Question #: 77

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You need to execute a batch prediction on 100 million records in a BigQuery table with a custom TensorFlow DNN regressor model, and then store the predicted results in a BigQuery table. You want to minimize the effort required to build this inference pipeline. What should you do?

A. Import the TensorFlow model with BigQuery ML, and run the ml.predict function.

B. Use the TensorFlow BigQuery reader to load the data, and use the BigQuery API to write the results to BigQuery.

C. Create a Dataflow pipeline to convert the data in BigQuery to TFRecords. Run a batch inference on Vertex AI Prediction, and write the results to BigQuery.

D. Load the TensorFlow SavedModel in a Dataflow pipeline. Use the BigQuery I/O connector with a custom function to perform the inference within the pipeline, and write the results to BigQuery.

Show Suggested Answer

Question #: 78

Topic #: 1

[All Professional Machine Learning Engineer Questions]

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 each categorical value into an integer value.

B. Convert the categorical string data to one-hot hash buckets.

C. Map the categorical variables into a vector of boolean values.

D. Convert each categorical value into a run-length encoded string.

Show Suggested Answer

Question #: 79

Topic #: 1

[All Professional Machine Learning Engineer Questions]

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. Create a hot-encoding of words, and feed the encodings into your model.

B. Identify word embeddings from a pre-trained model, and use the embeddings in your model.

C. Sort the words by frequency of occurrence, and use the frequencies as the encodings in your model.

D. Assign a numerical value to each word from 1 to 100,000 and feed the values as inputs in your model.

Show Suggested Answer

Question #: 81
Topic #: 1
[All Professional Machine Learning Engineer Questions]

Your data science team has requested a system that supports scheduled model retraining, Docker containers, and a service that supports autoscaling and monitoring for online prediction requests. Which platform components should you choose for this system?

A. Vertex AI Pipelines and App Engine

B. Vertex AI Pipelines, Vertex AI Prediction, and Vertex AI Model Monitoring

C. Cloud Composer, BigQuery ML, and Vertex AI Prediction

D. Cloud Composer, Vertex AI Training with custom containers, and App Engine

Show Suggested Answer

Question #: 82
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. Preprocess the input CSV file into a TFRecord file.

B. Randomly select a 10 gigabyte subset of the data to train your model.

C. Split into multiple CSV files and use a parallel interleave transformation.

D. Set the reshuffle_each_iteration parameter to true in the tf.data.Dataset.shuffle method.

Show Suggested Answer

Question #: 83
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. 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.

B. 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.

C. 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.

D. 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.

Show Suggested Answer

Question #: 84
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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

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

D. 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.

Show Suggested Answer

Question #: 85
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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

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

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

Show Suggested Answer

Question #: 86
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are a lead ML engineer at a retail company. You want to track and manage ML metadata in a centralized way so that your team can have reproducible experiments by generating artifacts. Which management solution should you recommend to your team?

A. Store your tf.logging data in BigQuery.

B. Manage all relational entities in the Hive Metastore.

C. Store all ML metadata in Google Cloud's operations suite.

D. Manage your ML workflows with Vertex ML Metadata.

Show Suggested Answer

Question #: 87
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You have been given a dataset with sales predictions based on your company's marketing activities. The data is structured and stored in BigQuery, and has been carefully managed by a team of data analysts. You need to prepare a report providing insights into the predictive capabilities of the data. You were asked to run several ML models with different levels of sophistication, including simple models and multilayered neural networks. You only have a few hours to gather the results of your experiments. Which Google Cloud tools should you use to complete this task in the most efficient and self-serviced way?

A. Use BigQuery ML to run several regression models, and analyze their performance.

B. Read the data from BigQuery using Dataproc, and run several models using SparkML.

C. Use Vertex AI Workbench user-managed notebooks with scikit-learn code for a variety of ML algorithms and performance metrics.

D. Train a custom TensorFlow model with Vertex AI, reading the data from BigQuery featuring a variety of ML algorithms.

Show Suggested Answer

Question #: 88
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are an ML engineer at a bank. You have developed a binary classification model using AutoML Tables to predict whether a customer will make loan payments on time. The output is used to approve or reject loan requests. One customer's loan request has been rejected by your model, and the bank's risks department is asking you to provide the reasons that contributed to the model's decision. What should you do?

A. Use local feature importance from the predictions.

B. Use the correlation with target values in the data summary page.

C. Use the feature importance percentages in the model evaluation page.

D. Vary features independently to identify the threshold per feature that changes the classification.

Show Suggested Answer

Question #: 89
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work for a magazine distributor and need to build a model that predicts which customers will renew their subscriptions for the upcoming year. Using your company's historical data as your training set, you created a TensorFlow model and deployed it to AI Platform. You need to determine which customer attribute has the most predictive power for each prediction served by the model. What should you do?

A. Use AI Platform notebooks to perform a Lasso regression analysis on your model, which will eliminate features that do not provide a strong signal.

B. Stream prediction results to BigQuery. Use BigQuery's CORR(X1, X2) function to calculate the Pearson correlation coefficient between each feature and the target variable.

C. Use the AI Explanations feature on AI Platform. Submit each prediction request with the 'explain' keyword to retrieve feature attributions using the sampled Shapley method.

D. Use the What-If tool in Google Cloud to determine how your model will perform when individual features are excluded. Rank the feature importance in order of those that caused the most significant performance drop when removed from the model.

