스 출력 레이어를 포함합니다.

**Answer: A**

Explanation:

A deep convolutional neural network (CNN) classifier is a suitable architecture for image classification tasks, as it can learn features from the images and reduce the dimensionality of the input. A linear output layer that outputs the probability that an image contains a car is appropriate for a binary classification problem, as it can produce a single scalar value between 0 and 1. A softmax output layer is more suitable for a multi-class classification problem, as it can produce a vector of probabilities that sum up to 1. A deep multilayer perceptron (MLP) classifier is not as effective as a CNN for image classification, as it does not exploit the spatial structure of the images and requires a large number of parameters to process the high-dimensional input. References:

AWS Certified Machine Learning - Specialty Exam Guide

AWS Training - Machine Learning on AWS

AWS Whitepaper - An Overview of Machine Learning on AWS

**QUESTION NO: 33**

금융 회사가 신용 카드 사기를 감지하려고 합니다. 회사는 평균적으로 신용 카드 거래의 2%가 사기임을 관찰했습니다. 데이터 과학자는 1년 동안의 신용 카드 거래 데이터에 대해 분류기를 훈련했습니다. 모델은 사기 거래(양수)와 일반 거래(음수)를 식별해야 합니다. 회사의 목표는 가능한 한 많은 긍정적인 점을 정확하게 포착하는 것입니다.

데이터 과학자는 모델을 최적화하기 위해 어떤 메트릭을 사용해야 합니까? (2개를 선택하세요.)

- A. 특이성
- B. 가양성 비율
- C. 정확도
- D. 정밀 재현율 곡선 아래 영역
- E. 참양성률

**Answer:** D E

Explanation:

The data scientist should use the area under the precision-recall curve and the true positive rate to optimize the model. These metrics are suitable for imbalanced classification problems, such as credit card fraud detection, where the positive class (fraudulent transactions) is much rarer than the negative class (non-fraudulent transactions).

The area under the precision-recall curve (AUPRC) is a measure of how well the model can identify the positive class among all the predicted positives. Precision is the fraction of predicted positives that are actually positive, and recall is the fraction of actual positives that are correctly predicted. A higher AUPRC means that the model can achieve a higher precision with a higher recall, which is desirable for fraud detection.

The true positive rate (TPR) is another name for recall. It is also known as sensitivity or hit rate. It measures the proportion of actual positives that are correctly identified by the model. A higher TPR means that the model can capture more positives, which is the company's goal.

Metrics for Imbalanced Classification in Python - Machine Learning Mastery Precision-Recall - scikit-learn