

# Amazon AWS-Certified-Machine-Learning-Specialty Pass Guaranteed, Test AWS-Certified-Machine-Learning-Specialty Registration



P.S. Free & New AWS-Certified-Machine-Learning-Specialty dumps are available on Google Drive shared by DumpExam  
<https://drive.google.com/open?id=1efs09djuprE7oc3H6RP7G86idoTM3BK0>

DumpExam's Amazon AWS-Certified-Machine-Learning-Specialty exam training materials are bring the greatest success rate to all the candidates who want to pass the exam. Amazon AWS-Certified-Machine-Learning-Specialty exam is a challenging Certification Exam. Besides the books, internet is considered to be a treasure house of knowledge. In DumpExam you can find your treasure house of knowledge. This is a site of great help to you. You will encounter the complex questions in the exam, but DumpExam can help you to pass the exam easily. DumpExam's Amazon AWS-Certified-Machine-Learning-Specialty Exam Training material includes all the knowledge that must be mastered for the purpose of passing the Amazon AWS-Certified-Machine-Learning-Specialty exam.

Amazon MLS-C01 (AWS Certified Machine Learning - Specialty) Certification Exam is designed for individuals who want to validate their expertise in machine learning technologies and AWS services. AWS Certified Machine Learning - Specialty certification is ideal for data scientists, software developers, and IT professionals who want to demonstrate their proficiency in designing, building, and deploying machine learning models on AWS.

>> Amazon AWS-Certified-Machine-Learning-Specialty Pass Guaranteed <<

## Pass AWS-Certified-Machine-Learning-Specialty Exam with First-grade AWS-Certified-Machine-Learning-Specialty Pass Guaranteed by DumpExam

With the help of AWS-Certified-Machine-Learning-Specialty study materials, you can conduct targeted review on the topics which to be tested before the exam, and then you no longer have to worry about the problems that you may encounter a question that you are not familiar with during the exam. With AWS-Certified-Machine-Learning-Specialty study materials, you will not need to purchase any other review materials. We have hired professional IT staff to maintain AWS-Certified-Machine-Learning-Specialty Study Materials and our team of experts also constantly updates and renew the question bank according to changes in the syllabus. With AWS-Certified-Machine-Learning-Specialty study materials, you can study at ease, and we will help you solve all the problems that you may encounter in the learning process.

Achieving the AWS Certified Machine Learning – Specialty certification can open up a range of job opportunities in the field of machine learning, including roles such as Machine Learning Engineer, Data Scientist, and AI Developer. It also provides a solid foundation for pursuing advanced certifications in the field of machine learning on AWS.

## Amazon AWS Certified Machine Learning - Specialty Sample Questions (Q198-Q203):

### NEW QUESTION # 198

A machine learning (ML) specialist wants to create a data preparation job that uses a PySpark script with complex window

aggregation operations to create data for training and testing. The ML specialist needs to evaluate the impact of the number of features and the sample count on model performance.

Which approach should the ML specialist use to determine the ideal data transformations for the model?

- A. Add an Amazon SageMaker Experiments tracker to the script to capture key parameters. Run the script as a SageMaker processing job.
- B. Add an Amazon SageMaker Debugger hook to the script to capture key parameters. Run the script as a SageMaker processing job.
- C. Add an Amazon SageMaker Experiments tracker to the script to capture key metrics. Run the script as an AWS Glue job.
- D. Add an Amazon SageMaker Debugger hook to the script to capture key metrics. Run the script as an AWS Glue job.

**Answer: C**

#### NEW QUESTION # 199

An obtain relator collects the following data on customer orders: demographics, behaviors, location, shipment progress, and delivery time. A data scientist joins all the collected datasets. The result is a single dataset that includes 980 variables.

The data scientist must develop a machine learning (ML) model to identify groups of customers who are likely to respond to a marketing campaign.

Which combination of algorithms should the data scientist use to meet this requirement? (Select TWO.)

- A. K-means
- B. Semantic segmentation
- C. Latent Dirichlet Allocation (LDA)
- D. Factorization machines (FM)
- E. Principal component analysis (PCA)

**Answer: A,E**

Explanation:

Explanation

The data scientist should use K-means and principal component analysis (PCA) to meet this requirement.

K-means is a clustering algorithm that can group customers based on their similarity in the feature space. PCA is a dimensionality reduction technique that can transform the original 980 variables into a smaller set of uncorrelated variables that capture most of the variance in the data. This can help reduce the computational cost and noise in the data, and improve the performance of the clustering algorithm.

