Machine Learning Implementation and Operations

Completed on 28-January-2021

Attempt	Marks Obtained	Your score	Time Taken	Result
01	0 / 7	0.0%	00 H 00 M 09 S	Failed

Domains wise Quiz Performance Report

No	Domain	Total Question	Correct	Incorrect	Unattempted	Marked as Review
1	ML Implementation and Operations	7	0	1	6	0
Total	All Domain	7	0	1	6	0

Review the Answers

Sorting by All

Question 1 **Unattempted**

Domain :ML Implementation and Operations

You are a machine learning specialist at a large bank where you are rolling out a new marketing campaign. The campaign is to determine whether a given customer will enroll in a new term deposit at the bank. You have decided to use the SageMaker Autopilot service to produce the best machine learning pipeline for your enrollment predictions. Your data set, which consists of 17 attributes and hundreds of millions of rows, is stored in the libsvm format on one of your S3 buckets. You plan to use SageMaker Autopilot on your dataset to get the most accurate machine learning pipeline by exploring several potential options, or candidate models. However, when you attempt to run AutoPilot using your Autopilot job, you get an error stating:
Could not complete the data builder processing job. The AutoML Job cannot continue.

What might be the cause of your Autopilot job failure?

A. Your data source is too large, 17 attributes x hundreds of millions of rows, you need to break your data source into subset files and run your Autopilot job over the subsets of data.

B. The format of your data source is not supported by the Autopilot service. ✓

C. Your data source is in a tabular format, which needs to be transformed using a machine learning tool such as sklearn or pandas.

D. Your data source has a header row in the first file in your set of source data files. You need to remove the header row.

Explanation:

Answer: B

Option A is incorrect. The AutoPilot service, as it is built upon the SageMaker service, has no limit to the size of your source data.

Option B is CORRECT. SageMaker AutoPilot only supports tabular data sets in the CSV format. Your source data is in the libsvm format. This will cause your Autopilot job to fail with the given error.

Option C is incorrect. SageMaker AutoPilot supports tabular data sets in the CSV format. If your data source is in a tabular format that cannot be the source of your error. Your source data is in the libsvm format, which will cause the given error.

Option D is incorrect. The AutoPilot service supports tabular data where either all files have a header row, or the first file of the dataset, when sorted in alphabetical/lexical order, has a header row. Therefore, your data source having a header row cannot be the source of your error.

Reference:

Please see the **Amazon Sagemaker developer guide** titled Samples: Explore modeling with Amazon SageMaker Autopilot.

Please refer to the GitHub repository titled **Direct Marketing with Amazon SageMaker Autopilot.**

Ask our Experts Rate this Question? ☺ ☹

View Queries open

Question 2 **Unattempted**

Domain :ML Implementation and Operations

As a machine learning specialist for a credit card company, you are building a model to extract information from scanned bank application documents. These forms generate datasets containing customer Personally Identifiable Information (PII) such as credit card numbers. The resulting datasets need to be highly accurate; your company has a very low tolerance for inaccurate bank application data.
How can you ensure the PII data remains encrypted and the credit card information is secure while also ensuring the highest level of quality from the scanned forms?

A. **Build a custom encryption algorithm to encrypt the data. Then store the data in your SageMaker instance in your private subnet in your VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers. Finally, use SageMaker GroundTruth to provide human review of the image scan data.**

B. **Use an IAM policy to encrypt the data on your S3 bucket. Then use Kinesis Data Analytics to obfuscate the credit card numbers. Finally, use SageMaker AutoPilot to provide human review of the image scan data.**

C. **Use a SageMaker custom image to encrypt the data when it is loaded into the SageMaker instance in your private subnet in your VPC. Use the SageMaker principal component analysis built-in algorithm to obfuscate the credit card numbers.**

D. **Use AWS KMS to encrypt the data on S3 and in your SageMaker environment. Then obfuscate the credit card numbers from the customer data using AWS Comprehend. Finally, use SageMaker Augmented AI to provide human review of the image scan data.** ✓

Explanation:

Answer: D

Option A is incorrect. Building a custom encryption algorithm will be time consuming and will most likely produce a less secure encryption technique than using AWS KMS. Also, the SageMaker DeepAR built-in algorithm is used for forecasting, not randomizing credit card numbers. Finally, the GroundTruth service is used to label datasets, not for verifying datasets.