Show Suggested Answer

Question #: 90
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are working on a binary classification ML algorithm that detects whether an image of a classified scanned document contains a company's logo. In the dataset, 96% of examples don't have the logo, so the dataset is very skewed. Which metrics would give you the most confidence in your model?

A. F-score where recall is weighed more than precision

B. RMSE

C. F1 score

D. F-score where precision is weighed more than recall

Show Suggested Answer

Question #: 90
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are working on a binary classification ML algorithm that detects whether an image of a classified scanned document contains a company's logo. In the dataset, 96% of examples don't have the logo, so the dataset is very skewed. Which metrics would give you the most confidence in your model?

A. F-score where recall is weighed more than precision

B. RMSE

C. F1 score

D. F-score where precision is weighed more than recall

Show Suggested Answer

Question #: 91
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work on the data science team for a multinational beverage company. You need to develop an ML model to predict the company's profitability for a new line of naturally flavored bottled waters in different locations. You are provided with historical data that includes product types, product sales volumes, expenses, and profits for all regions. What should you use as the input and output for your model?

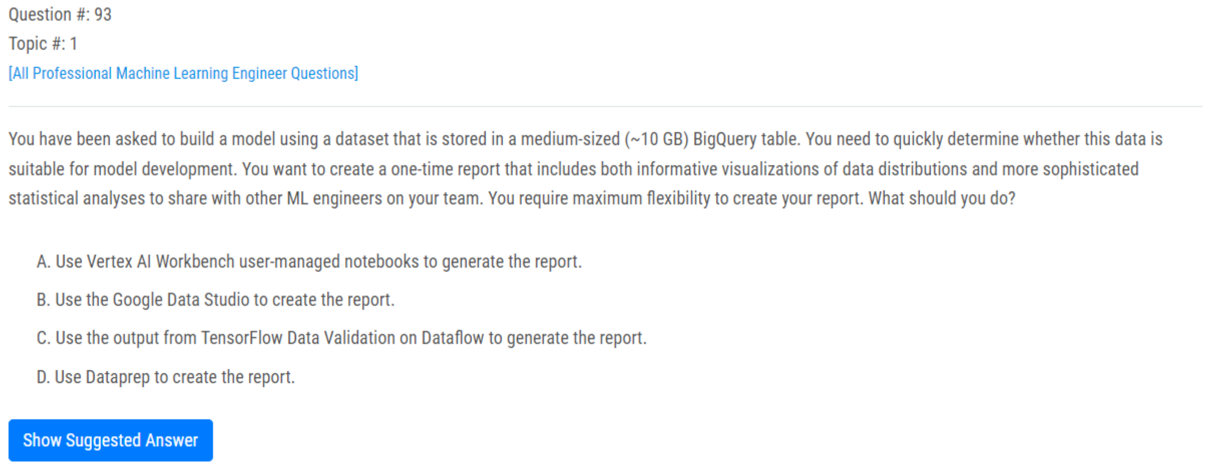
A. Use latitude, longitude, and product type as features. Use profit as model output.

B. Use latitude, longitude, and product type as features. Use revenue and expenses as model outputs.

C. Use product type and the feature cross of latitude with longitude, followed by binning, as features. Use profit as model output.

D. Use product type and the feature cross of latitude with longitude, followed by binning, as features. Use revenue and expenses as model outputs.

Show Suggested Answer



Question #: 93

Topic #: 1

[All Professional Machine Learning Engineer Questions]

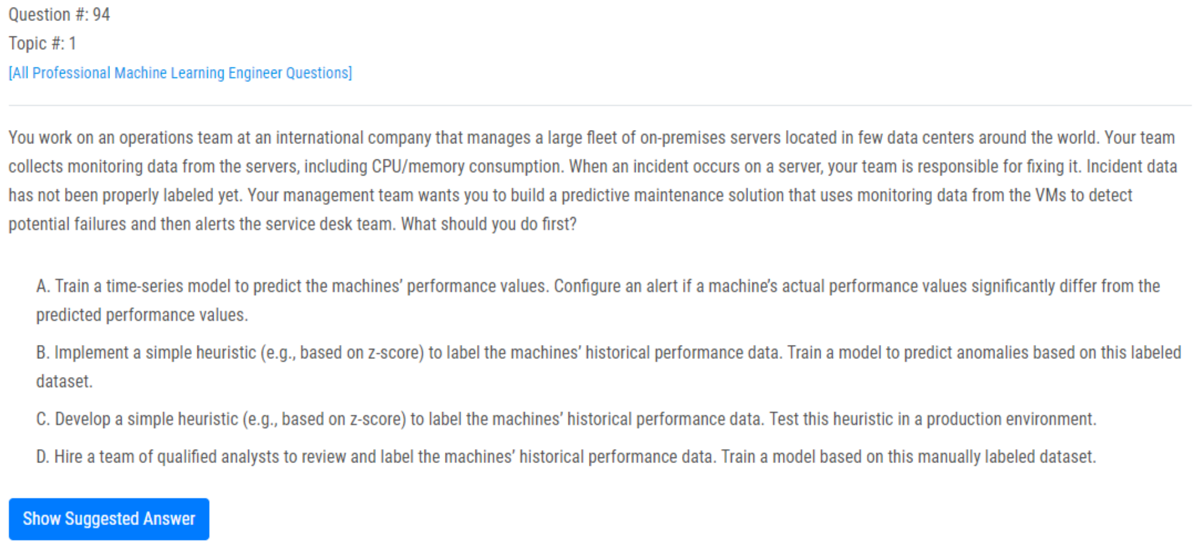
You have been asked to build a model using a dataset that is stored in a medium-sized (~10 GB) BigQuery table. You need to quickly determine whether this data is suitable for model development. You want to create a one-time report that includes both informative visualizations of data distributions and 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 Vertex AI Workbench user-managed notebooks to generate the report.

