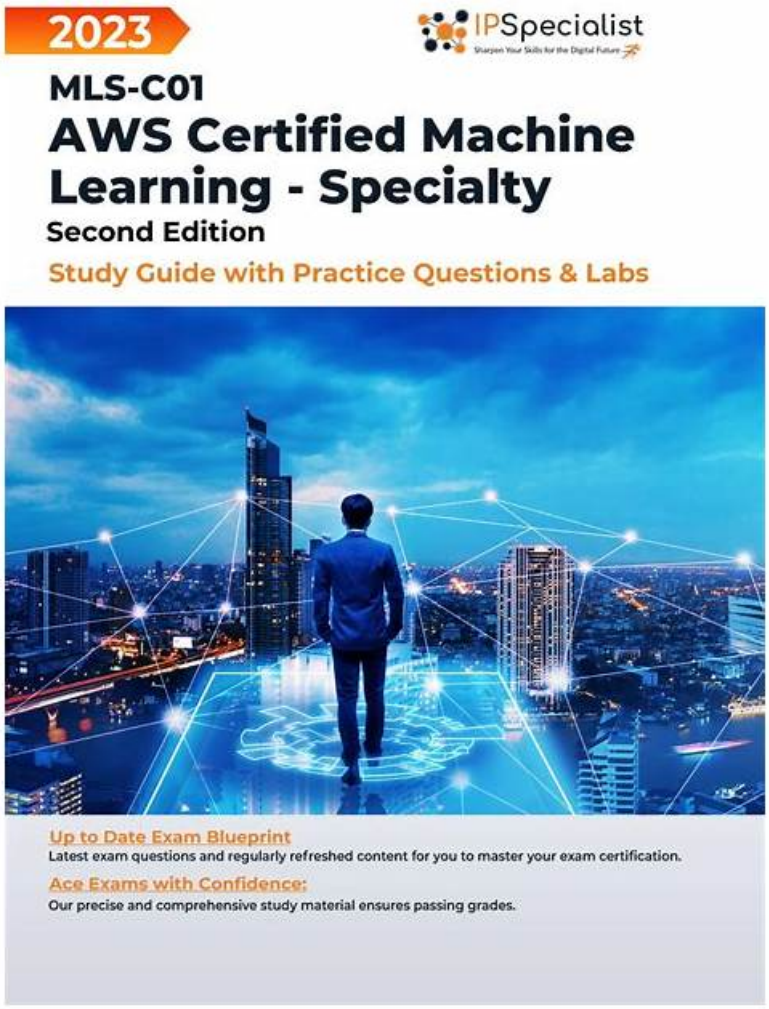
Amazon MLS-C01 Questions Tips For Better Preparation 2026



BTW, DOWNLOAD part of ITExamDownload MLS-C01 dumps from Cloud Storage: https://drive.google.com/open?id=1cj4wAA5A4F7_cVcFA0rJ18FqMIe9_bqO

Luckily, we are going to tell you a good new that the demo of the MLS-C01 study materials are easily available in our company. If you buy the study materials from our company, we are glad to offer you with the best demo of our study materials. You will have a deep understanding of the MLS-C01 Study Materials from our company, and then you will find that the study materials from our company will very useful and suitable for you to prepare for you MLS-C01 exam.

Career Path

In case you want to specialize in more specific AWS services, you can opt for other Amazon specialty certifications like the AWS Certified Advanced Networking Specialty, the AWS Certified Alexa Skill Builder Specialty, or the AWS Certified Database Specialty if to name a few.

Topics in AWS Certified Machine Learning - Specialty

The following will be discussed in **AMAZON MLS-C01 Practice Exam** and **AMAZON MLS-C01 practice exams**:

- Data Engineering
- Modeling

- Machine Learning Implementation and Operations
- Exploratory Data Analysis

>> MLS-C01 Latest Braindumps Free <<

MLS-C01 Discount Code, MLS-C01 Latest Dumps

How can you pass your exam and get your certificate in a short time? Our MLS-C01 exam torrent will be your best choice to help you achieve your aim. According to customers' needs, our product was revised by a lot of experts; the most functions of our MLS-C01 exam dumps are to help customers save more time, and make customers relaxed. If you choose to use our MLS-C01 Test Quiz, you will find it is very easy for you to pass your MLS-C01 exam in a short time. You just need to spend 20-30 hours on studying with our MLS-C01 exam questions; you will have more free time to do other things.