Option B is incorrect. An IAM policy cannot be used to encrypt your data, it is used to build policies to control access to your data on S3 and in SageMaker. You could use Kinesis Data Analytics to obfuscate your data but it would be a very inefficient way of achieving this goal. Also, the question doesn't state that the scanned data is streamed to your S3 bucket, so a Kinesis solution is not a good choice. Finally, SageMaker AutoPilot is a service that allows you to automate your SageMaker pipeline, it does not provide a human review capability.

Option C is incorrect. A SageMaker custom image, a component of SageMaker Studio

service, will not give you the capability to encrypt your data. Also, the PCA built-in algorithm is used for dimensionality reduction, it would not be a good choice for data obfuscation.

Option D is CORRECT. AWS KMS is the best choice provided for encrypting your data on your S3 bucket and in your SageMaker environment. Also, AWS Comprehend now provides a very efficient way to obfuscate your credit card data. Lastly, SageMaker Augmented AI can be used to provide a human review step in your data engineering step of your modeling workflow.

Reference:

Please see the **AWS SageMaker developer guide** titled Using Amazon Augmented AI for Human Review.

Please refer to the Amazon SageMaker GroundTruth overview page.

Please review the **AWS Machine Learning blog** titled Detecting and redacting PII using Amazon Comprehend.

Ask our Experts Rate this Question?

View Queries open

Question 3 **Unattempted**

Domain :ML Implementation and Operations

You are a machine learning specialist at a government agency that has gathered election data that they plan to use to predict election outcomes. Using the agile methodology, a minimum viable product for the model idea has been built using a relative data sample. Now your agency is ready to create a machine learning model using SageMaker Studio. The election data you plan to use for training data is stored in RDS SQL Server.
Which option gives you the highest performing and most efficient method of loading the data into your SageMaker Studio environment?

A. **Load the election data into your SageMaker Studio jupyter notebook using a direct connection to the SQL Server database from the code in your notebook cells.**

B. **Use AWS Data Pipeline to load the election data from your SQL Server instance to one of your S3 buckets. Load the data into your SageMaker Studio jupyter notebook from the S3 bucket.**

C. Use AWS Data Pipeline to load the election data from your SQL Server instance to a set of DynamoDB tables. Then connect to your DynamoDB tables from the code in your jupyter notebook cells to load the data into your SageMaker Studio environment.

D. Use AWS DMS to load the election data from your SQL Server instance to an ElastiCache cluster. Then connect to your Elasticache cluster from the code in your jupyter notebook cells to load the data into your SageMaker Studio environment.

Explanation:

Answer: B

Option A is incorrect. SageMaker requires that your model artifacts, including your training data, be stored in S3. This option describes using a direct connection to SQL Server in RDS to load the data.

Option B is CORRECT. This option correctly describes moving the training data from SQL Server to your S3 bucket. Then loading the data into your jupyter notebook from the S3 bucket.

Option C is incorrect. SageMaker requires that your model artifacts, including your training data, be stored in S3. This option describes using AWS Data Pipeline to load the data into DynamoDB then loading the data into your jupyter notebook from DynamoDB.

Option D is incorrect. SageMaker requires that your model artifacts, including your training data, be stored in S3. This option describes using AWS DMS to load the data into Elasticache then loading the data into your jupyter notebook from Elasticache.

Reference:

Please see the **Amazon SageMaker developer guide** titled Get Started with Amazon SageMaker Notebook Instances and SDKs.

Please refer to the **GitHub repository** titled Getting Started with Amazon SageMaker Studio.

Please review the **GitHub repository** titled Amazon SageMaker Studio Walkthrough.

Ask our Experts Rate this Question?

View Queries open

Question 4 **Unattempted**

Domain :ML Implementation and Operations

You are an engineering manager for a healthcare conglomerate that is expanding its web application suite of online patient tools and services. You have just started a new web user experience lab where you are responsible for creating a SageMaker environment for your analyst and engineering staff. Your corporation's data security policies require that communication within your web lab machine learning environment not traverse the internet. How can you provide the SageMaker service to your user experience labs without allowing internet access for your lab SageMaker notebook instances?