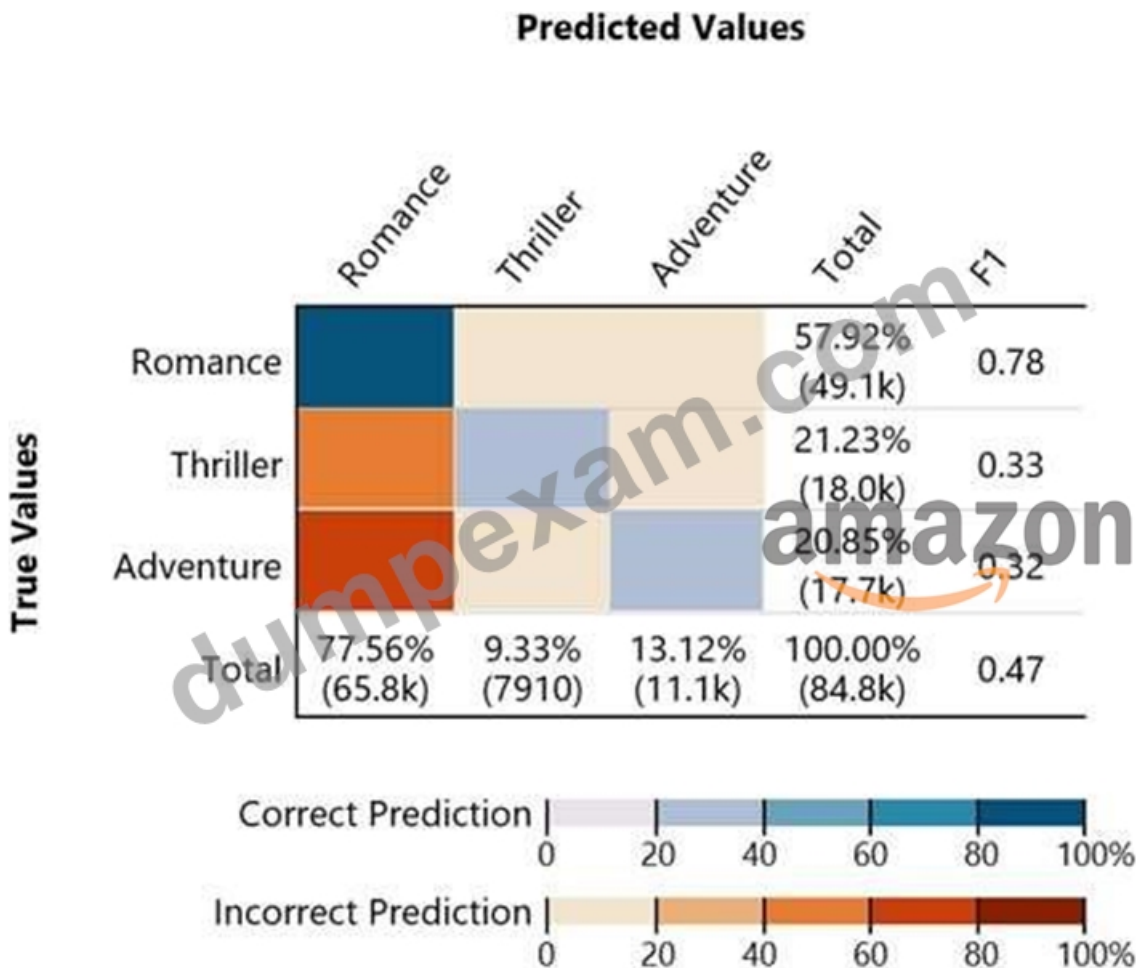
References:

Clustering - Amazon SageMaker

Dimensionality Reduction - Amazon SageMaker

#### NEW QUESTION # 200

Given the following confusion matrix for a movie classification model, what is the true class frequency for Romance and the predicted class frequency for Adventure?



- A. The true class frequency for Romance is 57.92% and the predicted class frequency for Adventure is 13.12%
- B. The true class frequency for Romance is 77.56% \* 0.78 and the predicted class frequency for Adventure is 20.85%\*0.32
- C. The true class frequency for Romance is 77.56% and the predicted class frequency for Adventure is 20.85%
- D. The true class frequency for Romance is 0.78 and the predicted class frequency for Adventure is (0.47-0.32)

**Answer: A**

#### NEW QUESTION # 201

A Machine Learning Specialist is preparing data for training on Amazon SageMaker. The Specialist is using one of the SageMaker built-in algorithms for the training. The dataset is stored in .CSV format and is transformed into a numpy.array, which appears to be negatively affecting the speed of the training.

What should the Specialist do to optimize the data for training on SageMaker?

- A. Use the SageMaker batch transform feature to transform the training data into a DataFrame.
- B. Use the SageMaker hyperparameter optimization feature to automatically optimize the data.
- C. Use AWS Glue to compress the data into the Apache Parquet format.
- D. Transform the dataset into the RecordIO protobuf format.

