Google.Professional-Machine-Learning-Engineer.v2023-04-15.q71

Exam Code:	Professional-Machine-Learning-Engineer
Exam Name:	Google Professional Machine Learning Engineer
Certification Provider:	Google
Free Question Number:	71
Version:	v2023-04-15
# of views:	1239
# of Questions views:	710
https://www.freepdfdumps.com/Google.Professional-Machine-Learning-Engineer.v2023-04-15.q71.html	

NEW QUESTION: 1
You are responsible for building a unified analytics environment across a variety of on-premises data marts. Your company is experiencing data quality and security challenges when integrating data across the servers, caused by the use of a wide range of disconnected tools and temporary solutions. You need a fully managed, cloud-native data integration service that will lower the total cost of work and reduce repetitive work. Some members on your team prefer a codeless interface for building Extract, Transform, Load (ETL) process. Which service should you use?
A. Dataflow
B. Dataprep
C. Apache Flink
D. Cloud Data Fusion
Answer: D (LEAVE A REPLY)
https://cloud.google.com/data-fusion/docs/concepts/overview#using_the_code-free_web_ui

NEW QUESTION: 2
You are building a real-time prediction engine that streams files which may contain Personally Identifiable Information (PII) to Google Cloud. You want to use the Cloud Data Loss Prevention (DLP) API to scan the files. How should you ensure that the PII is not accessible by unauthorized individuals?
A. Create two buckets of data Sensitive and Non-sensitive Write all data to the Non-sensitive bucket Periodically conduct a bulk scan of that bucket using the DLP API, and move the sensitive data to the Sensitive bucket
B. Create three buckets of data: Quarantine, Sensitive, and Non-sensitive Write all data to the Quarantine bucket.

C. Periodically conduct a bulk scan of that bucket using the DLP API, and move the data to either the Sensitive or Non-Sensitive bucket
D. Stream all files to Google CloudT and then write the data to BigQuery Periodically conduct a bulk scan of the table using the DLP API.
E. Stream all files to Google Cloud, and write batches of the data to BigQuery While the data is being written to BigQuery conduct a bulk scan of the data using the DLP API.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 3
While monitoring your model training's GPU utilization, you discover that you have a native synchronous implementation. The training data is split into multiple files. You want to reduce the execution time of your input pipeline. What should you do?
A. Add caching to the pipeline
B. Increase the network bandwidth
C. Add parallel interleave to the pipeline
D. Increase the CPU load
Answer: D (LEAVE A REPLY)

NEW QUESTION: 4
You are an ML engineer at a large grocery retailer with stores in multiple regions. You have been asked to create an inventory prediction model. Your models features include region, location, historical demand, and seasonal popularity. You want the algorithm to learn from new inventory data on a daily basis. Which algorithms should you use to build the model?
A. Classification
B. Reinforcement Learning
C. Recurrent Neural Networks (RNN)
D. Convolutional Neural Networks (CNN)
Answer: C (LEAVE A REPLY)
"algorithm to learn from new inventory data on a daily basis" = time series model , best option to deal with time series is forsure RNN
https://builtin.com/data-science/recurrent-neural-networks-and-lstm

NEW QUESTION: 5
You work for an online retail company that is creating a visual search engine. You have set up an end-to-end ML pipeline on Google Cloud to classify whether an image contains your company's product. Expecting the release of new products in the near future, you configured a retraining functionality in the pipeline so that new data can be fed into your ML models. You also want to use AI Platform's continuous evaluation service to ensure that the models have high accuracy on your test data set. What should you do?
A. Update your test dataset with images of the newer products when your evaluation metrics drop below a pre-decided threshold.

B. Replace your test dataset with images of the newer products when they are introduced to retraining.
C. Keep the original test dataset unchanged even if newer products are incorporated into retraining
D. Extend your test dataset with images of the newer products when they are introduced to retraining
Answer: D (LEAVE A REPLY)

NEW QUESTION: 6
You work for a toy manufacturer that has been experiencing a large increase in demand. You need to build an ML model to reduce the amount of time spent by quality control inspectors checking for product defects. Faster defect detection is a priority. The factory does not have reliable Wi-Fi. Your company wants to implement the new ML model as soon as possible. Which model should you use?
A. AutoML Vision model
B. AutoML Vision Edge mobile-high-accuracy-1 model
C. AutoML Vision Edge mobile-low-latency-1 model
D. AutoML Vision Edge mobile-versatile-1 model
Answer: A (LEAVE A REPLY)

NEW QUESTION: 7
You are training a deep learning model for semantic image segmentation with reduced training time. While using a Deep Learning VM Image, you receive the following error: The resource 'projects/deeplearning-platform/zones/europe-west4-c/acceleratorTypes/nvidia-tesla-k80' was not found. What should you do?
A. Ensure that you have preemptible GPU quota in the selected region.
B. Ensure that the required GPU is available in the selected region.
C. Ensure that you have GPU quota in the selected region.
D. Ensure that the selected GPU has enough GPU memory for the workload.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 8
You manage a team of data scientists who use a cloud-based backend system to submit training jobs. This system has become very difficult to administer, and you want to use a managed service instead. The data scientists you work with use many different frameworks, including Keras, PyTorch, theano. Scikit-team, and custom libraries. What should you do?
A. Use the Al Platform custom containers feature to receive training jobs using any framework
B. Configure Kubeflow to run on Google Kubernetes Engine and receive training jobs through TFJob
C. Create a library of VM images on Compute Engine; and publish these images on a centralized repository

D. Set up Slurm workload manager to receive jobs that can be scheduled to run on your cloud infrastructure.
Answer: A (LEAVE A REPLY)
because AI platform supported all the frameworks mentioned. And Kubeflow is not managed service in GCP. https://cloud.google.com/ai-platform/training/docs/getting-started-pytorch https://cloud.google.com/ai-platform/training/docs/containers-overview#advantages_of_custom_containers Use the ML framework of your choice. If you can't find an AI Platform Training runtime version that supports the ML framework you want to use, then you can build a custom container that installs your chosen framework and use it to run jobs on AI Platform Training.

NEW QUESTION: 9
You need to train a natural language model to perform text classification on product descriptions that contain millions of examples and 100,000 unique words. You want to preprocess the words individually so that they can be fed into a recurrent neural network. What should you do?
A. Assign a numerical value to each word from 1 to 100,000 and feed the values as inputs in your model.
B. Create a hot-encoding of words, and feed the encodings into your model.
C. Sort the words by frequency of occurrence, and use the frequencies as the encodings in your model.
D. Identify word embeddings from a pre-trained model, and use the embeddings in your model.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 10
You work at a subscription-based company. You have trained an ensemble of trees and neural networks to predict customer churn, which is the likelihood that customers will not renew their yearly subscription. The average prediction is a 15% churn rate, but for a particular customer the model predicts that they are 70% likely to churn. The customer has a product usage history of 30%, is located in New York City, and became a customer in 1997. You need to explain the difference between the actual prediction, a 70% churn rate, and the average prediction. You want to use Vertex Explainable AI. What should you do?
A. Configure integrated gradients explanations on Vertex Explainable AI.
B. Measure the effect of each feature as the weight of the feature multiplied by the feature value.
C. Train local surrogate models to explain individual predictions.
D. Configure sampled Shapley explanations on Vertex Explainable AI.
Answer: (SHOW ANSWER)

NEW QUESTION: 11
You are experimenting with a built-in distributed XGBoost model in Vertex AI Workbench user-managed notebooks. You use BigQuery to split your data into training and validation sets using the following queries:

```
CREATE OR REPLACE TABLE 'myproject.mydataset.training' AS
(SELECT * FROM 'myproject.mydataset.mytable' WHERE RAND() <= 0.8);
CREATE OR REPLACE TABLE 'myproject.mydataset.validation' AS
(SELECT * FROM 'myproject.mydataset.mytable' WHERE RAND() <= 0.2);
```

After training the model, you achieve an area under the receiver operating characteristic curve (AUC ROC) value of 0.8, but after deploying the model to production, you notice that your model performance has dropped to an AUC ROC value of 0.65. What problem is most likely occurring?