A. **Use a NAT gateway in the public subnet of your corporate VPC to enable the SageMaker service.**

B. **Use a NAT gateway in the private subnet of your corporate VPC to enable the SageMaker service.**

C. **Use a SageMaker VPC interface endpoint in your corporate VPC to enable the SageMaker service.** ✓

D. **Use VPC peering between the VPC hosting the SageMaker service and your corporate VPC to enable the SageMaker service.**

Explanation:

Answer: C

Option A is incorrect. You cannot enable the SageMaker service without a connection between your VPC and the Amazon SageMaker service, which runs in an Amazon managed service account. Using a NAT Gateway in your public subnet within your corporate VPC will send your traffic over the public internet.

Option B is incorrect. NAT Gateways are instantiated in your public subnet, not your private subnet. Also, using a NAT Gateway in your public subnet within your corporate VPC will send your traffic over the public internet.

Option C is CORRECT. Using a VPC interface endpoint to enable communication between your VPC and the SageMaker managed service account allows your traffic to flow entirely and securely within the AWS network.

Option D is incorrect. The VPC hosting the SageMaker service is an Amazon managed service account. VPC Peering is used for client account peering, not for communication with an Amazon managed service account.

Reference:

Please see the **AWS Machine Learning blog** titled Understanding Amazon SageMaker notebook instance networking configurations and advanced routing options.

Please refer to the **Amazon SageMaker developer guide** titled Connect to SageMaker

Through a VPC Interface Endpoint

Ask our Experts Rate this Question?

View Queries open

Question 5 **Unattempted**

Domain :ML Implementation and Operations

You are on a machine learning team working for a retail website. Your team is responsible for the shopping cart leg of the "complete purchase" journey. Your team is building an up-sell recommendation model to be used to show "customers who purchased this item also purchased these other items" recommendations on the shopping cart page. Your model will use a dataset containing customer credit card information and other Personally Identifiable Information (PII). The PII data needs to be protected in order to meet the Payment Card Industry Data Security Standard (PCI DSS) requirements.
What is the most efficient way to obfuscate the PII data, including the credit card information?

A. Use KMS to encrypt the PII data in transit and at rest

B. Tokenize the PII data ✓

C. Use AWS Shield Advanced for all website traffic

D. Use GuardDuty for website traffic

Explanation:

Answer: B

Option A is incorrect. Using KMS and encrypting your data in transit and at rest is more complex and costly than using tokenization on the specific PII data, including the credit card data.

Option B is CORRECT. You can use tokenization instead of encryption when you only need to protect specific highly sensitive data for regulatory compliance requirements, such as PCI DSS.

Option C is incorrect. You can use AWS Shield to protect your retail website from distributed denial of service (DDoS) attacks, and Shield Advanced gives you a DDoS response team. However, Shield and Shield advanced won't protect your PII credit card data from being exposed. You need to either encrypt your data or tokenize your customer's PII data.

Option D is incorrect. You can use GuardDuty to systematically monitor network traffic to detect anomalies in the behavior of your website users by using machine learning. However, GuardDuty won't protect your PII credit card data from being exposed. You need to either encrypt your data or tokenize your customer's PII data.

Reference:

Please see the **AWS Compute Blog** titled Building a serverless tokenization solution to mask sensitive data.

Please see the **AWS Big Data blog** titled Best practices for securing sensitive data in AWS data stores.

Please see the **Wikipedia page** titled Payment Card Industry Data Security Standard.

Please refer to the **AWS GuardDuty** overview page.

Please see the **AWS Shield** overview page.

Ask our Experts Rate this Question?

View Queries open

Question 6 **Unattempted**

Domain :ML Implementation and Operations

You are a machine learning specialist at a medical research institute where you are using deep learning techniques to analyze brain scan images. Your team has built and trained your deep learning model using one of the pre-built SageMaker container images for TensorFlow and you are now attempting to deploy it for inferences in production.
You have used the TensorFlow 2.3.0 framework using Horovod on GPU servers using python version 3.7. Your training jobs run well but when you deploy to your container image for serving inferences your container fails. Why is your inference container failing?

