IT-DUMPS Q&A

Accurate study guides, High passing rate!
IT-dumps provides update free of charge in one year!

Exam : **MLS-C01**

Title : AWS Certified Machine
Learning - Specialty

Version : DEMO

1.A Machine Learning Specialist is working with multiple data sources containing billions of records that need to be joined.

What feature engineering and model development approach should the Specialist take with a dataset this large?

A. Use an Amazon SageMaker notebook for both feature engineering and model development

B. Use an Amazon SageMaker notebook for feature engineering and Amazon ML for model development

C. Use Amazon EMR for feature engineering and Amazon SageMaker SDK for model development

D. Use Amazon ML for both feature engineering and model development.

Answer: C

Explanation:

Amazon EMR is a service that can process large amounts of data efficiently and cost-effectively. It can run distributed frameworks such as Apache Spark, which can perform feature engineering on big data. Amazon SageMaker SDK is a Python library that can interact with Amazon SageMaker service to train and deploy machine learning models. It can also use Amazon EMR as a data source for training data.

References:

Amazon EMR

Amazon SageMaker SDK

2.A Machine Learning Specialist has completed a proof of concept for a company using a small data sample and now the Specialist is ready to implement an end-to-end solution in AWS using Amazon SageMaker. The historical training data is stored in Amazon RDS

Which approach should the Specialist use for training a model using that data?

A. Write a direct connection to the SQL database within the notebook and pull data in

B. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.

C. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in

D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

Answer: B

Explanation:

Pushing the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and providing the S3 location within the notebook is the best approach for training a model using the data stored in Amazon RDS. This is because Amazon SageMaker can directly access data from Amazon S3 and train models on it. AWS Data Pipeline is a service that can automate the movement and transformation of data between different AWS services. It can also use Amazon RDS as a data source and Amazon S3 as a data destination. This way, the data can be transferred efficiently and securely without writing any code within the notebook.

References:

Amazon SageMaker

AWS Data Pipeline

3.Which of the following metrics should a Machine Learning Specialist generally use to compare/evaluate machine learning classification models against each other?

A. Recall

B. Misclassification rate

C. Mean absolute percentage error (MAPE)

D. Area Under the ROC Curve (AUC)

Answer: D

Explanation:

Area Under the ROC Curve (AUC) is a metric that measures the performance of a binary classifier across all possible thresholds. It is also known as the probability that a randomly chosen positive example will be ranked higher than a randomly chosen negative example by the classifier. AUC is a good metric to compare different classification models because it is independent of the class distribution and the decision threshold. It also captures both the sensitivity (true positive rate) and the specificity (true negative rate) of the model.

References:

AWS Machine Learning Specialty Exam Guide

AWS Machine Learning Specialty Sample Questions

4.A Machine Learning Specialist is using Amazon Sage Maker to host a model for a highly available customer-facing application.

The Specialist has trained a new version of the model, validated it with historical data, and now wants to deploy it to production To limit any risk of a negative customer experience, the Specialist wants to be able to monitor the model and roll it back, if needed

What is the SIMPLEST approach with the LEAST risk to deploy the model and roll it back, if needed?

A. Create a SageMaker endpoint and configuration for the new model version. Redirect production traffic to the new endpoint by updating the client configuration. Revert traffic to the last version if the model does not perform as expected.

B. Create a SageMaker endpoint and configuration for the new model version. Redirect production traffic to the new endpoint by using a load balancer Revert traffic to the last version if the model does not perform as expected.

C. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 5% of the traffic to the new variant. Revert traffic to the last version by resetting the weights if the model does not perform as expected.

D. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 100% of the traffic to the new variant Revert traffic to the last version by resetting the weights if the model does not perform as expected.

Answer: C

Explanation:

Updating the existing SageMaker endpoint to use a new configuration that is weighted to send 5% of the traffic to the new variant is the simplest approach with the least risk to deploy the model and roll it back, if needed. This is because SageMaker supports A/B testing, which allows the Specialist to compare the performance of different model variants by sending a portion of the traffic to each variant. The Specialist can monitor the metrics of each variant and adjust the weights accordingly. If the new variant does not perform as expected, the Specialist can revert traffic to the last version by resetting the weights to 100% for the old variant and 0% for the new variant. This way, the Specialist can deploy the model without affecting the customer experience and roll it back easily if needed.

References:

Amazon SageMaker

Deploying models to Amazon SageMaker hosting services

5.A manufacturing company has a large set of labeled historical sales data. The manufacturer would like to predict how many units of a particular part should be produced each quarter.

Which machine learning approach should be used to solve this problem?

A. Logistic regression

B. Random Cut Forest (RCF)

C. Principal component analysis (PCA)

D. Linear regression

Answer: D

Explanation:

Linear regression is a machine learning approach that can be used to solve this problem. Linear regression is a supervised learning technique that can model the relationship between one or more input variables (features) and an output variable (target). In this case, the input variables could be the historical sales data of the part, such as the quarter, the demand, the price, the inventory, etc. The output variable could be the number of units to be produced for the part. Linear regression can learn the coefficients (weights) of the input variables that best fit the output variable, and then use them to make predictions for new data. Linear regression is suitable for problems that involve continuous and numeric output variables, such as predicting house prices, stock prices, or sales volumes.

References:

AWS Machine Learning Specialty Exam Guide

Linear Regression