A. There is training-serving skew in your production environment.

B. There is not a sufficient amount of training data.

C. The tables that you created to hold your training and validation records share some records, and you may not be using all the data in your initial table.

D. The RAND() function generated a number that is less than 0.2 in both instances, so every record in the validation table will also be in the training table.

Answer: A (LEAVE A REPLY)

This is the most likely problem that is occurring based on the information provided. Training-serving skew occurs when the distribution of the data used for training and the data used for serving the model in production are different. This can result in a drop in model performance when the model is deployed to production. It's also possible that the model is overfitting during training.

It is not a problem of insufficient amount of data because the data is split by using the BigQuery and it's not a problem of sharing some records between tables because it is not mentioned that the data is shared in the question.

The problem D is also not correct as the RAND() function is used to split the data but it doesn't mean that every record in the validation table will also be in the training table.

NEW QUESTION: 12

You work for a gaming company that manages a popular online multiplayer game where teams with 6 players play against each other in 5-minute battles. There are many new players every day. You need to build a model that automatically assigns available players to teams in real time. User research indicates that the game is more enjoyable when battles have players with similar skill levels. Which business metrics should you track to measure your model's performance? (Choose One Correct Answer)

A. Average time players wait before being assigned to a team

B. Precision and recall of assigning players to teams based on their predicted versus actual ability

C. Rate of return as measured by additional revenue generated minus the cost of developing a new model

D. User engagement as measured by the number of battles played daily per user

Answer: D (LEAVE A REPLY)

NEW QUESTION: 13

You work for a large social network service provider whose users post articles and discuss news. Millions of comments are posted online each day, and more than 200 human moderators constantly review comments and flag those that are inappropriate. Your team is building an ML model to help human moderators check content on the platform. The model scores each comment and flags suspicious comments to be reviewed by a human. Which metric(s) should you use to monitor the model's performance?

A. Number of messages flagged by the model per minute

B. Precision and recall estimates based on a random sample of 0.1% of raw messages each minute sent to a human for review

C. Precision and recall estimates based on a sample of messages flagged by the model as potentially inappropriate each minute

D. Number of messages flagged by the model per minute confirmed as being inappropriate by humans.

Answer: (SHOW ANSWER)

NEW QUESTION: 14

You are an ML engineer in the contact center of a large enterprise. You need to build a sentiment analysis tool that predicts customer sentiment from recorded phone conversations. You need to identify the best approach to building a model while ensuring that the gender, age, and cultural differences of the customers who called the contact center do not impact any stage of the model development pipeline and results. What should you do?

A. Extract sentiment directly from the voice recordings

B. Convert the speech to text and build a model based on the words

C. Convert the speech to text and extract sentiments based on the sentences

D. Convert the speech to text and extract sentiment using syntactical analysis

Answer: (SHOW ANSWER)

To ensure that gender, age, and cultural differences of the customers who called the contact center do not impact any stage of the model development pipeline and results, it is important to focus on the meaning and context of the conversation, rather than the characteristics of the speaker.

Converting the speech to text and then using syntactical analysis to extract sentiment will allow you to focus on the meaning and context of the conversation, rather than characteristics of the speaker. This approach will also give you more data to work with, as you can analyze the entire conversation, rather than just the voice recordings.

NEW QUESTION: 15

You work for an online travel agency that also sells advertising placements on its website to other companies.

You have been asked to predict the most relevant web banner that a user should see next. Security is important to your company. The model latency requirements are 300ms@p99, the inventory is thousands of web banners, and your exploratory analysis has shown that navigation

context is a good predictor. You want to Implement the simplest solution. How should you configure the prediction pipeline?
A. Embed the client on the website, and then deploy the model on AI Platform Prediction.
B. Embed the client on the website, deploy the gateway on App Engine, and then deploy the model on AI Platform Prediction.
C. Embed the client on the website, deploy the gateway on App Engine, deploy the database on Cloud Bigtable for writing and for reading the user's navigation context, and then deploy the model on AI Platform Prediction.
D. Embed the client on the website, deploy the gateway on App Engine, deploy the database on Memorystore for writing and for reading the user's navigation context, and then deploy the model on Google Kubernetes Engine.
Answer: C (LEAVE A REPLY)
https://medium.com/google-cloud/secure-cloud-run-cloud-functions-and-app-engine-with-api-key-73c57bededd1

NEW QUESTION: 16
You need to build an ML model for a social media application to predict whether a user's submitted profile photo meets the requirements. The application will inform the user if the picture meets the requirements. How should you build a model to ensure that the application does not falsely accept a non-compliant picture?
A. Use Vertex AI Workbench user-managed notebooks to build a custom model that has three times as many examples of pictures that do not meet the profile photo requirements.
B. Use Vertex AI Workbench user-managed notebooks to build a custom model that has three times as many examples of pictures that meet the profile photo requirements.
C. Use AutoML to optimize the model's F1 score in order to balance the accuracy of false positives and false negatives.
D. Use AutoML to optimize the model's recall in order to minimize false negatives.
Answer: (SHOW ANSWER)

Valid Professional-Machine-Learning-Engineer Dumps shared by Actual4test.com for Helping Passing Professional-Machine-Learning-Engineer Exam! Actual4test.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the Actual4test.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com Professional-Machine-Learning-Engineer dumps with Test Engine here: https://www.actual4test.com/Professional-Machine-Learning-Engineer_examcollection.html (**300** Q&As Dumps, **30%OFF** **Special Discount:** **Freepdfdumps**)

NEW QUESTION: 17

You recently joined an enterprise-scale company that has thousands of datasets. You know that there are accurate descriptions for each table in BigQuery, and you are searching for the proper BigQuery table to use for a model you are building on AI Platform. How should you find the data that you need?

A. Use Data Catalog to search the BigQuery datasets by using keywords in the table description.

B. Tag each of your model and version resources on AI Platform with the name of the BigQuery table that was used for training.

C. Maintain a lookup table in BigQuery that maps the table descriptions to the table ID. Query the lookup table to find the correct table ID for the data that you need.

D. Execute a query in BigQuery to retrieve all the existing table names in your project using the

Answer: A (LEAVE A REPLY)

INFORMATION_SCHEMA metadata tables that are native to BigQuery. Use the result o find the table that you need.

Explanation:

A should be the way to go for large datasets --This is also good but it is legacy way of checking:- NFORMATION_SCHEMA contains these views for table metadata: TABLES and TABLE_OPTIONS for metadata about tables. COLUMNS and COLUMN_FIELD_PATHS for metadata about columns and fields. PARTITIONS for metadata about table partitions (Preview)

NEW QUESTION: 18

During batch training of a neural network, you notice that there is an oscillation in the loss. How should you adjust your model to ensure that it converges?

A. Increase the size of the training batch

B. Decrease the size of the training batch

C. Increase the learning rate hyperparameter

D. Decrease the learning rate hyperparameter

Answer: (SHOW ANSWER)

https://developers.google.com/machine-learning/crash-course/introduction-to-neural-networks/playground-exercises

NEW QUESTION: 19

You are working on a Neural Network-based project. The dataset provided to you has columns with different ranges. While preparing the data for model training, you discover that gradient optimization is having difficulty moving weights to a good solution. What should you do?

