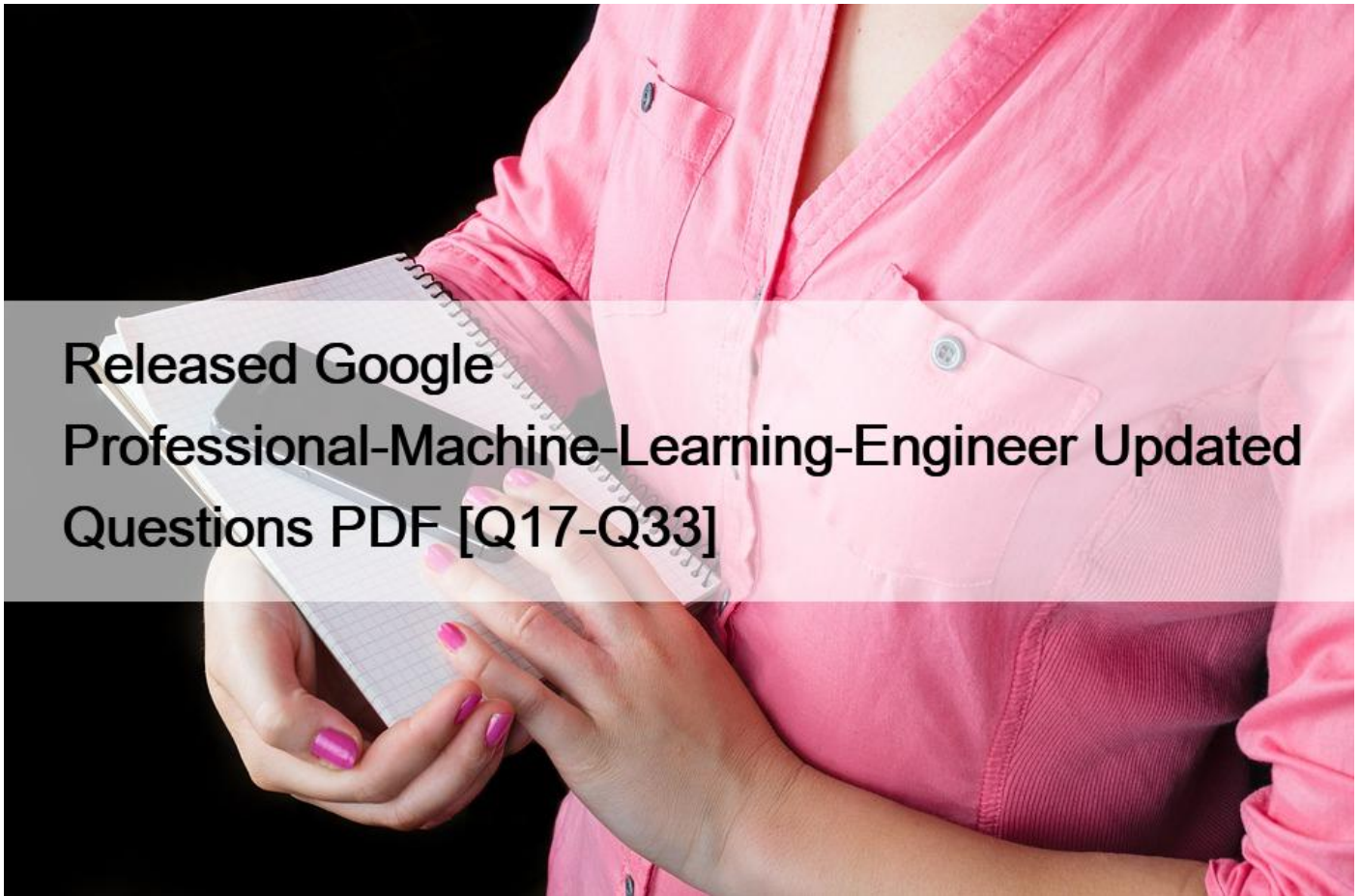


Released Google Professional-Machine-Learning-Engineer Updated Questions PDF [Q17-Q33]



Released Google Professional-Machine-Learning-Engineer Updated Questions PDF
Professional-Machine-Learning-Engineer Dumps and Practice Test (75 Exam Questions)

For more info read reference:
[Google Web Services Website](#)

NEW QUESTION 17

A company is using Amazon Textract to extract textual data from thousands of scanned text-heavy legal documents daily. The company uses this information to process loan applications automatically. Some of the documents fail business validation and are returned to human reviewers, who investigate the errors. This activity increases the time to process the loan applications.

What should the company do to reduce the processing time of loan applications?

* Configure Amazon Textract to route low-confidence predictions to Amazon SageMaker Ground Truth.

Perform a manual review on those words before performing a business validation.