B. Use the Google Data Studio to create the report.

C. Use the output from TensorFlow Data Validation on Dataflow to generate the report.

D. Use Dataprep to create the report.



Question #: 94

Topic #: 1

[All Professional Machine Learning Engineer Questions]

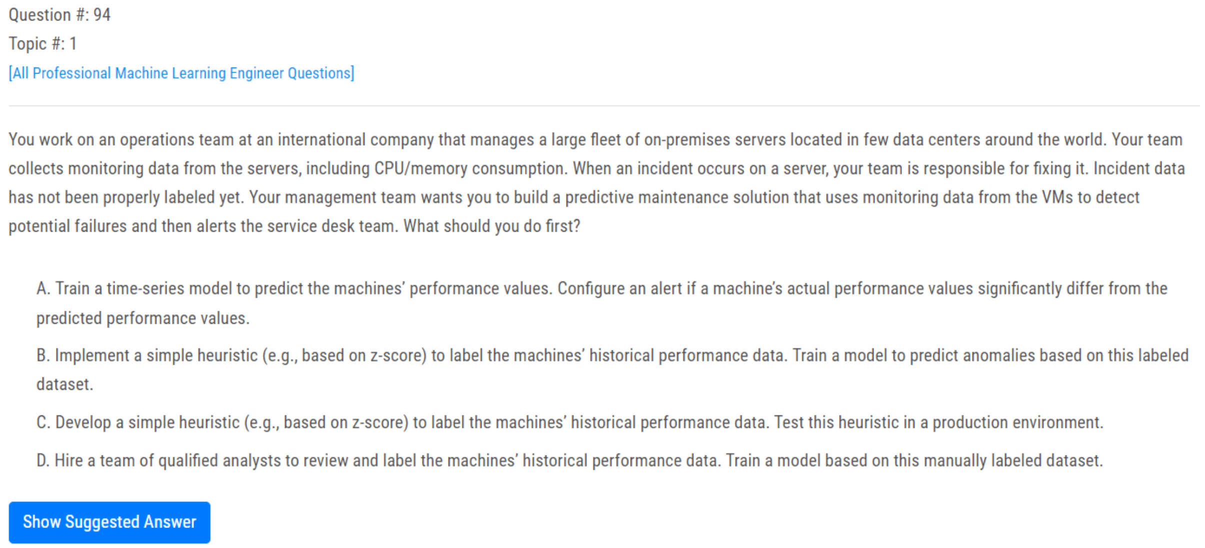
You work on an operations team at an international company that manages a large fleet of on-premises servers located in few data centers around the world. Your team collects monitoring data from the servers, including CPU/memory consumption. When an incident occurs on a server, your team is responsible for fixing it. Incident data has not been properly labeled yet. Your management team wants you to build a predictive maintenance solution that uses monitoring data from the VMs to detect potential failures and then alerts the service desk team. What should you do first?

A. Train a time-series model to predict the machines' performance values. Configure an alert if a machine's actual performance values significantly differ from the predicted performance values.

B. Implement a simple heuristic (e.g., based on z-score) to label the machines' historical performance data. Train a model to predict anomalies based on this labeled dataset.

C. Develop a simple heuristic (e.g., based on z-score) to label the machines' historical performance data. Test this heuristic in a production environment.

D. Hire a team of qualified analysts to review and label the machines' historical performance data. Train a model based on this manually labeled dataset.



Question #: 94

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work on an operations team at an international company that manages a large fleet of on-premises servers located in few data centers around the world. Your team collects monitoring data from the servers, including CPU/memory consumption. When an incident occurs on a server, your team is responsible for fixing it. Incident data has not been properly labeled yet. Your management team wants you to build a predictive maintenance solution that uses monitoring data from the VMs to detect potential failures and then alerts the service desk team. What should you do first?

A. Train a time-series model to predict the machines' performance values. Configure an alert if a machine's actual performance values significantly differ from the predicted performance values.

B. Implement a simple heuristic (e.g., based on z-score) to label the machines' historical performance data. Train a model to predict anomalies based on this labeled dataset.

C. Develop a simple heuristic (e.g., based on z-score) to label the machines' historical performance data. Test this heuristic in a production environment.

D. Hire a team of qualified analysts to review and label the machines' historical performance data. Train a model based on this manually labeled dataset.

Question #: 95
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are developing an ML model that uses sliced frames from video feed and creates bounding boxes around specific objects. You want to automate the following steps in your training pipeline: ingestion and preprocessing of data in Cloud Storage, followed by training and hyperparameter tuning of the object model using Vertex AI jobs, and finally deploying the model to an endpoint. You want to orchestrate the entire pipeline with minimal cluster management. What approach should you use?

