Free Questions for MLS-C01 by dumpshq

Shared by Vaughn on 15-04-2024

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

A wildlife research company has a set of images of lions and cheetahs. The company created a dataset of the images. The company labeled each image with a binary label that indicates whether an image contains a lion or cheetah. The company wants to train a model to identify whether new images contain a lion or cheetah.

.... Dh Amazon SageMaker algorithm will meet this requirement?

Options:

A- XGBoost

B- Image Classification - TensorFlow

C- Object Detection - TensorFlow

D- Semantic segmentation - MXNet

Answer:

B

Explanation:

The best Amazon SageMaker algorithm for this task is Image Classification - TensorFlow. This algorithm is a supervised learning algorithm that supports transfer learning with many pretrained models from the TensorFlow Hub. Transfer learning allows the company to fine-tune one of the available pretrained models on their own dataset, even if a large amount of image data is not available. The image classification algorithm takes an image as input and outputs a probability for each provided class label. The company can choose from a variety of models, such as MobileNet, ResNet, or Inception, depending on their accuracy and speed requirements. The algorithm also supports distributed training, data augmentation, and hyperparameter tuning.

References:

Image Classification - TensorFlow - Amazon SageMaker

Amazon SageMaker Provides New Built-in TensorFlow Image Classification Algorithm

Image Classification with ResNet :: Amazon SageMaker Workshop

Image classification on Amazon SageMaker | by Julien Simon - Medium

Question 2

Question Type: MultipleChoice

A company wants to create an artificial intelligence (AI) yoga instructor that can lead large classes of students. The company needs to create a feature that can accurately count the number of students who are in a class. The company also needs a feature that can differentiate students who are performing a yoga stretch correctly from students who are performing a stretch incorrectly.

...etermine whether students are performing a stretch correctly, the solution needs to measure the location and angle of each student's arms and legs A data scientist must use Amazon SageMaker to ...ss video footage of a yoga class by extracting image frames and applying computer vision models.

Which combination of models will meet these requirements with the LEAST effort? (Select TWO.)

Options:

A- Image Classification

B- Optical Character Recognition (OCR)

C- Object Detection

D- Pose estimation

E- Image Generative Adversarial Networks (GANs)

Answer:

C, D

Explanation:

To count the number of students who are in a class, the solution needs to detect and locate each student in the video frame. Object detection is a computer vision model that can identify and locate multiple objects in an image. To differentiate students who are

performing a stretch correctly from students who are performing a stretch incorrectly, the solution needs to measure the location and angle of each student's arms and legs. Pose estimation is a computer vision model that can estimate the pose of a person by detecting the position and orientation of key body parts. Image classification, OCR, and image GANs are not relevant for this use case.References:

Object Detection: A computer vision technique that identifies and locates objects within an image or video.

Pose Estimation: A computer vision technique that estimates the pose of a person by detecting the position and orientation of key body parts.

Amazon SageMaker: A fully managed service that provides every developer and data scientist with the ability to build, train, and deploy machine learning (ML) models quickly.

Question 3

Question Type: **MultipleChoice**

A company operates large cranes at a busy port. The company plans to use machine learning (ML) for predictive maintenance of the cranes to avoid unexpected breakdowns and to improve productivity.

The company already uses sensor data from each crane to monitor the health of the cranes in real time. The sensor data includes rotation speed, tension, energy consumption, vibration, pressure, and ...perature for each crane. The company contracts AWS ML experts to implement an ML solution.

Which potential findings would indicate that an ML-based solution is suitable for this scenario? (Select TWO.)

Options:

A- The historical sensor data does not include a significant number of data points and attributes for certain time periods.

B- The historical sensor data shows that simple rule-based thresholds can predict crane failures.

C- The historical sensor data contains failure data for only one type of crane model that is in operation and lacks failure data of most other types of crane that are in operation.

D- The historical sensor data from the cranes are available with high granularity for the last 3 years.