- * Use an Amazon Textract synchronous operation instead of an asynchronous operation.
- * Configure Amazon Textract to route low-confidence predictions to Amazon Augmented AI (Amazon A2I).

Perform a manual review on those words before performing a business validation.

- * Use Amazon Rekognition's feature to detect text in an image to extract the data from scanned images. Use this information to process the loan applications.

NEW QUESTION 18

You are an ML engineer at a bank that has a mobile application. Management has asked you to build an ML-based biometric authentication for the app that verifies a customer's identity based on their fingerprint. Fingerprints are considered highly sensitive personal information and cannot be downloaded and stored into the bank databases. Which learning strategy should you recommend to train and deploy this ML model?

- * Differential privacy
- * Federated learning
- * MD5 to encrypt data
- * Data Loss Prevention API

NEW QUESTION 19

Your data science team needs to rapidly experiment with various features, model architectures, and hyperparameters. They need to track the accuracy metrics for various experiments and use an API to query the metrics over time. What should they use to track and report their experiments while minimizing manual effort?

- * Use Kubeflow Pipelines to execute the experiments Export the metrics file, and query the results using the Kubeflow Pipelines API.
- * Use AI Platform Training to execute the experiments Write the accuracy metrics to BigQuery, and query the results using the BigQueryAPI.
- * Use AI Platform Training to execute the experiments Write the accuracy metrics to Cloud Monitoring, and query the results using the Monitoring API.
- * Use AI Platform Notebooks to execute the experiments. Collect the results in a shared Google Sheets file, and query the results using the Google Sheets API

NEW QUESTION 20

An online reseller has a large, multi-column dataset with one column missing 30% of its data. A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data.

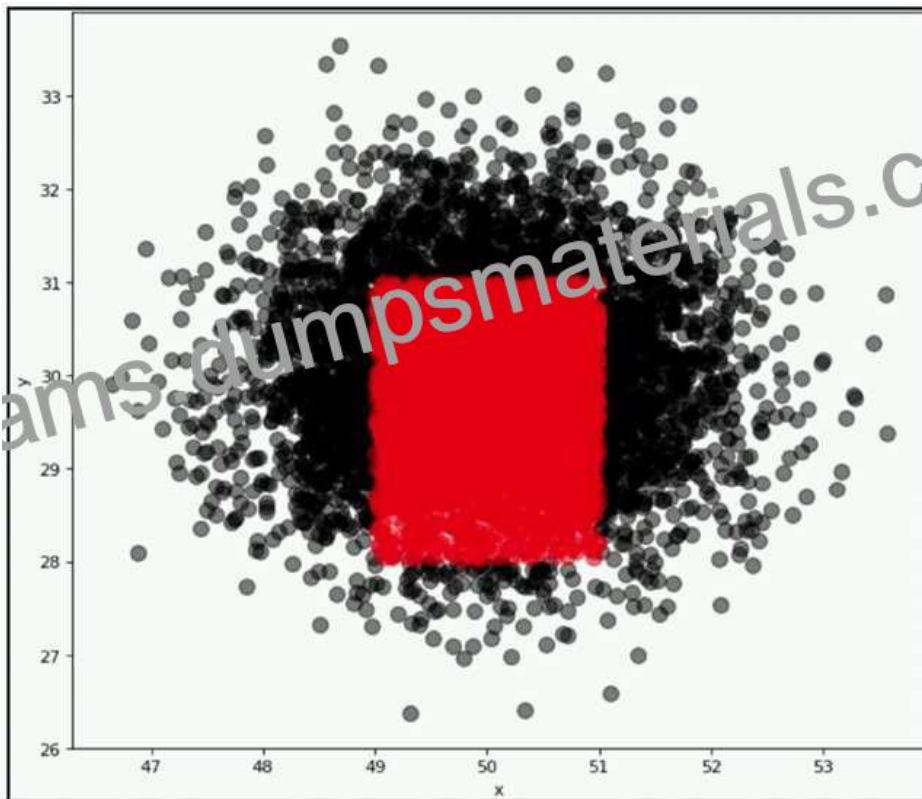
Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

- * Listwise deletion
- * Last observation carried forward
- * Multiple imputation
- * Mean substitution

Explanation/Reference: <https://worldwidescience.org/topicpages/i/imputing+missing+values.html>

NEW QUESTION 21

A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a machine learning specialist will build a binary classifier based on two features: age of account, denoted by x , and transaction month, denoted by y . The class distributions are illustrated in the provided figure. The positive class is portrayed in red, while the negative class is portrayed in black.



Which model would have the HIGHEST accuracy?