A. Use Kubeflow Pipelines on Google Kubernetes Engine.

B. Use Vertex AI Pipelines with TensorFlow Extended (TFX) SDK.

C. Use Vertex AI Pipelines with Kubeflow Pipelines SDK.

D. Use Cloud Composer for the orchestration.

Show Suggested Answer

Question #: 96
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are training an object detection machine learning model on a dataset that consists of three million X-ray images, each roughly 2 GB in size. You are using Vertex AI Training to run a custom training application on a Compute Engine instance with 32-cores, 128 GB of RAM, and 1 NVIDIA P100 GPU. You notice that model training is taking a very long time. You want to decrease training time without sacrificing model performance. What should you do?

A. Increase the instance memory to 512 GB and increase the batch size.

B. Replace the NVIDIA P100 GPU with a v3-32 TPU in the training job.

C. Enable early stopping in your Vertex AI Training job.

D. Use the tf.distribute.Strategy API and run a distributed training job.

Show Suggested Answer

Question #: 97
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are a data scientist at an industrial equipment manufacturing company. You are developing a regression model to estimate the power consumption in the company's manufacturing plants based on sensor data collected from all of the plants. The sensors collect tens of millions of records every day. You need to schedule daily training runs for your model that use all the data collected up to the current date. You want your model to scale smoothly and require minimal development work. What should you do?

A. Train a regression model using AutoML Tables.

B. Develop a custom TensorFlow regression model, and optimize it using Vertex AI Training.

C. Develop a custom scikit-learn regression model, and optimize it using Vertex AI Training.

D. Develop a regression model using BigQuery ML.

Show Suggested Answer

Question #: 98
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. Migrate your model to TensorFlow, and train it using Vertex AI Training.

B. Train your model in a distributed mode using multiple Compute Engine VMs.

C. Train your model with DLVM images on Vertex AI, and ensure that your code utilizes NumPy and SciPy internal methods whenever possible.

D. Train your model using Vertex AI Training with GPUs.

Show Suggested Answer

Question #: 99
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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 pipeline run in Kubeflow under an experiment for each season per year. Compare the results across the experiments in the Kubeflow UI.

D. 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.

Show Suggested Answer

Actual exam question from Google's Professional Machine Learning Engineer
Question #: 100
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are an ML engineer at a manufacturing company. You need to build a model that identifies defects in products based on images of the product taken at the end of the assembly line. You want your model to preprocess the images with lower computation to quickly extract features of defects in products. Which approach should you use to build the model?

A. Reinforcement learning

B. Recommender system

C. Recurrent Neural Networks (RNN)

D. Convolutional Neural Networks (CNN)

Show Suggested Answer

Question #: 101
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are developing an ML model intended to classify whether X-ray images indicate bone fracture risk. You have trained a ResNet architecture on Vertex AI using a TPU as an accelerator, however you are unsatisfied with the training time and 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 model's 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. Reduce the dimensions of the images used in the model.

D. Configure your model to use bfloat16 instead of float32.

Show Suggested Answer

Question #: 102
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You have successfully deployed to production a large and complex TensorFlow model trained on tabular data. 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. Implement continuous retraining of the model daily using Vertex AI Pipelines.

B. Add a model monitoring job where 10% of incoming predictions are sampled 24 hours.

C. Add a model monitoring job where 90% of incoming predictions are sampled 24 hours.

D. Add a model monitoring job where 10% of incoming predictions are sampled every hour.

Show Suggested Answer

Question #: 103
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You recently developed a deep learning model using Keras, and now you are experimenting with different training strategies. First, you trained the model using a single GPU, but the training process was too slow. Next, you distributed the training across 4 GPUs using tf.distribute.MirroredStrategy (with no other changes), but you did not observe a decrease in training time. What should you do?

A. Distribute the dataset with tf.distribute.Strategy.experimental_distribute_dataset

B. Create a custom training loop.

C. Use a TPU with tf.distribute.TPUStrategy.

D. Increase the batch size.

Show Suggested Answer

Question #: 104
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work for a gaming company that has millions of customers around the world. All games offer a chat feature that allows players to communicate with each other in real time. Messages can be typed in more than 20 languages and are translated in real time using the Cloud Translation API. You have been asked to build an ML system to moderate the chat in real time while assuring that the performance is uniform across the various languages and without changing the serving infrastructure.

You trained your first model using an in-house word2vec model for embedding the chat messages translated by the Cloud Translation API. However, the model has significant differences in performance across the different languages. How should you improve it?

A. Add a regularization term such as the Min-Diff algorithm to the loss function.

B. Train a classifier using the chat messages in their original language.

C. Replace the in-house word2vec with GPT-3 or T5.

D. Remove moderation for languages for which the false positive rate is too high.

Show Suggested Answer

Question #: 105
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work for a gaming company that develops massively multiplayer online (MMO) games. You built a TensorFlow model that predicts whether players will make in-app purchases of more than $10 in the next two weeks. The model's predictions will be used to adapt each user's game experience. User data is stored in BigQuery. How should you serve your model while optimizing cost, user experience, and ease of management?

A. Import the model into BigQuery ML. Make predictions using batch reading data from BigQuery, and push the data to Cloud SQL

B. Deploy the model to Vertex AI Prediction. Make predictions using batch reading data from Cloud Bigtable, and push the data to Cloud SQL.

C. Embed the model in the mobile application. Make predictions after every in-app purchase event is published in Pub/Sub, and push the data to Cloud SQL.

D. Embed the model in the streaming Dataflow pipeline. Make predictions after every in-app purchase event is published in Pub/Sub, and push the data to Cloud SQL.