A. Use feature construction to combine the strongest features.

B. Use the representation transformation (normalization) technique.

C. Improve the data cleaning step by removing features with missing values.

D. Change the partitioning step to reduce the dimension of the test set and have a larger training set.

Answer: B (LEAVE A REPLY)

https://developers.google.com/machine-learning/data-prep/transform/transform-numeric

- NN models needs features with close ranges
- SGD converges well using features in [0, 1] scale
- The question specifically mention "different ranges"
Documentation - https://developers.google.com/machine-learning/data-prep/transform/transform-numeric

NEW QUESTION: 20
Your team has been tasked with creating an ML solution in Google Cloud to classify support requests for one of your platforms. You analyzed the requirements and decided to use TensorFlow to build the classifier so that you have full control of the model's code, serving, and deployment. You will use Kubeflow pipelines for the ML platform. To save time, you want to build on existing resources and use managed services instead of building a completely new model. How should you build the classifier?
A. Use the Natural Language API to classify support requests
B. Use AutoML Natural Language to build the support requests classifier
C. Use an established text classification model on AI Platform to perform transfer learning
D. Use an established text classification model on AI Platform as-is to classify support requests
Answer: (SHOW ANSWER)
the model cannot work as-is as the classes to predict will likely not be the same; we need to use transfer learning to retrain the last layer and adapt it to the classes we need

NEW QUESTION: 21
You have been asked to develop an input pipeline for an ML training model that processes images from disparate sources at a low latency. You discover that your input data does not fit in memory. How should you create a dataset following Google-recommended best practices?
A. Create a tf.data.Dataset.prefetch transformation
B. Convert the images to tf .Tensor Objects, and then run Dataset. from_tensor_slices{).
C. Convert the images to tf .Tensor Objects, and then run tf. data. Dataset. from_tensors ().
D. Convert the images Into TFRecords, store the images in Cloud Storage, and then use the tf. data API to read the images for training
Answer: D (LEAVE A REPLY)
Cite from Google Pag: to construct a Dataset from data in memory, use tf.data.Dataset.from_tensors() or tf.data.Dataset.from_tensor_slices(). When input data is stored in a file (not in memory), the recommended TFRecord format, you can use tf.data.TFRecordDataset(). tf.data.Dataset is for data in memory. tf.data.TFRecordDataset is for data in non-memory storage.
https://cloud.google.com/architecture/ml-on-gcp-best-practices#store-image-video-audio-and-unstructured-data-on-cloud-storage
" Store image, video, audio and unstructured data on Cloud Storage Store these data in large container formats on Cloud Storage. This applies to sharded TFRecord files if you're using TensorFlow, or Avro files if you're using any other framework. Combine many individual images,

videos, or audio clips into large files, as this will improve your read and write throughput to Cloud Storage. Aim for files of at least 100mb, and between 100 and 10,000 shards. To enable data management, use Cloud Storage buckets and directories to group the shards. "

NEW QUESTION: 22
You recently built the first version of an image segmentation model for a self-driving car. After deploying the model, you observe a decrease in the area under the curve (AUC) metric. When analyzing the video recordings, you also discover that the model fails in highly congested traffic but works as expected when there is less traffic. What is the most likely reason for this result?
A. AUC is not the correct metric to evaluate this classification model.
B. Too much data representing congested areas was used for model training.
C. Gradients become small and vanish while backpropagating from the output to input nodes.
D. The model is overfitting in areas with less traffic and underfitting in areas with more traffic.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 23
You are the Director of Data Science at a large company, and your Data Science team has recently begun using the Kubeflow Pipelines SDK to orchestrate their training pipelines. Your team is struggling to integrate their custom Python code into the Kubeflow Pipelines SDK. How should you instruct them to proceed in order to quickly integrate their code with the Kubeflow Pipelines SDK?
A. Deploy the custom Python code to Cloud Functions, and use Kubeflow Pipelines to trigger the Cloud Function.
B. Package the custom Python code into Docker containers, and use the load_component_from_file function to import the containers into the pipeline.
C. Use the predefined components available in the Kubeflow Pipelines SDK to access Dataproc, and run the custom code there.
D. Use the func_to_container_op function to create custom components from the Python code.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 24
You were asked to investigate failures of a production line component based on sensor readings. After receiving the dataset, you discover that less than 1% of the readings are positive examples representing failure incidents. You have tried to train several classification models, but none of them converge. How should you resolve the class imbalance problem?
A. Use the class distribution to generate 10% positive examples
B. Use a convolutional neural network with max pooling and softmax activation
C. Downsample the data with upweighting to create a sample with 10% positive examples
D. Remove negative examples until the numbers of positive and negative examples are equal
Answer: C (LEAVE A REPLY)

https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data#downsampling-and-upweighting
https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data

NEW QUESTION: 25
You are training a TensorFlow model on a structured data set with 100 billion records stored in several CSV files. You need to improve the input/output execution performance. What should you do?
A. Convert the CSV files into shards of TFRecords, and store the data in Cloud Storage
B. Load the data into Cloud Bigtable, and read the data from Bigtable
C. Load the data into BigQuery and read the data from BigQuery.
D. Convert the CSV files into shards of TFRecords, and store the data in the Hadoop Distributed File System (HDFS)
Answer: A (LEAVE A REPLY)

NEW QUESTION: 26
You work for a company that manages a ticketing platform for a large chain of cinemas. Customers use a mobile app to search for movies they're interested in and purchase tickets in the app. Ticket purchase requests are sent to Pub/Sub and are processed with a Dataflow streaming pipeline configured to conduct the following steps:
1. Check for availability of the movie tickets at the selected cinema.
2. Assign the ticket price and accept payment.
3. Reserve the tickets at the selected cinema.
4. Send successful purchases to your database.
Each step in this process has low latency requirements (less than 50 milliseconds). You have developed a logistic regression model with BigQuery ML that predicts whether offering a promo code for free popcorn increases the chance of a ticket purchase, and this prediction should be added to the ticket purchase process. You want to identify the simplest way to deploy this model to production while adding minimal latency. What should you do?
A. Export your model in TensorFlow format, deploy it on Vertex AI, and query the prediction endpoint from your streaming pipeline.
B. Convert your model with TensorFlow Lite (TFLite), and add it to the mobile app so that the promo code and the incoming request arrive together in Pub/Sub.
C. Export your model in TensorFlow format, and add a tfx_bsl.public.beam.RunInference step to the Dataflow pipeline.
D. Run batch inference with BigQuery ML every five minutes on each new set of tickets issued.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 27

You built a custom ML model using scikit-learn. Training time is taking longer than expected. You decide to migrate your model to Vertex AI Training, and you want to improve the model's training time. What should you try out first?
A. Train your model in a distributed mode using multiple Compute Engine VMs.
B. Train your model with DLVM images on Vertex AI, and ensure that your code utilizes NumPy and SciPy internal methods whenever possible.
C. Train your model using Vertex AI Training with GPUs.
D. Migrate your model to TensorFlow, and train it using Vertex AI Training.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 28
Your task is classify if a company logo is present on an image. You found out that 96% of a data does not include a logo. You are dealing with data imbalance problem. Which metric do you use to evaluate to model?
A. F1 Score
B. F Score with higher precision weighting than recall
C. RMSE
D. F Score with higher recall weighted than precision
Answer: D (LEAVE A REPLY)

NEW QUESTION: 29
You lead a data science team at a large international corporation. Most of the models your team trains are large-scale models using high-level TensorFlow APIs on AI Platform with GPUs. Your team usually takes a few weeks or months to iterate on a new version of a model. You were recently asked to review your team's spending. How should you reduce your Google Cloud compute costs without impacting the model's performance?
A. Use AI Platform to run distributed training jobs without checkpoints.
B. Use AI Platform to run distributed training jobs with checkpoints.
C. Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs without checkpoints.
D. Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs with checkpoints.
Answer: (SHOW ANSWER)

NEW QUESTION: 30
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?
A. Translate the normalization algorithm into SQL for use with BigQuery

B. Normalize the data using Google Kubernetes Engine
C. Use the normalizer_fn argument in TensorFlow's Feature Column API
D. Normalize the data with Apache Spark using the Dataproc connector for BigQuery
Answer: (SHOW ANSWER)

NEW QUESTION: 31
You are an ML engineer at an ecommerce company and have been tasked with building a model that predicts how much inventory the logistics team should order each month. Which approach should you take?
A. Use a regression model to predict how much additional inventory should be purchased each month. Give the results to the logistics team at the beginning of the month so they can increase inventory by the amount predicted by the model.
B. Use a clustering algorithm to group popular items together. Give the list to the logistics team so they can increase inventory of the popular items.
C. Use a classification model to classify inventory levels as UNDER_STOCKED, OVER_STOCKED, and CORRECTLY_STOCKED. Give the report to the logistics team each month so they can fine-tune inventory levels.
D. Use a time series forecasting model to predict each item's monthly sales. Give the results to the logistics team so they can base inventory on the amount predicted by the model.
Answer: A (LEAVE A REPLY)

Valid Professional-Machine-Learning-Engineer Dumps shared by Actual4test.com for Helping Passing Professional-Machine-Learning-Engineer Exam! Actual4test.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the Actual4test.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com Professional-Machine-Learning-Engineer dumps with Test Engine here: https://www.actual4test.com/Professional-Machine-Learning-Engineer_examcollection.html (**300** Q&As Dumps, **30%OFF** **Special Discount: Freepdfdumps**)

NEW QUESTION: 32
You work for a company that is developing a new video streaming platform. You have been asked to create a recommendation system that will suggest the next video for a user to watch. After a review by an AI Ethics team, you are approved to start development. Each video asset in your company's catalog has useful metadata (e.g., content type, release date, country), but you do not have any historical user event dat a. How should you build the recommendation system for the first version of the product?
A. Launch the product with machine learning. Use a publicly available dataset such as MovieLens to train a model using the Recommendations AI, and then apply this trained model to your data.