A. **You cannot use TensorFlow in a SageMaker inference container image**

B. **The TensorFlow SageMaker inference container image only runs on CPU servers**

C. **You need to remove the Horovod operations from your inference container** ✓

D. **You should use the DeepAR SageMaker built-in algorithm**

Explanation:

Answer: C

Option A is incorrect. You can use TensorFlow in a SageMaker inference container, you just can't use the container with Horovod for inference.

Option B is incorrect. You can use TensorFlow in a SageMaker inference container on GPUs or CPUs, you just can't use the container with Horovod for inference.

Option C is CORRECT. When running inference on a SageMaker inference container for TensorFlow that was trained with Horovod you need to remove the references to Horovod before deploying for inference.

Option D is incorrect. The DeepAR SageMaker built-in algorithm is an algorithm for forecasting one-dimensional time series using recurrent neural networks (RNN). You would not use this algorithm for image analysis.

Reference:

Please see the **Amazon SageMaker developer guide** titled Using Docker containers with SageMaker.

Please see the **Amazon SageMaker developer guide** titled Prebuilt SageMaker Docker Images for TensorFlow, MXNet, Chainer, and PyTorch.

Please see the **GitHub repository** titled Available Deep Learning Containers Images.

Please refer to the **GitHub repository** overview page titled Horovod.

Please see the **GitHub repository** titled Horovod Inference.

Please see the **Amazon SageMaker developer guide** titled DeepAR Forecasting Algorithm.

Ask our Experts Rate this Question?

View Queries open

Question 7 **Incorrect**

Domain :ML Implementation and Operations

You are a machine learning specialist at a financial services firm. Your team has been tasked with developing a forecasting model to predict the price movement of the S&P 500 emini futures contract on the CME trading exchange. You have historical data from the past year as

your training data. You are also considering using trading data from the closely related NASDAQ 100 emini futures contract as part of your training data.
Which AWS machine learning service and algorithm should you use for your model that gives your team the most expeditious result with the least amount of administrative overhead?

A. SageMaker using the DeepAR Forecasting algorithm

B. Amazon Forecast using the Prophet algorithm

✓ C. Amazon Forecast using the NTPS algorithm

D. Amazon Forecast using the ARIMA algorithm

E. Amazon Forecast using the DeepAR+ algorithm

Explanation:

Answer: E

Option A is incorrect. While the SageMaker DeepAR Forecasting algorithm would be a good choice for this type of forecasting, it requires much more administrative effort than using Amazon Forecast.

Option B is incorrect. The Amazon Forecast Prophet algorithm is best suited for time series with strong seasonal effects and several seasons of historical data. Stock index futures data is not seasonal in nature.

Option C is incorrect. The Amazon Forecast NTPS algorithm is best suited for parse or intermittent time series. Stock index futures data is not sparse or intermittent in nature.

Option D is incorrect. The Amazon Forecast ARIMA algorithm is best suited for simple datasets with under 100 time series. Stock index futures data is complex with many features and millions of observations across many time series.

Option E is CORRECT. The Amazon Forecast DeepAR+ algorithm works best with large datasets containing hundreds of feature time series. It also works with forward-looking related time series. Additionally, Amazon Forecast requires much less administrative overhead than SageMaker, so this is the best option for the given scenario.

Reference:

Please see the **Amazon Forecast developer guide** titled Getting Started (Console).

Please see the **Amazon Forecast developer guide** titled Choosing an Amazon Forecast Algorithm.

Please see the **Amazon SageMaker developer guide** titled Use Amazon SageMaker Built-in Algorithms.

Ask our Experts

Rate this Question?

View Queries open

Finish Review

Certification

Cloud Certification

Java Certification

PM Certification

Big Data Certification

Company

Become Our Instructor

Support

Discussions

Blog

Business

Support

Contact Us

Help Topics

Join us on Slack!

Join our open **Slack community** and get your queries answered instantly! Our experts are online to answer your questions!

Follow us