Show Suggested Answer

Question #: 107
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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 mode?

A. Data Loss Prevention API

B. Federated learning

C. MD5 to encrypt data

D. Differential privacy

Show Suggested Answer

Question #: 108
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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.

Show Suggested Answer

Question #: 109
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. Decrease the size of the training batch.

B. Decrease the learning rate hyperparameter.

C. Increase the learning rate hyperparameter.

D. Increase the size of the training batch.

Show Suggested Answer

Question #: 110
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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 Edge mobile-high-accuracy-1 model

B. AutoML Vision Edge mobile-low-latency-1 model

C. AutoML Vision model

D. AutoML Vision Edge mobile-versatile-1 model

Show Suggested Answer

Question #: 112

Topic #: 1

[All Professional Machine Learning Engineer Questions]

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

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

C. Extract sentiment directly from the voice recordings.

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

Show Suggested Answer

Question #: 113

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You need to analyze user activity data from your company's mobile applications. Your team will use BigQuery for data analysis, transformation, and experimentation with ML algorithms. You need to ensure real-time ingestion of the user activity data into BigQuery. What should you do?

A. Configure Pub/Sub to stream the data into BigQuery.

B. Run an Apache Spark streaming job on Dataproc to ingest the data into BigQuery.

C. Run a Dataflow streaming job to ingest the data into BigQuery.

D. Configure Pub/Sub and a Dataflow streaming job to ingest the data into BigQuery,

Show Suggested Answer

Question #: 114

Topic #: 1

[All Professional Machine Learning Engineer Questions]

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?

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

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

Show Suggested Answer

Question #: 115
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are building an ML model to predict trends in the stock market based on a wide range of factors. While exploring the data, you notice that some features have a large range. You want to ensure that the features with the largest magnitude don't overfit the model. What should you do?

A. Standardize the data by transforming it with a logarithmic function.

B. Apply a principal component analysis (PCA) to minimize the effect of any particular feature.

C. Use a binning strategy to replace the magnitude of each feature with the appropriate bin number.

D. Normalize the data by scaling it to have values between 0 and 1.

Show Suggested Answer

Question #: 116
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work for a biotech startup that is experimenting with deep learning ML models based on properties of biological organisms. Your team frequently works on early-stage experiments with new architectures of ML models, and writes custom TensorFlow ops in C++. You train your models on large datasets and large batch sizes. Your typical batch size has 1024 examples, and each example is about 1 MB in size. The average size of a network with all weights and embeddings is 20 GB. What hardware should you choose for your models?

A. A cluster with 2 n1-highcpu-64 machines, each with 8 NVIDIA Tesla V100 GPUs (128 GB GPU memory in total), and a n1-highcpu-64 machine with 64 vCPUs and 58 GB RAM

B. A cluster with 2 a2-megagpu-16g machines, each with 16 NVIDIA Tesla A100 GPUs (640 GB GPU memory in total), 96 vCPUs, and 1.4 TB RAM

C. A cluster with an n1-highcpu-64 machine with a v2-8 TPU and 64 GB RAM

D. A cluster with 4 n1-highcpu-96 machines, each with 96 vCPUs and 86 GB RAM

Show Suggested Answer

Question #: 117
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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?

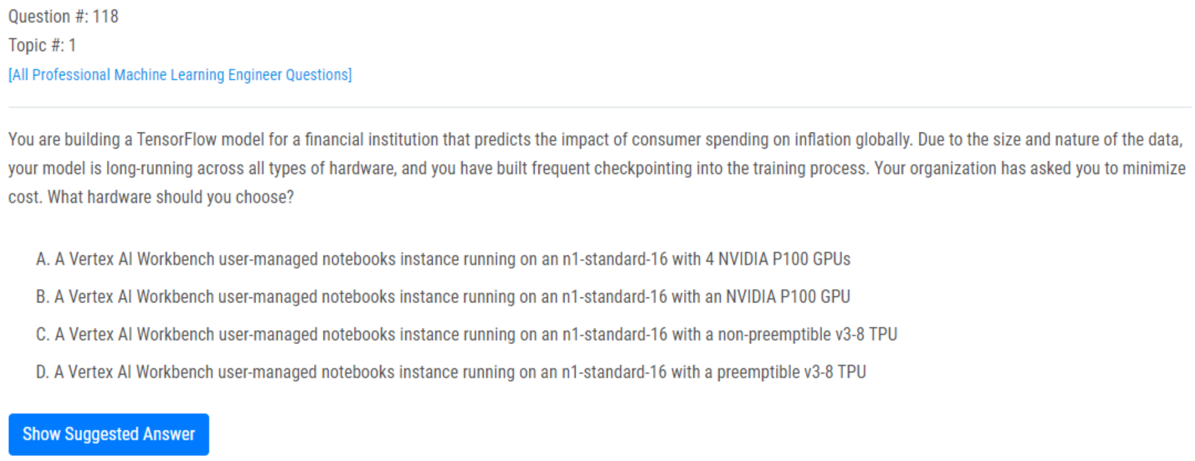
A. Use a clustering algorithm to group popular items together. Give the list to the logistics team so they can increase inventory of the popular items.

B. 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.

C. 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.

D. Use a classification model to classify inventory levels as UNDER_STOCKED, OVER_STOCKED, and CORRECTLY_STOCKEGive the report to the logistics team each month so they can fine-tune inventory levels.