B. Launch the product without machine learning. Present videos to users alphabetically, and start collecting user event data so you can develop a recommender model in the future.
C. Launch the product with machine learning. Generate embeddings for each video by training an autoencoder on the content metadata using TensorFlow. Cluster content based on the similarity of these embeddings, and then recommend videos from the same cluster.
D. Launch the product without machine learning. Use simple heuristics based on content metadata to recommend similar videos to users, and start collecting user event data so you can develop a recommender model in the future.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 33
While performing exploratory data analysis on a dataset, you find that an important categorical feature has 5% null values. You want to minimize the bias that could result from the missing values. How should you handle the missing values?
A. Replace the missing values with a placeholder category indicating a missing value.
B. Move the rows with missing values to your validation dataset.
C. Replace the missing values with the feature's mean.
D. Remove the rows with missing values, and upsample your dataset by 5%.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 34
You are building a model to predict daily temperatures. You split the data randomly and then transformed the training and test datasets. Temperature data for model training is uploaded hourly. During testing, your model performed with 97% accuracy; however, after deploying to production, the model's accuracy dropped to 66%. How can you make your production model more accurate?
A. Normalize the data for the training, and test datasets as two separate steps.
B. Split the training and test data based on time rather than a random split to avoid leakage
C. Add more data to your test set to ensure that you have a fair distribution and sample for testing
D. Apply data transformations before splitting, and cross-validate to make sure that the transformations are applied to both the training and test sets.
Answer: (SHOW ANSWER)
https://community.rapidminer.com/discussion/32592/normalising-data-before-data-split-or-after

NEW QUESTION: 35
You are developing an ML model intended to classify whether X-Ray images indicate bone fracture risk. You have trained on Api Resnet architecture on Vertex AI using a TPU as an accelerator, however you are unsatisfied with the trainning time and use memory usage. You want to quickly iterate your training code but make minimal changes to the code. You also want to minimize impact on the models accuracy. What should you do?
A. Reduce the number of layers in the model architecture

B. Reduce the dimensions of the images used un the model
C. Reduce the global batch size from 1024 to 256
D. Configure your model to use bfloat16 instead float32
Answer: C (LEAVE A REPLY)

NEW QUESTION: 36
You are creating a deep neural network classification model using a dataset with categorical input values. Certain columns have a cardinality greater than 10,000 unique values. How should you encode these categorical values as input into the model?
A. Convert the categorical string data to one-hot hash buckets.
B. Map the categorical variables into a vector of boolean values.
C. Convert each categorical value into a run-length encoded string.
D. Convert each categorical value into an integer value.
Answer: B (LEAVE A REPLY)

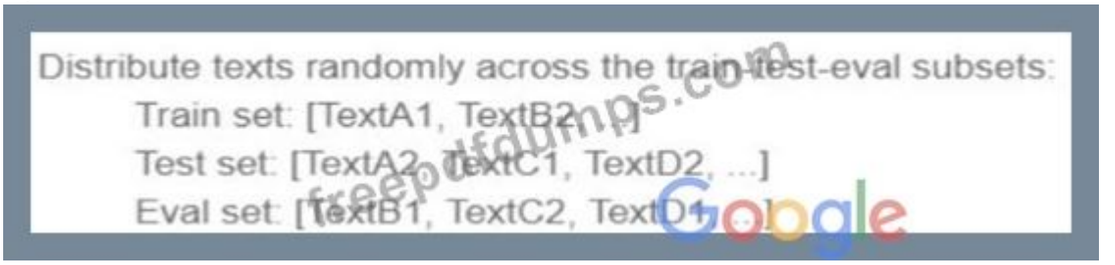
NEW QUESTION: 37
Your team is working on an NLP research project to predict political affiliation of authors based on articles they have written. You have a large training dataset that is structured like this:

```
AuthorA:Political Party A
    TextA1: [SentenceA11, SentenceA12, SentenceA13, ...]
    TextA2: [SentenceA21, SentenceA22, SentenceA23, ...]
    ...
AuthorB:Political Party B
    TextB1: [SentenceB11, SentenceB12, SentenceB13, ...]
    TextB2: [SentenceB21, SentenceB22, SentenceB23, ...]
    ...
AuthorC:Political Party B
    TextC1: [SentenceC11, SentenceC12, SentenceC13, ...]
    TextC2: [SentenceC21, SentenceC22, SentenceC23, ...]
    ...
AuthorD:Political Party A
    TextD1: [SentenceD11, SentenceD12, SentenceD13, ...]
    TextD2: [SentenceD21, SentenceD22, SentenceD23, ...]
    ...
...
```

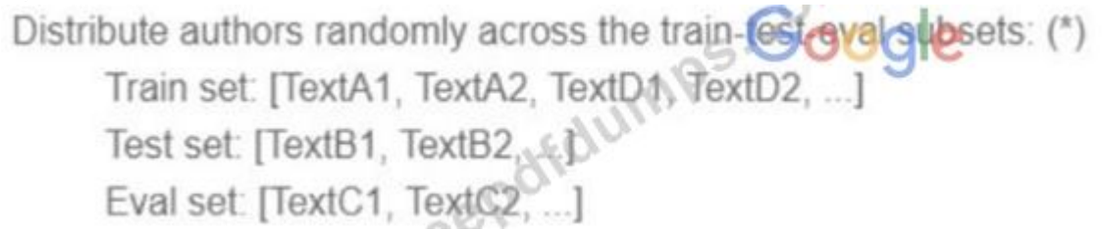
freepdfdumps.com

You followed the standard 80%-10%-10% data distribution across the training, testing, and evaluation subsets. How should you distribute the training examples across the train-test-eval subsets while maintaining the 80-10-10 proportion?

A)

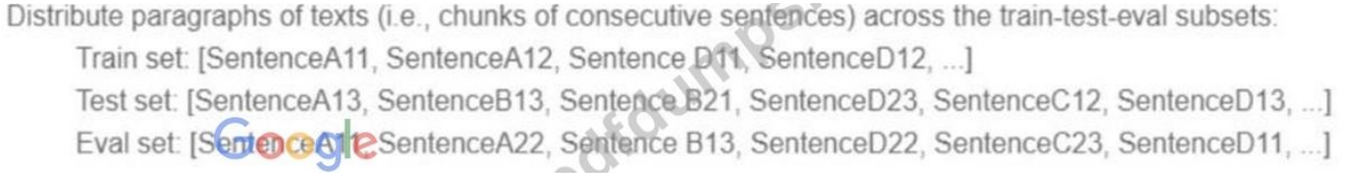


B)



C)

D)



A. Option A

B. Option B

C. Option C

D. Option D

Answer: B (LEAVE A REPLY)