E- The historical sensor data contains most common types of crane failures that the company wants to predict.

Answer:

D, E

Explanation:

The best indicators that an ML-based solution is suitable for this scenario are D and E, because they imply that the historical sensor data is sufficient and relevant for building a predictive maintenance model.This model can use machine learning techniques such as regression, classification, or anomaly detection to learn from the past data and forecast future failures or issues12.Having high granularity and diversity of data can improve the accuracy and generalization of the model, as well as enable the detection of complex patterns and relationships that are not captured by simple rule-based thresholds3.

The other options are not good indicators that an ML-based solution is suitable, because they suggest that the historical sensor data is incomplete, inconsistent, or inadequate for building a predictive maintenance model.These options would require additional data collection, preprocessing, or augmentation to overcome the data quality issues and ensure that the model can handle different scenarios and types of cranes4.

References:

1:Machine Learning Techniques for Predictive Maintenance

2:A Guide to Predictive Maintenance & Machine Learning

3:Machine Learning for Predictive Maintenance: Reinventing Asset Upkeep

4:Predictive Maintenance with Machine Learning: A Complete Guide

: [Machine Learning for Predictive Maintenance - AWS Online Tech Talks]

Question 4

Question Type: **MultipleChoice**

A social media company wants to develop a machine learning (ML) model to detect Inappropriate or offensive content in images. The company has collected a large dataset of labeled images and plans to use the built-in Amazon SageMaker image classification algorithm to train the model. The company also intends to use SageMaker pipe mode to speed up the training.

...company splits the dataset into training, validation, and testing datasets. The company stores the training and validation images in folders that are named Training and Validation, respectively. The folder ...ain subfolders that correspond to the names of the dataset classes. The company resizes the images to the same sue and generates two input manifest files named training.1st and validation.1st, for the ..ing dataset and the validation dataset. respectively. Finally, the company creates two separate Amazon S3 buckets for uploads of the training dataset and the validation dataset.

...h additional data preparation steps should the company take before uploading the files to Amazon S3?

Options:

A- Generate two Apache Parquet files, training.parquet and validation.parquet. by reading the images into a Pandas data frame and storing the data frame as a Parquet file. Upload the Parquet files to the training S3 bucket

B- Compress the training and validation directories by using the Snappy compression library Upload the manifest and compressed files to the training S3 bucket

C- Compress the training and validation directories by using the gzip compression library. Upload the manifest and compressed files to the training S3 bucket.

D- Generate two RecordIO files, training rec and validation.rec. from the manifest files by using the im2rec Apache MXNet utility tool. Upload the RecordlO files to the training S3 bucket.

Answer:

D

Explanation:

The SageMaker image classification algorithm supports both RecordIO and image content types for training in file mode, and supports the RecordIO content type for training in pipe mode1.However, the algorithm also supports training in pipe mode using the image files without creating RecordIO files, by using the augmented manifest format2. In this case, the company should generate

Question 5

Question Type: MultipleChoice

A data engineer is preparing a dataset that a retail company will use to predict the number of visitors to stores. The data engineer created an Amazon S3 bucket. The engineer subscribed the S3 bucket to an AWS Data Exchange data product for general economic indicators. The data engineer wants to join the economic indicator data to an existing table in Amazon Athena to merge with the business dat

a. All these transformations must finish running in 30-60 minutes.

Which solution will meet these requirements MOST cost-effectively?

Options:

A- Configure the AWS Data Exchange product as a producer for an Amazon Kinesis data stream. Use an Amazon Kinesis Data Firehose delivery stream to transfer the data to Amazon S3 Run an AWS Glue job that will merge the existing business data with the Athena table. Write the result set back to Amazon S3.

B- Use an S3 event on the AWS Data Exchange S3 bucket to invoke an AWS Lambda function. Program the Lambda function to use Amazon SageMaker Data Wrangler to merge the existing business data with the Athena table. Write the result set back to Amazon S3.