Show Suggested Answer



Question #: 118

Topic #: 1

[All Professional Machine Learning Engineer Questions]

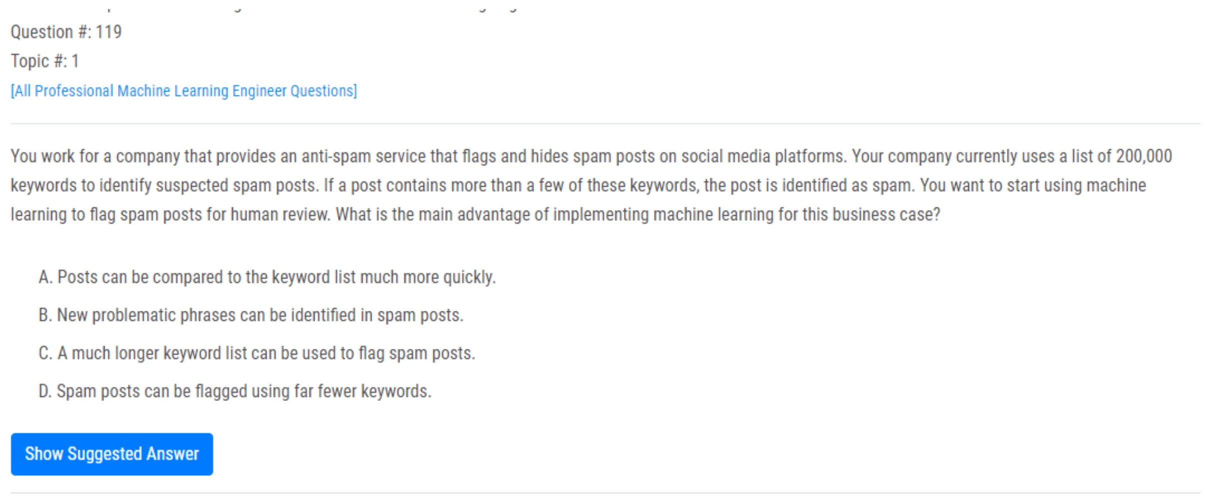
You are building a TensorFlow model for a financial institution that predicts the impact of consumer spending on inflation globally. Due to the size and nature of the data, your model is long-running across all types of hardware, and you have built frequent checkpointing into the training process. Your 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 4 NVIDIA P100 GPUs

B. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with an NVIDIA P100 GPU

C. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with a non-preemptible v3-8 TPU

D. A Vertex AI Workbench user-managed notebooks instance running on an n1-standard-16 with a preemptible v3-8 TPU



Question #: 119

Topic #: 1

[All Professional Machine Learning Engineer Questions]

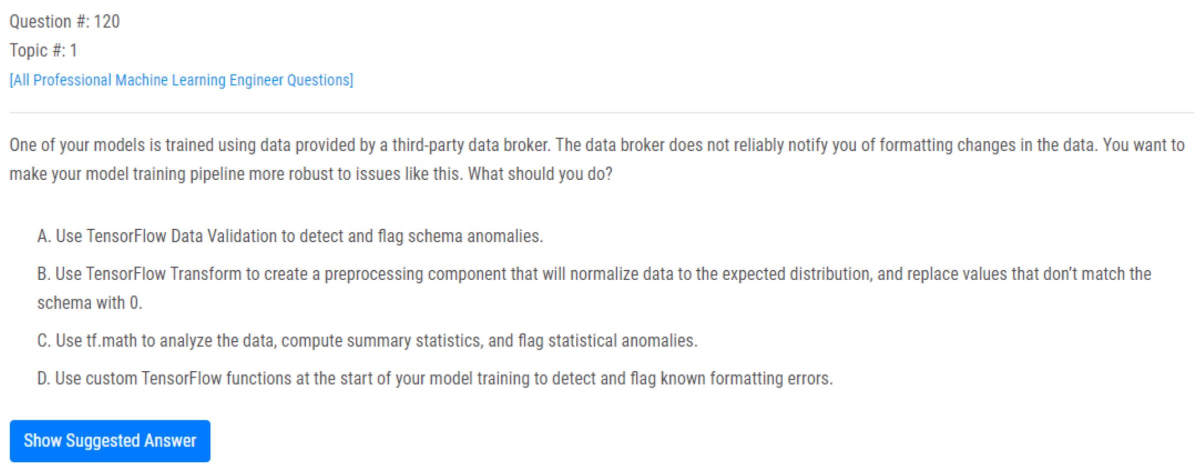
You work for a company that provides an anti-spam service that flags and hides spam posts on social media platforms. Your company currently uses a list of 200,000 keywords to identify suspected spam posts. If a post contains more than a few of these keywords, the post is identified as spam. You want to start using machine learning to flag spam posts for human review. What is the main advantage of implementing machine learning for this business case?

A. Posts can be compared to the keyword list much more quickly.

B. New problematic phrases can be identified in spam posts.

C. A much longer keyword list can be used to flag spam posts.

D. Spam posts can be flagged using far fewer keywords.



Question #: 120

Topic #: 1

[All Professional Machine Learning Engineer Questions]

One of your models is trained using data provided by a third-party data broker. The data broker does not reliably notify you of formatting changes in the data. You want to make your model training pipeline more robust to issues like this. What should you do?