If we just put inside the Training set , Validation set and Test set , randomly Text, Paragraph or sentences the model will have the ability to learn specific qualities about The Author's use of language beyond just his own articles. Therefore the model will mixed up different opinions. Rather if we divided things up a the author level, so that given authors were only on the training data, or only in the test data or only in the validation data. The model will find more difficult to get a high accuracy on the test validation (What is correct and have more sense!). Because it will need to really focus in author by author articles rather than get a single political affiliation based on a bunch of mixed articles from different authors. https://developers.google.com/machine-learning/crash-course/18th-century-literature For example, suppose you are training a model with purchase data from a number of stores. You know, however, that the model will be used primarily to make predictions for stores that are not in the training data. To ensure that the model can

generalize to unseen stores, you should segregate your data sets by stores. In other words, your test set should include only stores different from the evaluation set, and the evaluation set should include only stores different from the training set. https://cloud.google.com/automl-tables/docs/prepare#ml-use

NEW QUESTION: 38
Your organization manages an online message board A few months ago, you discovered an increase in toxic language and bullying on the message board. You deployed an automated text classifier that flags certain comments as toxic or harmful. Now some users are reporting that benign comments referencing their religion are being misclassified as abusive Upon further inspection, you find that your classifier's false positive rate is higher for comments that reference certain underrepresented religious groups. Your team has a limited budget and is already overextended. What should you do?
A. Add synthetic training data where those phrases are used in non-toxic ways
B. Remove the model and replace it with human moderation.
C. Replace your model with a different text classifier.
D. Raise the threshold for comments to be considered toxic or harmful
Answer: A (LEAVE A REPLY)
This approach would help to improve the performance of the classifier by providing it with more examples of the religious phrases being used in non-toxic ways. This would allow the classifier to better differentiate between toxic and non-toxic comments that reference these religious groups. Additionally, synthetic data is a cost-effective way to improve the performance of an existing model without requiring a significant investment in human resources.

NEW QUESTION: 39
Your organization wants to make its internal shuttle service route more efficient. The shuttles currently stop at all pick-up points across the city every 30 minutes between 7 am and 10 am. The development team has already built an application on Google Kubernetes Engine that requires users to confirm their presence and shuttle station one day in advance. What approach should you take?
A. 1. Build a tree-based regression model that predicts how many passengers will be picked up at each shuttle station.
2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the prediction.
B. 1. Build a tree-based classification model that predicts whether the shuttle should pick up passengers at each shuttle station.
2. Dispatch an available shuttle and provide the map with the required stops based on the prediction
C. 1. Define the optimal route as the shortest route that passes by all shuttle stations with confirmed attendance at the given time under capacity constraints.
2 Dispatch an appropriately sized shuttle and indicate the required stops on the map

D. 1. Build a reinforcement learning model with tree-based classification models that predict the presence of passengers at shuttle stops as agents and a reward function around a distance-based metric
2. Dispatch an appropriately sized shuttle and provide the map with the required stops based on the simulated outcome.

Answer: C (LEAVE A REPLY)

This is a case where machine learning would be terrible, as it would not be 100% accurate and some passengers would not get picked up. A simple algorith works better here, and the question confirms customers will be indicating when they are at the stop so no ML required.

NEW QUESTION: 40

You are an ML engineer at a manufacturing company You are creating a classification model for a predictive maintenance use case You need to predict whether a crucial machine will fail in the next three days so that the repair crew has enough time to fix the machine before it breaks. Regular maintenance of the machine is relatively inexpensive, but a failure would be very costly You have trained several binary classifiers to predict whether the machine will fail. where a prediction of 1 means that the ML model predicts a failure.

You are now evaluating each model on an evaluation dataset. You want to choose a model that prioritizes detection while ensuring that more than 50% of the maintenance jobs triggered by your model address an imminent machine failure. Which model should you choose?

A. The model with the highest area under the receiver operating characteristic curve (AUC ROC) and precision greater than 0 5

B. The model with the lowest root mean squared error (RMSE) and recall greater than 0.5.

C. The model with the highest recall where precision is greater than 0.5.

D. The model with the highest precision where recall is greater than 0.5.

Answer: C (LEAVE A REPLY)

In predictive maintenance, the goal is to identify which machines are likely to fail soon, so that the repair crew can fix them before they break. In this context, it is important to prioritize detection, while also ensuring that more than 50% of the maintenance jobs triggered by your model address an imminent machine failure.

Recall is a metric that measures the proportion of actual positive observations that are correctly predicted as such by the model. In this case, recall is a good metric to use because it measures how well the model is able to identify the machines that are likely to fail soon.

Precision is a metric that measures the proportion of positive predictions that are actually true. In this case, precision is also important because it measures how many of the machines that the model predicts will fail soon, actually do fail soon.

By combining these two metrics, you can ensure that your model is able to identify the machines that are likely to fail soon with a high degree of accuracy. In this case, the model with the highest recall where precision is greater than 0.5 will be the best model, as it will have a high ability to identify the machines that are likely to fail soon and also it will have a high degree of accuracy.

Reference:

Recall and Precision
Predictive Maintenance
Metrics for classification

NEW QUESTION: 41
You are building a linear model with over 100 input features, all with values between -1 and 1. You suspect that many features are non-informative. You want to remove the non-informative features from your model while keeping the informative ones in their original form. Which technique should you use?
A. Use Principal Component Analysis to eliminate the least informative features.
B. Use L1 regularization to reduce the coefficients of uninformative features to 0.
C. After building your model, use Shapley values to determine which features are the most informative.
D. Use an iterative dropout technique to identify which features do not degrade the model when removed.
Answer: (SHOW ANSWER)
https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/overview#sampled-shapley

NEW QUESTION: 42
You are an ML engineer at a travel company. You have been researching customers' travel behavior for many years, and you have deployed models that predict customers' vacation patterns. You have observed that customers' vacation destinations vary based on seasonality and holidays; however, these seasonal variations are similar across years. You want to quickly and easily store and compare the model versions and performance statistics across years. What should you do?
A. Store the performance statistics in Cloud SQL. Query that database to compare the performance statistics across the model versions.
B. Store the performance statistics of each version of your models using seasons and years as events in Vertex ML Metadata. Compare the results across the slices.
C. Create versions of your models for each season per year in Vertex AI. Compare the performance statistics across the models in the Evaluate tab of the Vertex AI UI.
D. Store the performance statistics of each pipeline run in Kubeflow under an experiment for each season per year. Compare the results across the experiments in the Kubeflow UI.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 43
You have successfully deployed to production a large and complex TensorFlow model trained on tabular dat a. You want to predict the lifetime value (LTV) field for each subscription stored in the BigQuery table named subscription. subscriptionPurchase in the project named my-fortune500-company-project.

You have organized all your training code, from preprocessing data from the BigQuery table up to deploying the validated model to the Vertex AI endpoint, into a TensorFlow Extended (TFX) pipeline. You want to prevent prediction drift, i.e., a situation when a feature data distribution in production changes significantly over time. What should you do?
A. Add a model monitoring job where 10% of incoming predictions are sampled 24 hours.
B. Add a model monitoring job where 10% of incoming predictions are sampled every hour.
C. Implement continuous retraining of the model daily using Vertex AI Pipelines.
D. Add a model monitoring job where 90% of incoming predictions are sampled 24 hours.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 44
You are designing an ML recommendation model for shoppers on your company's ecommerce website. You will use Recommendations AI to build, test, and deploy your system. How should you develop recommendations that increase revenue while following best practices?
A. Use the "Other Products You May Like" recommendation type to increase the click-through rate
B. Use the "Frequently Bought Together' recommendation type to increase the shopping cart size for each order.
C. Import your user events and then your product catalog to make sure you have the highest quality event stream
D. Because it will take time to collect and record product data, use placeholder values for the product catalog to test the viability of the model.
Answer: (SHOW ANSWER)
Frequently bought together' recommendations aim to up-sell and cross-sell customers by providing product.

