AWS-Certified-Machine-Learning-Specialty考題，最新AWS-Certified-Machine-Learning-Specialty考古題

BONUS!!! 免費下載PDFExamDumps AWS-Certified-Machine-Learning-Specialty考試題庫的完整版：https://drive.google.com/open?id=1cbm1bCCHeYihT-u5KjO85S_vlAfBnLWY

看著這麼多種IT認證考試和這麼多考試資料，你是否感到頭疼了呢？到底要怎麼辦才好呢？要選擇哪種考試哪種資料呢？如果你不知道應該怎麼選擇，那麼我來替你選擇吧。你可以選擇參加最近很有人氣的Amazon的AWS-Certified-Machine-Learning-Specialty認證考試。得到這個考試的認證資格，你可以得到很大的好處。另外，為了更有效率地準備考試，你可以選擇PDFExamDumps的AWS-Certified-Machine-Learning-Specialty考古題。這是你輕鬆通過考試的最好的方法。

AWS Certified Machine Learning - Specialty 考試涵蓋廣泛的主題，包括數據工程、探索性數據分析、特徵工程、模型選擇和評估等。考生需要展示在使用 AWS 工具和服務方面的熟練程度，例如 Amazon SageMaker、AWS Deep Learning AMIs、Amazon Rekognition 和 Amazon Comprehend。

>> AWS-Certified-Machine-Learning-Specialty考題 <<

高質量的AWS-Certified-Machine-Learning-Specialty考題和資格考試中的領先供應平臺&有效的AWS-Certified-Machine-Learning-Specialty：AWS Certified Machine Learning - Specialty

選擇PDFExamDumps可以100%幫助你通過考試。我們根據Amazon AWS-Certified-Machine-Learning-Specialty的考試科目的不斷變化，也會不斷的更新我們的培訓資料，會提供最新的考試內容。PDFExamDumps可以為你免費提供24小時線上客戶服務，如果你沒有通過Amazon AWS-Certified-Machine-Learning-Specialty的認證考試，我們會全額退款給您。

Amazon MLS -C01（AWS認證的機器學習 - 專業）認證考試專為對機器學習概念，技術和最佳實踐有深入了解的個人而設計。該考試旨在驗證個人在AWS平台上構建和部署機器學習模型方面的技術專業知識。該認證適用於使用機器學習技術的任何人，包括數據科學家，開發人員和軟件工程師。

最新的 AWS Certified Machine Learning AWS-Certified-Machine-Learning-Specialty 免費考試真題 (Q276-Q281):

問題 #276
A company is building a new supervised classification model in an AWS environment. The company's data science team notices that the dataset has a large quantity of variables Ail the variables are numeric. The model accuracy for training and validation is low. The model's processing time is affected by high latency The data science team needs to increase the accuracy of the model and decrease the processing.
How it should the data science team do to meet these requirements?

- A. Use a principal component analysis (PCA) model.

- B. Create new features and interaction variables.
- C. Apply normalization on the feature set.
- D. Use a multiple correspondence analysis (MCA) model

答案：A

解題說明：
The best way to meet the requirements is to use a principal component analysis (PCA) model, which is a technique that reduces the dimensionality of the dataset by transforming the original variables into a smaller set of new variables, called principal components, that capture most of the variance and information in the data1. This technique has the following advantages:
It can increase the accuracy of the model by removing noise, redundancy, and multicollinearity from the data, and by enhancing the interpretability and generalization of the model23.
It can decrease the processing time of the model by reducing the number of features and the computational complexity of the model, and by improving the convergence and stability of the model45.
It is suitable for numeric variables, as it relies on the covariance or correlation matrix of the data, and it can handle a large quantity of variables, as it can extract the most relevant ones16.
The other options are not effective or appropriate, because they have the following drawbacks:
A: Creating new features and interaction variables can increase the accuracy of the model by capturing more complex and nonlinear relationships in the data, but it can also increase the processing time of the model by adding more features and increasing the computational complexity of the model7. Moreover, it can introduce more noise, redundancy, and multicollinearity in the data, which can degrade the performance and interpretability of the model8.
C: Applying normalization on the feature set can increase the accuracy of the model by scaling the features to a common range and avoiding the dominance of some features over others, but it can also decrease the processing time of the model by reducing the numerical instability and improving the convergence of the model . However, normalization alone is not enough to address the high dimensionality and high latency issues of the dataset, as it does not reduce the number of features or the variance in the data.
D: Using a multiple correspondence analysis (MCA) model is not suitable for numeric variables, as it is a technique that reduces the dimensionality of the dataset by transforming the original categorical variables into a smaller set of new variables, called factors, that capture most of the inertia and information in the data.
MCA is similar to PCA, but it is designed for nominal or ordinal variables, not for continuous or interval variables.
1: Principal Component Analysis - Amazon SageMaker
2: How to Use PCA for Data Visualization and Improved Performance in Machine Learning | by Pratik Shukla | Towards Data Science
3: Principal Component Analysis (PCA) for Feature Selection and some of its Pitfalls | by Nagesh Singh Chauhan | Towards Data Science
4: How to Reduce Dimensionality with PCA and Train a Support Vector Machine in Python | by James Briggs | Towards Data Science
5: Dimensionality Reduction and Its Applications | by Aniruddha Bhandari | Towards Data Science
6: Principal Component Analysis (PCA) in Python | by Susan Li | Towards Data Science
7: Feature Engineering for Machine Learning | by Dipanjan (DJ) Sarkar | Towards Data Science
8: Feature Engineering - How to Engineer Features and How to Get Good at It | by Parul Pandey | Towards Data Science
[Feature Scaling for Machine Learning: Understanding the Difference Between Normalization vs. Standardization | by Benjamin Obi Tayo Ph.D. | Towards Data Science]
[Why, How and When to Scale your Features | by George Seif | Towards Data Science]
[Normalization vs Dimensionality Reduction | by Saurabh Annadate | Towards Data Science]
[Multiple Correspondence Analysis - Amazon SageMaker]
[Multiple Correspondence Analysis (MCA) | by Raul Eulogio | Towards Data Science]