**Answer: D**

#### NEW QUESTION # 202

A company wants to segment a large group of customers into subgroups based on shared characteristics. The company's data scientist is planning to use the Amazon SageMaker built-in k-means clustering algorithm for this task. The data scientist needs to

determine the optimal number of subgroups (k) to use.

Which data visualization approach will MOST accurately determine the optimal value of k?

- A. Calculate the principal component analysis (PCA) components. Create a line plot of the number of components against the explained variance. The optimal value of k is the number of PCA components after which the curve starts decreasing in a linear fashion.
- **B. Run the k-means clustering algorithm for a range of k. For each value of k, calculate the sum of squared errors (SSE). Plot a line chart of the SSE for each value of k. The optimal value of k is the point after which the curve starts decreasing in a linear fashion.**
- C. Create a t-distributed stochastic neighbor embedding (t-SNE) plot for a range of perplexity values. The optimal value of k is the value of perplexity, where the clusters start to look reasonably separated.
- D. Calculate the principal component analysis (PCA) components. Run the k-means clustering algorithm for a range of k by using only the first two PCA components. For each value of k, create a scatter plot with a different color for each cluster. The optimal value of k is the value where the clusters start to look reasonably separated.

**Answer: B**

Explanation:

Explanation

The solution D is the best data visualization approach to determine the optimal value of k for the k-means clustering algorithm. The solution D involves the following steps:

Run the k-means clustering algorithm for a range of k. For each value of k, calculate the sum of squared errors (SSE). The SSE is a measure of how well the clusters fit the data. It is calculated by summing the squared distances of each data point to its closest cluster center. A lower SSE indicates a better fit, but it will always decrease as the number of clusters increases. Therefore, the goal is to find the smallest value of k that still has a low SSE1.

Plot a line chart of the SSE for each value of k. The line chart will show how the SSE changes as the value of k increases. Typically, the line chart will have a shape of an elbow, where the SSE drops rapidly at first and then levels off. The optimal value of k is the point after which the curve starts decreasing in a linear fashion. This point is also known as the elbow point, and it represents the balance between the number of clusters and the SSE1.

The other options are not suitable because:

Option A: Calculating the principal component analysis (PCA) components, running the k-means clustering algorithm for a range of k by using only the first two PCA components, and creating a scatter plot with a different color for each cluster will not accurately determine the optimal value of k. PCA is a technique that reduces the dimensionality of the data by transforming it into a new set of features that capture the most variance in the data. However, PCA may not preserve the original structure and distances of the data, and it may lose some information in the process. Therefore, running the k-means clustering algorithm on the PCA components may not reflect the true clusters in the data. Moreover, using only the first two PCA components may not capture enough variance to represent the data well. Furthermore, creating a scatter plot may not be reliable, as it depends on the subjective judgment of the data scientist to decide when the clusters look reasonably separated2.

Option B: Calculating the PCA components and creating a line plot of the number of components against the explained variance will not determine the optimal value of k. This approach is used to determine the optimal number of PCA components to use for dimensionality reduction, not for clustering. The explained variance is the ratio of the variance of each PCA component to the total variance of the data. The optimal number of PCA components is the point where adding more components does not significantly increase the explained variance. However, this number may not correspond to the optimal number of clusters, as PCA and k-means clustering have different objectives and assumptions2.

Option C: Creating a t-distributed stochastic neighbor embedding (t-SNE) plot for a range of perplexity values will not determine the optimal value of k. t-SNE is a technique that reduces the dimensionality of the data by embedding it into a lower-dimensional space, such as a two-dimensional plane. t-SNE preserves the local structure and distances of the data, and it can reveal clusters and patterns in the data.

However, t-SNE does not assign labels or centroids to the clusters, and it does not provide a measure of how well the clusters fit the data. Therefore, t-SNE cannot determine the optimal number of clusters, as it only visualizes the data. Moreover, t-SNE depends on the perplexity parameter, which is a measure of how many neighbors each point considers. The perplexity parameter can affect the shape and size of the clusters, and there is no optimal value for it. Therefore, creating a t-SNE plot for a range of perplexity values may not be consistent or reliable3.

References:

1: How to Determine the Optimal K for K-Means?

2: Principal Component Analysis

3: t-Distributed Stochastic Neighbor Embedding

**NEW QUESTION # 203**

.....

**Test AWS-Certified-Machine-Learning-Specialty Registration:** <https://www.dumpexam.com/AWS-Certified-Machine-Learning-Specialty-valid-torrent.html>

- [illegible]

P.S. Free & New AWS-Certified-Machine-Learning-Specialty dumps are available on Google Drive shared by DumpExam: <https://drive.google.com/open?id=1efS09diuprE7oc3H6RP7G86idoTM3BK0>