NEW QUESTION: 45
You need to design an architecture that serves asynchronous predictions to determine whether a particular mission-critical machine part will fail. Your system collects data from multiple sensors from the machine. You want to build a model that will predict a failure in the next N minutes, given the average of each sensor's data from the past 12 hours. How should you design the architecture?
A. 1. Export the data to Cloud Storage using the BigQuery command-line tool
2. Submit a Vertex AI batch prediction job that uses your trained model in Cloud Storage to perform scoring on the preprocessed data.
3. Export the batch prediction job outputs from Cloud Storage and import them into BigQuery.
B. 1. Export your data to Cloud Storage using Dataflow.
2. Submit a Vertex AI batch prediction job that uses your trained model in Cloud Storage to perform scoring on the preprocessed data.
3. Export the batch prediction job outputs from Cloud Storage and import them into Cloud SQL.

C. 1. HTTP requests are sent by the sensors to your ML model, which is deployed as a microservice and exposes a REST API for prediction
2. Your application queries a Vertex AI endpoint where you deployed your model.
3. Responses are received by the caller application as soon as the model produces the prediction.
D. 1. Events are sent by the sensors to Pub/Sub, consumed in real time, and processed by a Dataflow stream processing pipeline.
2. The pipeline invokes the model for prediction and sends the predictions to another Pub/Sub topic.
3. Pub/Sub messages containing predictions are then consumed by a downstream system for monitoring.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 46
You are an ML engineer at a travel company. You have been researching customers' travel behavior for many years, and you have deployed models that predict customers' vacation patterns. You have observed that customers' vacation destinations vary based on seasonality and holidays; however, these seasonal variations are similar across years. You want to quickly and easily store and compare the model versions and performance statistics across years. What should you do?
A. Store the performance statistics in Cloud SQL. Query that database to compare the performance statistics across the model versions.
B. Create versions of your models for each season per year in Vertex AI. Compare the performance statistics across the models in the Evaluate tab of the Vertex AI UI.
C. Store the performance statistics of each version of your models using seasons and years as events in Vertex ML Metadata. Compare the results across the slices.
D. Store the performance statistics of each pipeline run in Kubeflow under an experiment for each season per year. Compare the results across the experiments in the Kubeflow UI.
Answer: B (LEAVE A REPLY)

Valid Professional-Machine-Learning-Engineer Dumps shared by Actual4test.com for Helping Passing Professional-Machine-Learning-Engineer Exam! Actual4test.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the Actual4test.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com Professional-Machine-Learning-Engineer dumps with Test Engine here: https://www.actual4test.com/Professional-Machine-Learning-Engineer_examcollection.html (**300** Q&As Dumps, **30%OFF** **Special Discount:** **Freepdfdumps**)

NEW QUESTION: 47
You are building an ML model to detect anomalies in real-time sensor dat a. You will use Pub/Sub to handle incoming requests. You want to store the results for analytics and visualization. How should you configure the pipeline?

A. 1 = Dataflow, 2 - AI Platform, 3 = BigQuery
B. 1 = DataProc, 2 = AutoML, 3 = Cloud Bigtable
C. 1 = BigQuery, 2 = AI Platform, 3 = Cloud Storage
D. 1 = BigQuery, 2 = AutoML, 3 = Cloud Functions
Answer: A (LEAVE A REPLY)

NEW QUESTION: 48
You have been asked to build a model using a dataset that is stored in a medium-sized (~10 GB) BigQuery table. You need to quickly determine whether this data is suitable for model development. You want to create a one-time report that includes both informative visualizations of data distributions and more sophisticated statistical analyses to share with other ML engineers on your team. You require maximum flexibility to create your report. What should you do?
A. Use the output from TensorFlow Data Validation on Dataflow to generate the report.
B. Use Vertex AI Workbench user-managed notebooks to generate the report.
C. Use Dataprep to create the report.
D. Use the Google Data Studio to create the report.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 49
You are building a TensorFlow model for a financial institution that predicts the impact of consumer spending on inflation globally. Due to the size and nature of the data, your model is long-running across all types of hardware, and you have built frequent checkpointing into the training process. Your organization has asked you to minimize cost. What hardware should you choose?
A. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with a preemptible v3-8 TPU
B. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with 4 NVIDIA P100 GPUs
C. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with an NVIDIA P100 GPU
D. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with a non-preemptible v3-8 TPU

Answer: C (LEAVE A REPLY)

NEW QUESTION: 50

You trained a text classification model. You have the following SignatureDefs:

```
signature_def['serving_default']:
  The given SavedModel SignatureDef contains the following input(s):
    inputs['text'] tensor_info:
        dtype: DT_STRING
        shape: (-1, 2)
        name: serving_default_text:0
  The given SavedModel SignatureDef contains the following output(s):
    outputs['Softmax'] tensor_info:
        dtype: DT_FLOAT
        shape: (-1, 2)
        name: StatefulPartitionedCall:0
  Method name is: tensorflow/serving/predict
```

You started a TensorFlow-serving component server and tried to send an HTTP request to get a prediction using:

```
headers = {"content-type": "application/json"}
json_response = requests.post('http://localhost:8501/v1/models/text_model:predict', data=data,
headers=headers)
```

freepdfdumps.com

What is the correct way to write the predict request?

A. data = json.dumps({"signature_name": "serving_default'\ "instances": [fab', 'be1, 'cd']]})

B. data = json dumps({"signature_name": "serving_default"! "instances": [['a', 'b', "c", 'd', 'e', 'f']]})

C. data = json.dumps({"signature_name": "serving_default, "instances": [['a', 'b\ 'c'1, [d\ 'e\ T]]})

D. data = json dumps({"signature_name": f,serving_default", "instances": [['a', 'b'], [c\ 'd'], ['e\ T]]})

Answer: (SHOW ANSWER)

https://stackoverflow.com/questions/37956197/what-is-the-negative-index-in-shape-arrays-used-for-tensorflow

NEW QUESTION: 51

You need to quickly build and train a model to predict the sentiment of customer reviews with custom categories without writing code. You do not have enough data to train a model from scratch. The resulting model should have high predictive performance. Which service should you use?

A. AI Platform Training built-in algorithms

B. AI Hub pre-made Jupyter Notebooks

C. AutoML Natural Language

D. Cloud Natural Language API

Answer: C (LEAVE A REPLY)

NEW QUESTION: 52

Your data science team is training a PyTorch model for image classification based on a pre-trained RestNet model. You need to perform hyperparameter tuning to optimize for several parameters. What should you do?
A. Create a Kuberflow Pipelines instance, and run a hyperparameter tuning job on Katib.
B. Convert the model to a TensorFlow model, and run a hyperparameter tuning job on AI Platform.
C. Run a hyperparameter tuning job on AI Platform using custom containers.
D. Convert the model to a Keras model, and run a Keras Tuner job.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 53
You built and manage a production system that is responsible for predicting sales numbers. Model accuracy is crucial, because the production model is required to keep up with market changes. Since being deployed to production, the model hasn't changed; however the accuracy of the model has steadily deteriorated. What issue is most likely causing the steady decline in model accuracy?
A. Poor data quality
B. Lack of model retraining
C. Too few layers in the model for capturing information
D. Incorrect data split ratio during model training, evaluation, validation, and test
Answer: B (LEAVE A REPLY)
Retraining is needed as the market is changing. its how the Model keep updated and predictions accuracy.