A. Use TensorFlow Data Validation to detect and flag schema anomalies.

B. Use TensorFlow Transform to create a preprocessing component that will normalize data to the expected distribution, and replace values that don't match the schema with 0.

C. Use tf.math to analyze the data, compute summary statistics, and flag statistical anomalies.

D. Use custom TensorFlow functions at the start of your model training to detect and flag known formatting errors.

Question #: 121
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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 data. How should you build the recommendation system for the first version of the product?

A. 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.

B. 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.

C. 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.

D. 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.

Show Suggested Answer

Question #: 122
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. The model is overfitting in areas with less traffic and underfitting in areas with more traffic.

B. AUC is not the correct metric to evaluate this classification model.

C. Too much data representing congested areas was used for model training.

D. Gradients become small and vanish while backpropagating from the output to input nodes.

Show Suggested Answer

Question #: 123
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are developing an ML model to predict house prices. While preparing the data, you discover that an important predictor variable, distance from the closest school, is often missing and does not have high variance. Every instance (row) in your data is important. How should you handle the missing data?

A. Delete the rows that have missing values.

B. Apply feature crossing with another column that does not have missing values.

C. Predict the missing values using linear regression.

D. Replace the missing values with zeros.

Show Suggested Answer

Question #: 124
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are an ML engineer responsible for designing and implementing training pipelines for ML models. You need to create an end-to-end training pipeline for a TensorFlow model. The TensorFlow model will be trained on several terabytes of structured data. You need the pipeline to include data quality checks before training and model quality checks after training but prior to deployment. You want to minimize development time and the need for infrastructure maintenance. How should you build and orchestrate your training pipeline?

A. Create the pipeline using Kubeflow Pipelines domain-specific language (DSL) and predefined Google Cloud components. Orchestrate the pipeline using Vertex AI Pipelines.

B. Create the pipeline using TensorFlow Extended (TFX) and standard TFX components. Orchestrate the pipeline using Vertex AI Pipelines.

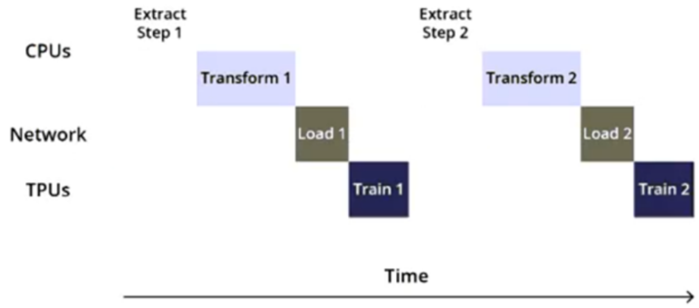
C. Create the pipeline using Kubeflow Pipelines domain-specific language (DSL) and predefined Google Cloud components. Orchestrate the pipeline using Kubeflow Pipelines deployed on Google Kubernetes Engine.

D. Create the pipeline using TensorFlow Extended (TFX) and standard TFX components. Orchestrate the pipeline using Kubeflow Pipelines deployed on Google Kubernetes Engine.

Show Suggested Answer

Question #: 126
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are training an object detection model using a Cloud TPU v2. Training time is taking longer than expected. Based on this simplified trace obtained with a Cloud TPU profile, what action should you take to decrease training time in a cost-efficient way?



A. Move from Cloud TPU v2 to Cloud TPU v3 and increase batch size.

B. Move from Cloud TPU v2 to 8 NVIDIA V100 GPUs and increase batch size.

C. Rewrite your input function to resize and reshape the input images.

D. Rewrite your input function using parallel reads, parallel processing, and prefetch.

Show Suggested Answer

Question #: 127
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. Remove the rows with missing values, and upsample your dataset by 5%.

B. Replace the missing values with the feature's mean.

C. Replace the missing values with a placeholder category indicating a missing value.

D. Move the rows with missing values to your validation dataset.

Show Suggested Answer

Question #: 128
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are an ML engineer on an agricultural research team working on a crop disease detection tool to detect leaf rust spots in images of crops to determine the presence of a disease. These spots, which can vary in shape and size, are correlated to the severity of the disease. You want to develop a solution that predicts the presence and severity of the disease with high accuracy. What should you do?

A. Create an object detection model that can localize the rust spots.

B. Develop an image segmentation ML model to locate the boundaries of the rust spots.

C. Develop a template matching algorithm using traditional computer vision libraries.

D. Develop an image classification ML model to predict the presence of the disease.

Show Suggested Answer

Question #: 129
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You have been asked to productionize a proof-of-concept ML model built using Keras. The model was trained in a Jupyter notebook on a data scientist's local machine. The notebook contains a cell that performs data validation and a cell that performs model analysis. You need to orchestrate the steps contained in the notebook and automate the execution of these steps for weekly retraining. You expect much more training data in the future. You want your solution to take advantage of managed services while minimizing cost. What should you do?

A. Move the Jupyter notebook to a Notebooks instance on the largest N2 machine type, and schedule the execution of the steps in the Notebooks instance using Cloud Scheduler.

B. Write the code as a TensorFlow Extended (TFX) pipeline orchestrated with Vertex AI Pipelines. Use standard TFX components for data validation and model analysis, and use Vertex AI Pipelines for model retraining.

C. Rewrite the steps in the Jupyter notebook as an Apache Spark job, and schedule the execution of the job on ephemeral Dataproc clusters using Cloud Scheduler.