C- Use an S3 event on the AWS Data Exchange S3 bucket to invoke an AWS Lambda Function Program the Lambda function to run an AWS Glue job that will merge the existing business data with the Athena table Write the results back to Amazon S3.

D- Provision an Amazon Redshift cluster. Subscribe to the AWS Data Exchange product and use the product to create an Amazon Redshift Table Merge the data in Amazon Redshift. Write the results back to Amazon S3.

Answer:

B

Explanation:

The most cost-effective solution is to use an S3 event to trigger a Lambda function that uses SageMaker Data Wrangler to merge the data. This solution avoids the need to provision and manage any additional resources, such as Kinesis streams, Firehose delivery streams, Glue jobs, or Redshift clusters. SageMaker Data Wrangler provides a visual interface to import, prepare, transform, and analyze data from various sources, including AWS Data Exchange products. It can also export the data preparation workflow to a Python script that can be executed by a Lambda function. This solution can meet the time requirement of 30-60 minutes, depending on the size and complexity of the data.

References:

Using Amazon S3 Event Notifications

Prepare ML Data with Amazon SageMaker Data Wrangler

AWS Lambda Function

Question 6

Question Type: **MultipleChoice**

A retail company wants to build a recommendation system for the company's website. The system needs to provide recommendations for existing users and needs to base those recommendations on each user's past browsing history. The system also must filter out any items that the user previously purchased.

Which solution will meet these requirements with the LEAST development effort?

Options:

A- Train a model by using a user-based collaborative filtering algorithm on Amazon SageMaker. Host the model on a SageMaker real-time endpoint. Configure an Amazon API Gateway API and an AWS Lambda function to handle real-time inference requests that the web application sends. Exclude the items that the user previously purchased from the results before sending the results back to the web application.

B- Use an Amazon Personalize PERSONALIZED_RANKING recipe to train a model. Create a real-time filter to exclude items that the user previously purchased. Create and deploy a campaign on Amazon Personalize. Use the GetPersonalizedRanking API operation to get the real-time recommendations.

C- Use an Amazon Personalize USER_ PERSONAL IZATION recipe to train a model Create a real-time filter to exclude items that the user previously purchased. Create and deploy a campaign on Amazon Personalize. Use the GetRecommendations API operation to get the real-time recommendations.

D- Train a neural collaborative filtering model on Amazon SageMaker by using GPU instances. Host the model on a SageMaker real-time endpoint. Configure an Amazon API Gateway API and an AWS Lambda function to handle real-time inference requests that the web application sends. Exclude the items that the user previously purchased from the results before sending the results back to the web application.

Answer:

C

Explanation:

Amazon Personalize is a fully managed machine learning service that makes it easy for developers to create personalized user experiences at scale. It uses the same recommender system technology that Amazon uses to create its own personalized recommendations. Amazon Personalize provides several pre-built recipes that can be used to train models for different use cases. The USER_PERSONALIZATION recipe is designed to provide personalized recommendations for existing users based on their past interactions with items. The PERSONALIZED_RANKING recipe is designed to re-rank a list of items for a user based on their preferences. The USER_PERSONALIZATION recipe is more suitable for this use case because it can generate recommendations for

each user without requiring a list of candidate items. To filter out the items that the user previously purchased, a real-time filter can be created and applied to the campaign. A real-time filter is a dynamic filter that uses the latest interaction data to exclude items from the recommendations. By using Amazon Personalize, the development effort is minimized because it handles the data processing, model training, and deployment automatically. The web application can use the GetRecommendations API operation to get the real-time recommendations from the campaign.References:

Amazon Personalize

What is Amazon Personalize?

USER_PERSONALIZATION recipe

PERSONALIZED_RANKING recipe

Filtering recommendations

GetRecommendations API operation

Question 7

Question Type: **MultipleChoice**

A company processes millions of orders every day. The company uses Amazon DynamoDB tables to store order information. When customers submit new orders, the new orders are immediately added to the DynamoDB tables. New orders arrive in the DynamoDB

tables continuously.