NEW QUESTION: 54
You are developing an image recognition model using PyTorch based on ResNet50 architecture. Your code is working fine on your local laptop on a small subsample. Your full dataset has 200k labeled images You want to quickly scale your training workload while minimizing cost. You plan to use 4 V100 GPUs. What should you do? (Choose Correct Answer and Give Reference and Explanation)
A. Configure a Compute Engine VM with all the dependencies that launches the training Train your model with Vertex AI using a custom tier that contains the required GPUs.
B. Package your code with Setuptools. and use a pre-built container Train your model with Vertex AI using a custom tier that contains the required GPUs.
C. Create a Vertex AI Workbench user-managed notebooks instance with 4 V100 GPUs, and use it to train your model
D. Create a Google Kubernetes Engine cluster with a node pool that has 4 V100 GPUs Prepare and submit a TFJob operator to this node pool.
Answer: D (LEAVE A REPLY)
Google Kubernetes Engine (GKE) is a powerful and easy-to-use platform for deploying and managing containerized applications. It allows you to create a cluster of virtual machines that are

pre-configured with the necessary dependencies and resources to run your machine learning workloads. By creating a GKE cluster with a node pool that has 4 V100 GPUs, you can take advantage of the powerful processing capabilities of these GPUs to train your model quickly and efficiently.
You can then use the Kubernetes Framework such as TFJob operator to submit the job of training your model, which will automatically distribute the workload across the available GPUs.
Reference:
Google Kubernetes Engine
TFJob operator
Vertex AI

NEW QUESTION: 55
You are developing an ML model intended to classify whether X-Ray images indicate bone fracture risk. You have trained on Api Resnet architecture on Vertex AI using a TPU as an accelerator, however you are unsatisfied with the trainning time and use memory usage. You want to quickly iterate your training code but make minimal changes to the code. You also want to minimize impact on the models accuracy. What should you do?
A. Reduce the number of layers in the model architecture
B. Reduce the global batch size from 1024 to 256
C. Configure your model to use bfloat16 instead float32
D. Reduce the dimensions of the images used un the model
Answer: B (LEAVE A REPLY)

NEW QUESTION: 56
You work for a large technology company that wants to modernize their contact center. You have been asked to develop a solution to classify incoming calls by product so that requests can be more quickly routed to the correct support team. You have already transcribed the calls using the Speech-to-Text API. You want to minimize data preprocessing and development time. How should you build the model?
A. Build a custom model to identify the product keywords from the transcribed calls, and then run the keywords through a classification algorithm
B. Use the Cloud Natural Language API to extract custom entities for classification
C. Use the AI Platform Training built-in algorithms to create a custom model
D. Use AutoML Natural Language to extract custom entities for classification
Answer: D (LEAVE A REPLY)

NEW QUESTION: 57
You are training a Resnet model on AI Platform using TPUs to visually categorize types of defects in automobile engines. You capture the training profile using the Cloud TPU profiler plugin and observe that it is highly input-bound. You want to reduce the bottleneck and speed up your model training process. Which modifications should you make to the tf .data dataset?

Choose 2 answers

A. Use the interleave option for reading data

B. Reduce the value of the repeat parameter

C. Increase the buffer size for the shuffle option.

D. Set the prefetch option equal to the training batch size

E. Decrease the batch size argument in your transformation

Answer: (SHOW ANSWER)

https://towardsdatascience.com/overcoming-data-preprocessing-bottlenecks-with-tensorflow-data-service-nvidia-dali-and-other-d6321917f851

NEW QUESTION: 58

You are going to train a DNN regression model with Keras APIs using this code:

```
model = tf.keras.Sequential()
model.add(tf.keras.layers.Dense(
    256,
    use_bias=True,
    activation='relu',
    kernel_initializer=None,
    kernel_regularizer=None,
    input_shape=(500,)))
model.add(tf.keras.layers.Dropout(rate=0.25))
model.add(tf.keras.layers.Dense(
    128, use_bias=True,
    activation='relu',
    kernel_initializer='uniform',
    kernel_regularizer='l2'))
model.add(tf.keras.layers.Dropout(rate=0.25))
model.add(tf.keras.layers.Dense(
    2, use_bias=False,
    activation='softmax'))
model.compile(loss='mse')
```

Google
freepdfdumps.com

How many trainable weights does your model have? (The arithmetic below is correct.)

A. 501*256+257*128+128*2=161408

B. 500*256+256*128+128*2 = 161024

C. 501*256+257*128+2 = 161154

D. 500*256*0 25+256*128*0 25+128*2 = 40448

Answer: (SHOW ANSWER)

NEW QUESTION: 59

You are developing models to classify customer support emails. You created models with TensorFlow Estimators using small datasets on your on-premises system, but you now need to train the models using large datasets to ensure high performance. You will port your models to Google Cloud and want to minimize code refactoring and infrastructure overhead for easier migration from on-prem to cloud. What should you do?

A. Use Vertex AI Platform for distributed training

B. Create a cluster on Dataproc for training

C. Create a Managed Instance Group with autoscaling

D. Use Kubeflow Pipelines to train on a Google Kubernetes Engine cluster.

Answer: A (LEAVE A REPLY)

AI platform also contains kubeflow pipelines. you don't need to set up infrastructure to use it. For D you need to set up a kubernetes cluster engine. The question asks us to minimize infrastructure overheard.

NEW QUESTION: 60

You have trained a DNN regressor with TensorFlow to predict housing prices using a set of predictive features. Your default precision is tf.float64, and you use a standard TensorFlow estimator; estimator = tf.estimator.DNNRegressor(feature_columns=[YOUR_LIST_OF_FEATURES], hidden_units-[1024, 512, 256], dropout=None) Your model performs well, but Just before deploying it to production, you discover that your current serving latency is 10ms @ 90 percentile and you currently serve on CPUs. Your production requirements expect a model latency of 8ms @ 90 percentile. You are willing to accept a small decrease in performance in order to reach the latency requirement Therefore your plan is to improve latency while evaluating how much the model's prediction decreases. What should you first try to quickly lower the serving latency?

A. Increase the dropout rate to 0.8 in_PREDICT mode by adjusting the TensorFlow Serving parameters

B. Increase the dropout rate to 0.8 and retrain your model.

C. Switch from CPU to GPU serving

D. Apply quantization to your SavedModel by reducing the floating point precision to tf.float16.

Answer: D (LEAVE A REPLY)

Applying quantization to your SavedModel by reducing the floating point precision can help reduce the serving latency by decreasing the amount of memory and computation required to make a prediction. TensorFlow provides tools such as the tf.quantization module that can be used to quantize models and reduce their precision, which can significantly reduce serving latency without a significant decrease in model performance.

NEW QUESTION: 61

Your team trained and tested a DNN regression model with good results. Six months after deployment, the model is performing poorly due to a change in the distribution of the input dat a. How should you address the input differences in production?

A. Create alerts to monitor for skew, and retrain the model.
B. Perform feature selection on the model, and retrain the model with fewer features
C. Retrain the model, and select an L2 regularization parameter with a hyperparameter tuning service
D. Perform feature selection on the model, and retrain the model on a monthly basis with fewer features
Answer: A (LEAVE A REPLY)
Data drift doesn't necessarily require feature reselection (e.g. by L2 regularization).
https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning#challenges Data values skews: These skews are significant changes in the statistical properties of data, which means that data patterns are changing, and you need to trigger a retraining of the model to capture these changes.
https://developers.google.com/machine-learning/guides/rules-of-ml/#rule_37_measure_trainingserving_skew

Valid Professional-Machine-Learning-Engineer Dumps shared by Actual4test.com for Helping Passing Professional-Machine-Learning-Engineer Exam! Actual4test.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the Actual4test.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com Professional-Machine-Learning-Engineer dumps with Test Engine here: https://www.actual4test.com/Professional-Machine-Learning-Engineer_examcollection.html (**300** Q&As Dumps, **30%OFF** **Special Discount:** **Freepdfdumps**)

NEW QUESTION: 62
You are building a linear regression model on BigQuery ML to predict a customer's likelihood of purchasing your company's products. Your model uses a city name variable as a key predictive component. In order to train and serve the model, your data must be organized in columns. You want to prepare your data using the least amount of coding while maintaining the predictable variables. What should you do?
A. Use Dataprep to transform the state column using a one-hot encoding method, and make each city a column with binary values.
B. Create a new view with BigQuery that does not include a column with city information
C. Use TensorFlow to create a categorical variable with a vocabulary list Create the vocabulary file, and upload it as part of your model to BigQuery ML.
D. Use Cloud Data Fusion to assign each city to a region labeled as 1, 2, 3, 4, or 5r and then use that number to represent the city in the model.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 63
You work as an ML engineer at a social media company, and you are developing a visual filter for users' profile photos. This requires you to train an ML model to detect bounding boxes around human faces. You want to use this filter in your company's iOS-based mobile phone application. You want to minimize code development and want the model to be optimized for inference on mobile phones. What should you do?
A. Train a model using AutoML Vision and use the "export for TensorFlow.js" option.
B. Train a model using AutoML Vision and use the "export for Coral" option.
C. Train a model using AutoML Vision and use the "export for Core ML" option.
D. Train a custom TensorFlow model and convert it to TensorFlow Lite (TFLite).
Answer: (SHOW ANSWER)