- * Linear support vector machine (SVM)
- * Decision tree
- * Support vector machine (SVM) with a radial basis function kernel
- * Single perceptron with a Tanh activation function

NEW QUESTION 22

You work for a social media company. You need to detect whether posted images contain cars. Each training example is a member of exactly one class. You have trained an object detection neural network and deployed the model version to AI Platform Prediction for evaluation. Before deployment, you created an evaluation job and attached it to the AI Platform Prediction model version. You notice that the precision is lower than your business requirements allow. How should you adjust the model's final layer softmax threshold to increase precision?

- * Increase the recall
- * Decrease the recall.
- * Increase the number of false positives
- * Decrease the number of false negatives

NEW QUESTION 23

A Machine Learning Specialist kicks off a hyperparameter tuning job for a tree-based ensemble model using Amazon SageMaker with Area Under the ROC Curve (AUC) as the objective metric. This workflow will eventually be deployed in a pipeline that retrains and tunes hyperparameters each night to model click-through on data that goes stale every 24 hours.

With the goal of decreasing the amount of time it takes to train these models, and ultimately to decrease costs, the Specialist wants to reconfigure the input hyperparameter range(s).

Which visualization will accomplish this?

- * A histogram showing whether the most important input feature is Gaussian.
- * A scatter plot with points colored by target variable that uses t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the large number of input variables in an easier-to-read dimension.
- * A scatter plot showing the performance of the objective metric over each training iteration.
- * A scatter plot showing the correlation between maximum tree depth and the objective metric.

NEW QUESTION 24

You have written unit tests for a Kubeflow Pipeline that require custom libraries. You want to automate the execution of unit tests with each new push to your development branch in Cloud Source Repositories. What should you do?

- * Write a script that sequentially performs the push to your development branch and executes the unit tests on Cloud Run
- * Using Cloud Build, set an automated trigger to execute the unit tests when changes are pushed to your development branch.
- * Set up a Cloud Logging sink to a Pub/Sub topic that captures interactions with Cloud Source Repositories. Configure a Pub/Sub trigger for Cloud Run, and execute the unit tests on Cloud Run.
- * Set up a Cloud Logging sink to a Pub/Sub topic that captures interactions with Cloud Source Repositories. Execute the unit tests using a Cloud Function that is triggered when messages are sent to the Pub/Sub topic

NEW QUESTION 25

A company is observing low accuracy while training on the default built-in image classification algorithm in Amazon SageMaker. The Data Science team wants to use an Inception neural network architecture instead of a ResNet architecture.

Which of the following will accomplish this? (Choose two.)

- * Customize the built-in image classification algorithm to use Inception and use this for model training.
- * Create a support case with the SageMaker team to change the default image classification algorithm to Inception.
- * Bundle a Docker container with TensorFlow Estimator loaded with an Inception network and use this for model training.
- * Use custom code in Amazon SageMaker with TensorFlow Estimator to load the model with an Inception network, and use this for model training.
- * Download and apt-get install the inception network code into an Amazon EC2 instance and use this instance as a Jupyter notebook in Amazon SageMaker.

NEW QUESTION 26

You have a demand forecasting pipeline in production that uses Dataflow to preprocess raw data prior to model training and prediction. During preprocessing, you employ Z-score normalization on data stored in BigQuery and write it back to BigQuery. New training data is added every week. You want to make the process more efficient by minimizing computation time and manual intervention. What should you do?

- * Normalize the data using Google Kubernetes Engine
- * Translate the normalization algorithm into SQL for use with BigQuery
- * Use the normalizer_fn argument in TensorFlow's Feature Column API
- * Normalize the data with Apache Spark using the Dataproc connector for BigQuery

NEW QUESTION 27

You work for a public transportation company and need to build a model to estimate delay times for multiple transportation routes. Predictions are served directly to users in an app in real time. Because different seasons and population increases impact the data

relevance, you will retrain the model every month. You want to follow Google-recommended best practices. How should you configure the end-to-end architecture of the predictive model?

- * Configure Kubeflow Pipelines to schedule your multi-step workflow from training to deploying your model.
- * Use a model trained and deployed on BigQuery ML and trigger retraining with the scheduled query feature in BigQuery
- * Write a Cloud Functions script that launches a training and deploying job on AI Platform that is triggered by Cloud Scheduler
- * Use Cloud Composer to programmatically schedule a Dataflow job that executes the workflow from training to deploying your model

NEW QUESTION 28

You have a functioning end-to-end ML pipeline that involves tuning the hyperparameters of your ML model using AI Platform, and then using the best-tuned parameters for training. Hypertuning is taking longer than expected and is delaying the downstream processes. You want to speed up the tuning job without significantly compromising its effectiveness. Which actions should you take?

Choose 2 answers

- * Decrease the number of parallel trials
- * Decrease the range of floating-point values
- * Set the early stopping parameter to TRUE
- * Change the search algorithm from Bayesian search to random search.
- * Decrease the maximum number of trials during subsequent training phases.