問題 #277
A data science team is planning to build a natural language processing (NLP) application. The application's text preprocessing stage will include part-of-speech tagging and key phase extraction. The preprocessed text will be input to a custom classification algorithm that the data science team has already written and trained using Apache MXNet.
Which solution can the team build MOST quickly to meet these requirements?

- A. Use an NLP library in Amazon SageMaker for the part-of-speech tagging. Use Amazon Comprehend for the key phase extraction. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.
- B. Use Amazon Comprehend for the part-of-speech tagging and key phase extraction tasks. Use AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier.
- C. Use Amazon Comprehend for the part-of-speech tagging, key phase extraction, and classification tasks.
- D. Use Amazon Comprehend for the part-of-speech tagging and key phase extraction tasks. Use Amazon SageMaker built-in Latent Dirichlet Allocation (LDA) algorithm to build the custom classifier.

答案：B

解題說明：
Amazon Comprehend is a natural language processing (NLP) service that can perform part-of-speech tagging and key phrase extraction tasks. AWS Deep Learning Containers are Docker images that are pre-installed with popular deep learning frameworks such as Apache MXNet. Amazon SageMaker is a fully managed service that can help build, train, and deploy machine learning models. Using Amazon Comprehend for the text preprocessing tasks and AWS Deep Learning Containers with Amazon SageMaker to build the custom classifier is the solution that can be built most quickly to meet the requirements.
Amazon Comprehend
AWS Deep Learning Containers
Amazon SageMaker

問題 #278
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]
Considering the graph, what is a reasonable selection for the optimal choice of k?

- A. 0
- B. 1
- C. 2
- D. 3

答案：B

解題說明：
The elbow method is a technique that we use to determine the number of centroids (k) to use in a k-means clustering algorithm. In this method, we plot the within-cluster sum of squares (WCSS) against the number of clusters (k) and look for the point where the curve bends sharply. This point is called the elbow point and it indicates that adding more clusters does not improve the model significantly. The graph in the question shows that the elbow point is at k = 4, which means that 4 is a reasonable choice for the optimal number of clusters. References:
Elbow Method for optimal value of k in KMeans: A tutorial on how to use the elbow method with Amazon SageMaker.
K-Means Clustering: A video that explains the concept and benefits of k-means clustering.

問題 #279
An aircraft engine manufacturing company is measuring 200 performance metrics in a time-series. Engineers want to detect critical manufacturing defects in near-real time during testing. All of the data needs to be stored for offline analysis.
What approach would be the MOST effective to perform near-real time defect detection?

- A. Use Amazon S3 for ingestion, storage, and further analysis. Use the Amazon SageMaker Random Cut Forest (RCF) algorithm to determine anomalies.
- B. Use Amazon Kinesis Data Firehose for ingestion and Amazon Kinesis Data Analytics Random Cut Forest (RCF) to perform anomaly detection. Use Kinesis Data Firehose to store data in Amazon S3 for further analysis.
- C. Use Amazon S3 for ingestion, storage, and further analysis. Use an Amazon EMR cluster to carry out Apache Spark ML k-means clustering to determine anomalies.
- D. Use AWS IoT Analytics for ingestion, storage, and further analysis. Use Jupyter notebooks from within AWS IoT Analytics to carry out analysis for anomalies.

答案：B

解題說明：
Explanation
The company wants to perform near-real time defect detection on a time-series of 200 performance metrics, and store all the data for offline analysis. The best approach for this scenario is to use Amazon Kinesis Data Firehose for ingestion and Amazon Kinesis Data Analytics Random Cut Forest (RCF) to perform anomaly detection. Use Kinesis Data Firehose to store data in Amazon S3 for further analysis.
Amazon Kinesis Data Firehose is a service that can capture, transform, and deliver streaming data to destinations such as Amazon S3, Amazon Redshift, Amazon OpenSearch Service, and Splunk. Kinesis Data Firehose can handle any amount and frequency of data, and automatically scale to match the throughput. Kinesis Data Firehose can also compress, encrypt, and batch the data before delivering it to the destination, reducing the storage cost and enhancing the security.
Amazon Kinesis Data Analytics is a service that can analyze streaming data in real time using SQL or Apache Flink applications.