The AWS Certified Machine Learning - Specialty Exam is intended for individuals who have a strong understanding of machine learning, including deep learning and neural networks, and who have experience designing, implementing, and deploying machine learning solutions on the AWS platform. AWS Certified Machine Learning - Specialty certification is particularly valuable for data scientists, software developers, and other IT professionals who want to demonstrate their expertise in machine learning and differentiate themselves in a competitive job market. With this certification, candidates can showcase their skills to potential employers and clients, as well as gain access to exclusive AWS resources and networking opportunities.

Amazon AWS Certified Machine Learning - Specialty Sample Questions (Q194-Q199):

NEW QUESTION # 194
A Data Scientist needs to analyze employment data. The dataset contains approximately 10 million observations on people across 10 different features. During the preliminary analysis, the Data Scientist notices that income and age distributions are not normal. While income levels shows a right skew as expected, with fewer individuals having a higher income, the age distribution also show a right skew, with fewer older individuals participating in the workforce.
Which feature transformations can the Data Scientist apply to fix the incorrectly skewed data? (Choose two.)

- A. Cross-validation
- B. Logarithmic transformation
- C. One hot encoding
- D. High-degree polynomial transformation
- E. Numerical value binning

Answer: B,E

Explanation:
To fix the incorrectly skewed data, the Data Scientist can apply two feature transformations: numerical value binning and logarithmic transformation. Numerical value binning is a technique that groups continuous values into discrete bins or categories. This can help reduce the skewness of the data by creating more balanced frequency distributions. Logarithmic transformation is a technique that applies the natural logarithm function to each value in the data. This can help reduce the right skewness of the data by compressing the large values and expanding the small values. Both of these transformations can make the data more suitable for machine learning algorithms that assume normality of the data. References:
* Data Transformation - Amazon SageMaker
* Transforming Skewed Data for Machine Learning

NEW QUESTION # 195
A data scientist needs to identify fraudulent user accounts for a company's ecommerce platform. The company wants the ability to determine if a newly created account is associated with a previously known fraudulent user.
The data scientist is using AWS Glue to cleanse the company's application logs during ingestion.
Which strategy will allow the data scientist to identify fraudulent accounts?

- A. Search for duplicate accounts in the AWS Glue Data Catalog.
- B. Create an AWS Glue crawler to infer duplicate accounts in the source data.
- C. Create a FindMatches machine learning transform in AWS Glue.
- D. Execute the built-in FindDuplicates Amazon Athena query.

Answer: C

Explanation:
Explanation/Reference: https://docs.aws.amazon.com/glue/latest/dg/machine-learning.html

NEW QUESTION # 196
A data scientist is building a forecasting model for a retail company by using the most recent 5 years of sales records that are stored in a data warehouse. The dataset contains sales records for each of the company's stores across five commercial regions The data scientist creates a working dataset with StorelD. Region. Date, and Sales Amount as columns. The data scientist wants to analyze yearly average sales for each region. The scientist also wants to compare how each region performed compared to average sales across all commercial regions.
Which visualization will help the data scientist better understand the data trend?

- A. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each region Create a bar plot, faceted by year, of average sales for each region Add a horizontal line in each facet to represent average sales.
- B. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each store. Create a bar plot, colored by region and faceted by year, of average sales for each store.
  Add a horizontal line in each facet to represent average sales.
- C. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each region Create a bar plot of average sales for each region. Add an extra bar in each facet to represent average sales.
- D. Create an aggregated dataset by using the Pandas GroupBy function to get average sales for each year for each store. Create a bar plot, faceted by year, of average sales for each store. Add an extra bar in each facet to represent average sales.