NEW QUESTION: 64
You work for a credit card company and have been asked to create a custom fraud detection model based on historical data using AutoML Tables. You need to prioritize detection of fraudulent transactions while minimizing false positives. Which optimization objective should you use when training the model?
A. An optimization objective that minimizes Log loss
B. An optimization objective that maximizes the Precision at a Recall value of 0.50
C. An optimization objective that maximizes the area under the precision-recall curve (AUC PR) value
D. An optimization objective that maximizes the area under the receiver operating characteristic curve (AUC ROC) value
Answer: C (LEAVE A REPLY)
https://stats.stackexchange.com/questions/262616/roc-vs-precision-recall-curves-on-imbalanced-dataset
https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc
https://icaiit.org/proceedings/6th_ICAIIT/1_3Fayzrakhmanov.pdf The problem of fraudulent transactions detection, which is an imbalanced classification problem (most transactions are not fraudulent), you want to maximize both precision and recall; so the area under the PR curve. As a matter of fact, the question asks you to focus on detecting fraudulent transactions (maximize true positive rate, a.k.a. Recall) while minimizing false positives (a.k.a. maximizing Precision). Another way to see it is this: for imbalanced problems like this one you'll get a lot of true negatives even from a bad model (it's easy to guess a transaction as "non-fraudulent" because most of them are!), and with high TN the ROC curve goes high fast, which would be misleading. So you wanna avoid dealing with true negatives in your evaluation, which is precisely what the PR curve allows you to do.

NEW QUESTION: 65
You work for a retailer that sells clothes to customers around the world. You have been tasked with ensuring that ML models are built in a secure manner. Specifically, you need to protect

sensitive customer data that might be used in the models. You have identified four fields containing sensitive data that are being used by your data science team: AGE, IS_EXISTING_CUSTOMER, LATITUDE_LONGITUDE, and SHIRT_SIZE. What should you do with the data before it is made available to the data science team for training purposes?
A. Remove all sensitive data fields, and ask the data science team to build their models using non-sensitive data.
B. Use principal component analysis (PCA) to reduce the four sensitive fields to one PCA vector.
C. Coarsen the data by putting AGE into quantiles and rounding LATITUDE_LONGTTUDE into single precision. The other two fields are already as coarse as possible.
D. Tokenize all of the fields using hashed dummy values to replace the real values.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 66
You need to train a computer vision model that predicts the type of government ID present in a given image using a GPU-powered virtual machine on Compute Engine. You use the following parameters:
* Optimizer: SGD
* Image shape = 224x224
* Batch size = 64
* Epochs = 10
* Verbose = 2
During training you encounter the following error: ResourceExhaustedError: out of Memory (oom) when allocating tensor. What should you do?
A. Change the optimizer
B. Reduce the batch size
C. Change the learning rate
D. Reduce the image shape
Answer: B (LEAVE A REPLY)
Reference:
https://stackoverflow.com/questions/59394947/how-to-fix-resourceexhaustederror-oom-when-allocating-tensor/59395251#:~:text=OOM%20stands%20for%20%22out%20of,in%20your%20Dense%20%2C%20Conv2D%20layers

NEW QUESTION: 67
You work for a large hotel chain and have been asked to assist the marketing team in gathering predictions for a targeted marketing strategy. You need to make predictions about user lifetime value (LTV) over the next 30 days so that marketing can be adjusted accordingly. The customer dataset is in BigQuery, and you are preparing the tabular data for training with AutoML Tables. This data has a time signal that is spread across multiple columns. How should you ensure that AutoML fits the best model to your data?

A. Manually combine all columns that contain a time signal into an array Allow AutoML to interpret this array appropriately Choose an automatic data split across the training, validation, and testing sets
B. Submit the data for training without performing any manual transformations Allow AutoML to handle the appropriate transformations Choose an automatic data split across the training, validation, and testing sets
C. Submit the data for training without performing any manual transformations, and indicate an appropriate column as the Time column Allow AutoML to split your data based on the time signal provided, and reserve the more recent data for the validation and testing sets
D. Submit the data for training without performing any manual transformations Use the columns that have a time signal to manually split your data Ensure that the data in your validation set is from 30 days after the data in your training set and that the data in your testing set is from 30 days after your validation set
Answer: D (LEAVE A REPLY)
https://cloud.google.com/automl-tables/docs/data-best-practices#time

NEW QUESTION: 68
As the lead ML Engineer for your company, you are responsible for building ML models to digitize scanned customer forms. You have developed a TensorFlow model that converts the scanned images into text and stores them in Cloud Storage. You need to use your ML model on the aggregated data collected at the end of each day with minimal manual intervention. What should you do?
A. Use the batch prediction functionality of AI Platform
B. Create a serving pipeline in Compute Engine for prediction
C. Use Cloud Functions for prediction each time a new data point is ingested
D. Deploy the model on AI Platform and create a version of it for online inference.
Answer: (SHOW ANSWER)
https://cloud.google.com/ai-platform/prediction/docs/batch-predict

NEW QUESTION: 69
You are profiling the performance of your TensorFlow model training time and notice a performance issue caused by inefficiencies in the input data pipeline for a single 5 terabyte CSV file dataset on Cloud Storage. You need to optimize the input pipeline performance. Which action should you try first to increase the efficiency of your pipeline?
A. Randomly select a 10 gigabyte subset of the data to train your model.
B. Set the reshuffle_each_iteration parameter to true in the tf.data.Dataset.shuffle method.
C. Preprocess the input CSV file into a TFRecord file.
D. Split into multiple CSV files and use a parallel interleave transformation.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 70

You are developing an ML model using a dataset with categorical input variables. You have randomly split half of the data into training and test sets. After applying one-hot encoding on the categorical variables in the training set, you discover that one categorical variable is missing from the test set. What should you do?

A. Randomly redistribute the data, with 70% for the training set and 30% for the test set

B. Use sparse representation in the test set

C. Apply one-hot encoding on the categorical variables in the test data.

D. Collect more data representing all categories

Answer: (SHOW ANSWER)

This approach ensures that the model is able to accurately interpret the categorical data in the test set. As the training set already contains one-hot encoded data, it is important to apply the same encoding to the test set so the model can interpret the data accurately. Reference: https://machinelearningmastery.com/how-to-one-hot-encode-sequence-data-in-python/https://machinelearningmastery.com/how-to-use-one-hot-encoding-for-categorical-data/. When working with categorical input variables, it's important to ensure that the same preprocessing steps are applied to both the training and test sets. One-hot encoding is a common method used to convert categorical variables into numerical values, which can then be used as inputs to machine learning models. By applying one-hot encoding to the test set, you will ensure that the test data has the same format as the training data and that the model can make accurate predictions.

NEW QUESTION: 71

Your company manages an application that aggregates news articles from many different online sources and sends them to users. You need to build a recommendation model that will suggest articles to readers that are similar to the articles they are currently reading. Which approach should you use?

A. Create a collaborative filtering system that recommends articles to a user based on the user's past behavior.

B. Manually label a few hundred articles, and then train an SVM classifier based on the manually classified articles that categorizes additional articles into their respective categories.

C. Encode all articles into vectors using word2vec, and build a model that returns articles based on vector similarity.

D. Build a logistic regression model for each user that predicts whether an article should be recommended to a user.

Answer: A (LEAVE A REPLY)

Valid Professional-Machine-Learning-Engineer Dumps shared by Actual4test.com for Helping Passing Professional-Machine-Learning-Engineer Exam! Actual4test.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the Actual4test.com

Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com Professional-Machine-Learning-Engineer dumps with Test Engine here: https://www.actual4test.com/Professional-Machine-Learning-Engineer_examcollection.html (**300** Q&As Dumps, **30%OFF** **Special Discount:** **Freepdfdumps**)