Kinesis Data Analytics can use built-in functions and algorithms to perform various analytics tasks, such as aggregations, joins, filters, windows, and anomaly detection. One of the built-in algorithms that Kinesis Data Analytics supports is Random Cut Forest (RCF), which is a supervised learning algorithm for forecasting scalar time series using recurrent neural networks. RCF can detect anomalies in streaming data by assigning an anomaly score to each data point, based on how distant it is from the rest of the data. RCF can handle multiple related time series, such as the performance metrics of the aircraft engine, and learn a global model that captures the common patterns and trends across the time series.
Therefore, the company can use the following architecture to build the near-real time defect detection solution:
Use Amazon Kinesis Data Firehose for ingestion: The company can use Kinesis Data Firehose to capture the streaming data from the aircraft engine testing, and deliver it to two destinations:
Amazon S3 and Amazon Kinesis Data Analytics. The company can configure the Kinesis Data Firehose delivery stream to specify the source, the buffer size and interval, the compression and encryption options, the error handling and retry logic, and the destination details.
Use Amazon Kinesis Data Analytics Random Cut Forest (RCF) to perform anomaly detection:
The company can use Kinesis Data Analytics to create a SQL application that can read the streaming data from the Kinesis Data Firehose delivery stream, and apply the RCF algorithm to detect anomalies. The company can use the RANDOM_CUT_FOREST or RANDOM_CUT_FOREST_WITH_EXPLANATION functions to compute the anomaly scores and attributions for each data point, and use the WHERE clause to filter out the normal data points. The company can also use the CURSOR function to specify the input stream, and the PUMP function to write the output stream to another destination, such as Amazon Kinesis Data Streams or AWS Lambda.
Use Kinesis Data Firehose to store data in Amazon S3 for further analysis: The company can use Kinesis Data Firehose to store the raw and processed data in Amazon S3 for offline analysis. The company can use the S3 destination of the Kinesis Data Firehose delivery stream to store the raw data, and use another Kinesis Data Firehose delivery stream to store the output of the Kinesis Data Analytics application. The company can also use AWS Glue or Amazon Athena to catalog, query, and analyze the data in Amazon S3.
References:
What Is Amazon Kinesis Data Firehose?
What Is Amazon Kinesis Data Analytics for SQL Applications?
DeepAR Forecasting Algorithm - Amazon SageMaker

問題 #280
A company's machine learning (ML) specialist is building a computer vision model to classify 10 different traffic signs. The company has stored 100 images of each class in Amazon S3, and the company has another
10.000 unlabeled images. All the images come from dash cameras and are a size of 224 pixels * 224 pixels.
After several training runs, the model is overfitting on the training data.
Which actions should the ML specialist take to address this problem? (Select TWO.)

- A. Use image preprocessing to transform the images into grayscale images.
- B. Use Amazon SageMaker Ground Truth to label the unlabeled images
- C. Use the Amazon SageMaker k-nearest neighbors (k-NN) algorithm to label the unlabeled images.
- D. Replace the activation of the last layer with a sigmoid.
- E. Use data augmentation to rotate and translate the labeled images.

答案：C,E

解題說明：
* Data augmentation is a technique to increase the size and diversity of the training data by applying random transformations such as rotation, translation, scaling, flipping, etc. This can help reduce overfitting and improve the generalization of the model. Data augmentation can be done using the Amazon SageMaker image classification algorithm, which supports various augmentation options such as horizontal_flip, vertical_flip, rotate, brightness, contrast, etc1
* The Amazon SageMaker k-nearest neighbors (k-NN) algorithm is a supervised learning algorithm that can be used to label unlabeled data based on the similarity to the labeled data. The k-NN algorithm assigns a label to an unlabeled instance by finding the k closest labeled instances in the feature space and taking a majority vote among their labels. This can help increase the size and diversity of the training data and reduce overfitting. The k-NN algorithm can be used with the Amazon SageMaker image classification algorithm by extracting features from the images using a pre-trained model and then applying the k-NN algorithm on the feature vectors2
* Using Amazon SageMaker Ground Truth to label the unlabeled images is not a good option because it is a manual and costly process that requires human annotators. Moreover, it does not address the issue of overfitting on the existing labeled data.
* Using image preprocessing to transform the images into grayscale images is not a good option because it reduces the amount of information and variation in the images, which can degrade the performance of the model. Moreover, it does not address the issue of overfitting on the existing labeled data.

* Replacing the activation of the last layer with a sigmoid is not a good option because it is not suitable for a multi-class classification problem. A sigmoid activation function outputs a value between 0 and 1, which can be interpreted as a probability of belonging to a single class. However, for a multi-class classification problem, the output should be a vector of probabilities that sum up to 1, which can be achieved by using a softmax activation function.
References:
* 1: Image classification algorithm - Amazon SageMaker
* 2: k-nearest neighbors (k-NN) algorithm - Amazon SageMaker

問題 #281
.....