A data scientist must build a peak-time prediction solution. The data scientist must also create an Amazon OuickSight dashboard to display near real-lime order insights. The data scientist needs to build a solution that will give QuickSight access to the data as soon as new order information arrives.

Which solution will meet these requirements with the LEAST delay between when a new order is processed and when QuickSight can access the new order information?

Options:

A- Use AWS Glue to export the data from Amazon DynamoDB to Amazon S3. Configure OuickSight to access the data in Amazon S3.

B- Use Amazon Kinesis Data Streams to export the data from Amazon DynamoDB to Amazon S3. Configure OuickSight to access the data in Amazon S3.

C- Use an API call from OuickSight to access the data that is in Amazon DynamoDB directly

D- Use Amazon Kinesis Data Firehose to export the data from Amazon DynamoDB to Amazon S3. Configure OuickSight to access the data in Amazon S3.

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 scale1.

It enables QuickSight to access the data in S3 using the Athena connector, which supports federated queries to multiple data sources, including Kinesis Data Streams2.

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 configuration3.

Using an API call from QuickSight to access the data in DynamoDB directly is not possible, as QuickSight does not support direct querying of DynamoDB4.

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 seconds5.

References:

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 8

Question Type: **MultipleChoice**

A credit card company wants to identify fraudulent transactions in real time. A data scientist builds a machine learning model for this purpose. The transactional data is captured and stored in Amazon S3. The historic data is already labeled with two classes: fraud (positive) and fair transactions (negative). The data scientist removes all the missing data and builds a classifier by using the XGBoost algorithm in Amazon SageMaker. The model produces the following results:

* True positive rate (TPR): 0.700

* False negative rate (FNR): 0.300

* True negative rate (TNR): 0.977

* False positive rate (FPR): 0.023

* Overall accuracy: 0.949

Which solution should the data scientist use to improve the performance of the model?

Options:

A- Apply the Synthetic Minority Oversampling Technique (SMOTE) on the minority class in the training dataset. Retrain the model with the updated training data.

B- Apply the Synthetic Minority Oversampling Technique (SMOTE) on the majority class in the training dataset. Retrain the model with the updated training data.

C- Undersample the minority class.

D- Oversample the majority class.

Answer:

A

Explanation:

The solution that the data scientist should use to improve the performance of the model is to apply the Synthetic Minority Oversampling Technique (SMOTE) on the minority class in the training dataset, and retrain the model with the updated training data. This solution can address the problem of class imbalance in the dataset, which can affect the model's ability to learn from the rare but important positive class (fraud).

Class imbalance is a common issue in machine learning, especially for classification tasks. It occurs when one class (usually the positive or target class) is significantly underrepresented in the dataset compared to the other class (usually the negative or non-target class). For example, in the credit card fraud detection problem, the positive class (fraud) is much less frequent than the negative class (fair transactions). This can cause the model to be biased towards the majority class, and fail to capture the characteristics and patterns of the minority class. As a result, the model may have a high overall accuracy, but a low recall or true positive rate for the minority class, which means it misses many fraudulent transactions.

SMOTE is a technique that can help mitigate the class imbalance problem by generating synthetic samples for the minority class. SMOTE works by finding the k-nearest neighbors of each minority class instance, and randomly creating new instances along the line segments connecting them. This way, SMOTE can increase the number and diversity of the minority class instances, without duplicating or losing any information.By applying SMOTE on the minority class in the training dataset, the data scientist can balance the classes and improve the model's performance on the positive class1.

The other options are either ineffective or counterproductive. Applying SMOTE on the majority class would not balance the classes, but increase the imbalance and the size of the dataset. Undersampling the minority class would reduce the number of instances available for the model to learn from, and potentially lose some important information. Oversampling the majority class would also increase the imbalance and the size of the dataset, and introduce redundancy and overfitting.

References:

1: SMOTE for Imbalanced Classification with Python - Machine Learning Mastery