NEW QUESTION 29

A Machine Learning Specialist receives customer data for an online shopping website. The data includes demographics, past visits, and locality information. The Specialist must develop a machine learning approach to identify the customer shopping patterns, preferences, and trends to enhance the website for better service and smart recommendations.

Which solution should the Specialist recommend?

- * Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in the customer database.
- * A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database.
- * Collaborative filtering based on user interactions and correlations to identify patterns in the customer database.
- * Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database.

Explanation

NEW QUESTION 30

You work for a global footwear retailer and need to predict when an item will be out of stock based on historical inventory data. Customer behavior is highly dynamic since footwear demand is influenced by many different factors. You want to serve models that are trained on all available data, but track your performance on specific subsets of data before pushing to production. What is the most streamlined and reliable way to perform this validation?

- * Use the TFX ModelValidator tools to specify performance metrics for production readiness
- * Use k-fold cross-validation as a validation strategy to ensure that your model is ready for production.
- * Use the last relevant week of data as a validation set to ensure that your model is performing accurately on current data
- * Use the entire dataset and treat the area under the receiver operating characteristics curve (AUC ROC) as the main metric.

NEW QUESTION 31

You work on a growing team of more than 50 data scientists who all use AI Platform. You are designing a strategy to organize your jobs, models, and versions in a clean and scalable way. Which strategy should you choose?

- * Set up restrictive IAM permissions on the AI Platform notebooks so that only a single user or group can access a given instance.
- * Separate each data scientist's work into a different project to ensure that the jobs, models, and versions created by each data scientist are accessible only to that user.
- * Use labels to organize resources into descriptive categories. Apply a label to each created resource so that users can filter the results by label when viewing or monitoring the resources
- * Set up a BigQuery sink for Cloud Logging logs that is appropriately filtered to capture information about AI Platform resource usage In BigQuery create a SQL view that maps users to the resources they are using.

NEW QUESTION 32

A Machine Learning Specialist is working with a large cybersecurity company that manages security events in real time for companies around the world. The cybersecurity company wants to design a solution that will allow it to use machine learning to score malicious events as anomalies on the data as it is being ingested. The company also wants be able to save the results in its data lake for later processing and analysis.

What is the MOST efficient way to accomplish these tasks?

- * Ingest the data using Amazon Kinesis Data Firehose, and use Amazon Kinesis Data Analytics Random Cut Forest (RCF) for anomaly detection. Then use Kinesis Data Firehose to stream the results to Amazon S3.
- * Ingest the data into Apache Spark Streaming using Amazon EMR, and use Spark MLlib with k-means to perform anomaly detection. Then store the results in an Apache Hadoop Distributed File System (HDFS) using Amazon EMR with a replication factor of three as the data lake.
- * Ingest the data and store it in Amazon S3. Use AWS Batch along with the AWS Deep Learning AMIs to train a k-means model using TensorFlow on the data in Amazon S3.
- * Ingest the data and store it in Amazon S3. Have an AWS Glue job that is triggered on demand transform the new data. Then use the built-in Random Cut Forest (RCF) model within Amazon SageMaker to detect anomalies in the data.

NEW QUESTION 33

A Data Scientist is training a multilayer perception (MLP) on a dataset with multiple classes. The target class of interest is unique compared to the other classes within the dataset, but it does not achieve an acceptable recall metric. The Data Scientist has already tried varying the number and size of the MLP's hidden layers, which has not significantly improved the results. A solution to improve recall must be implemented as quickly as possible.

Which techniques should be used to meet these requirements?

- * Gather more data using Amazon Mechanical Turk and then retrain
- * Train an anomaly detection model instead of an MLP
- * Train an XGBoost model instead of an MLP
- * Add class weights to the MLP's loss function and then retrain

How to Prepare For Professional Machine Learning Engineer - Google Preparation Guide for Professional Machine Learning Engineer - Google Introduction for Professional Machine Learning Engineer - Google

A Professional Machine Learning Engineer designs, builds, and productionizes ML models to solve business challenges using Google Cloud technologies and knowledge of proven ML models and techniques. The ML Engineer is proficient in all aspects of model architecture, data pipeline interaction, and metrics interpretation and needs familiarity with application development, infrastructure management, data engineering, and security.

The Professional Machine Learning Engineer exam assesses your ability to:

- Automate & orchestrate ML pipelines- Prepare and process data- Frame ML problems- Monitor, optimize, and maintain ML solutions

We prepare **Google Professional-Machine-Learning-Engineer practice exams** and **Google Professional-Machine-Learning-Engineer practice exams** to prepare you for all these requirements.

Professional-Machine-Learning-Engineer Exam Dumps Pass with Updated 2022 Certified Exam Questions:

<https://www.dumpsmaterials.com/Professional-Machine-Learning-Engineer-real-torrent.html>]