Answer: A

Explanation:
The best visualization for this task is to create a bar plot, faceted by year, of average sales for each region and add a horizontal line in each facet to represent average sales. This way, the data scientist can easily compare the yearly average sales for each region with the overall average sales and see the trends over time. The bar plot also allows the data scientist to see the relative performance of each region within each year and across years. The other options are less effective because they either do not show the yearly trends, do not show the overall average sales, or do not group the data by region.
pandas.DataFrame.groupby - pandas 2.1.4 documentation
pandas.DataFrame.plot.bar - pandas 2.1.4 documentation
Matplotlib - Bar Plot - Online Tutorials Library

NEW QUESTION # 197
A data scientist at a financial services company used Amazon SageMaker to train and deploy a model that predicts loan defaults. The model analyzes new loan applications and predicts the risk of loan default. To train the model, the data scientist manually extracted loan data from a database. The data scientist performed the model training and deployment steps in a Jupyter notebook that is hosted on SageMaker Studio notebooks.
The model's prediction accuracy is decreasing over time. Which combination of slept in the MOST operationally efficient way for the data scientist to maintain the model's accuracy? (Select TWO.)

- A. Export the training and deployment code from the SageMaker Studio notebooks into a Python script.Package the script into an Amazon Elastic Container Service (Amazon ECS) task that an AWS Lambda function can initiate.
- B. Use SageMaker Pipelines to create an automated workflow that extracts fresh data, trains the model, and deploys a new version of the model.
- C. Store the model predictions in Amazon S3 Create a daily SageMaker Processing job that reads the predictions from Amazon S3, checks for changes in model prediction accuracy, and sends an email notification if a significant change is detected.
- D. Configure SageMaker Model Monitor with an accuracy threshold to check for model drift. Initiate an Amazon CloudWatch alarm when the threshold is exceeded. Connect the workflow in SageMaker Pipelines with the CloudWatch alarm to automatically initiate retraining.
- E. Rerun the steps in the Jupyter notebook that is hosted on SageMaker Studio notebooks to retrain the model and redeploy a new version of the model.

Answer: B,D

Explanation:
Option A is correct because SageMaker Pipelines is a service that enables you to create and manage automated workflows for your machine learning projects. You can use SageMaker Pipelines to orchestrate the steps of data extraction, model training, and model deployment in a repeatable and scalable way1.
Option B is correct because SageMaker Model Monitor is a service that monitors the quality of your models in production and alerts you when there are deviations in the model quality. You can use SageMaker Model Monitor to set an accuracy threshold for your model and configure a CloudWatch alarm that triggers when the threshold is exceeded. You can then connect the alarm to the workflow in SageMaker Pipelines to automatically initiate retraining and deployment of a new version of the model2.
Option C is incorrect because it is not the most operationally efficient way to maintain the model's accuracy.
Creating a daily SageMaker Processing job that reads the predictions from Amazon S3 and checks for changes in model prediction accuracy is a manual and time-consuming process. It also requires you to write custom code to perform the data analysis and send the email notification. Moreover, it does not automatically retrain and deploy the model when the accuracy drops.
Option D is incorrect because it is not the most operationally efficient way to maintain the model's accuracy.
Rerunning the steps in the Jupyter notebook that is hosted on SageMaker Studio notebooks to retrain the model and redeploy a new version of the model is a manual and error-prone process. It also requires you to monitor the model's performance and initiate the retraining and deployment steps yourself. Moreover, it does not leverage the benefits of SageMaker Pipelines and SageMaker Model Monitor to automate and streamline the workflow.
Option E is incorrect because it is not the most operationally efficient way to maintain the model's accuracy.
Exporting the training and deployment code from the SageMaker Studio notebooks into a Python script and packaging the script into an Amazon ECS task that an AWS Lambda function can initiate is a complex and cumbersome process. It also requires you to manage the infrastructure and resources for the Amazon ECS task and the AWS Lambda function. Moreover, it does not leverage the benefits of SageMaker Pipelines and SageMaker Model Monitor to automate and streamline the workflow.
1: SageMaker Pipelines - Amazon SageMaker
2: Monitor data and model quality - Amazon SageMaker

NEW QUESTION # 198
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. Apply normalization on the feature set.
- B. Create new features and interaction variables.
- C. Use a principal component analysis (PCA) model.
- D. Use a multiple correspondence analysis (MCA) model

Answer: C

Explanation:
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]

NEW QUESTION # 199

......

P.S. Free 2026 Amazon MLS-C01 dumps are available on Google Drive shared by ITExamDownload: https://drive.google.com/open?id=1cj4wAA5A4F7_cVcFA0rJ18FqMIe9_bqO