D. Extract the steps contained in the Jupyter notebook as Python scripts, wrap each script in an Apache Airflow BashOperator, and run the resulting directed acyclic graph (DAG) in Cloud Composer.

Show Suggested Answer

Question #: 130
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are working on a system log anomaly detection model for a cybersecurity organization. You have developed the model using TensorFlow, and you plan to use it for real-time prediction. You need to create a Dataflow pipeline to ingest data via Pub/Sub and write the results to BigQuery. You want to minimize the serving latency as much as possible. What should you do?

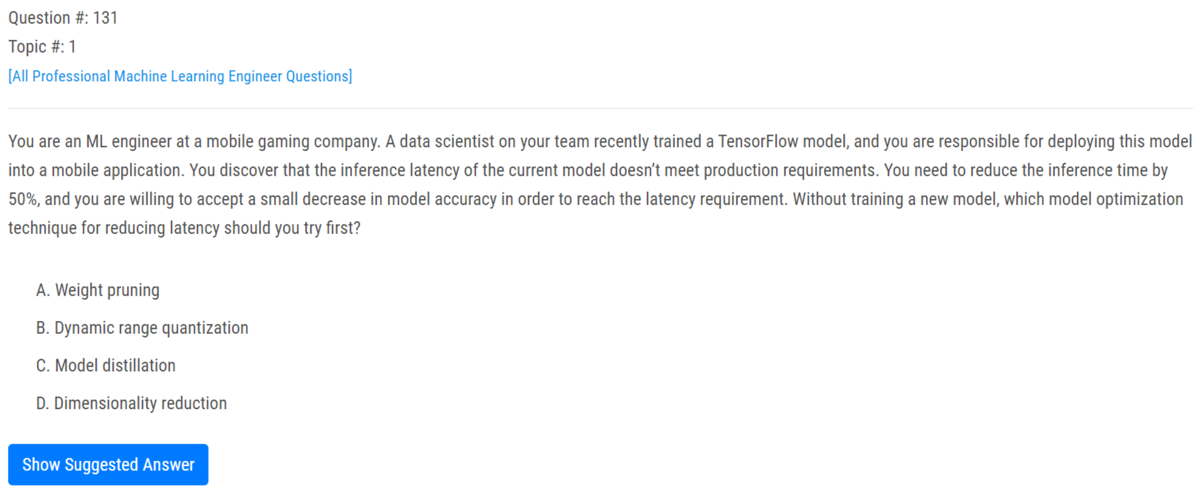
A. Containerize the model prediction logic in Cloud Run, which is invoked by Dataflow.

B. Load the model directly into the Dataflow job as a dependency, and use it for prediction.

C. Deploy the model to a Vertex AI endpoint, and invoke this endpoint in the Dataflow job.

D. Deploy the model in a TFServing container on Google Kubernetes Engine, and invoke it in the Dataflow job.

Show Suggested Answer



Question #: 131

Topic #: 1

[All Professional Machine Learning Engineer Questions]

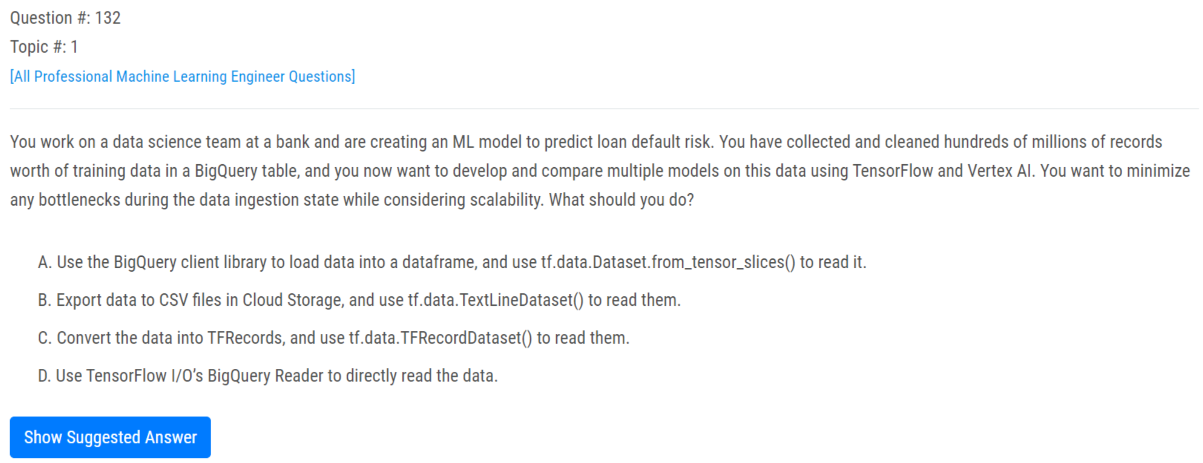
You are an ML engineer at a mobile gaming company. A data scientist on your team recently trained a TensorFlow model, and you are responsible for deploying this model into a mobile application. You discover that the inference latency of the current model doesn't meet production requirements. You need to reduce the inference time by 50%, and you are willing to accept a small decrease in model accuracy in order to reach the latency requirement. Without training a new model, which model optimization technique for reducing latency should you try first?

A. Weight pruning

B. Dynamic range quantization

C. Model distillation

D. Dimensionality reduction



Question #: 132

Topic #: 1

[All Professional Machine Learning Engineer Questions]

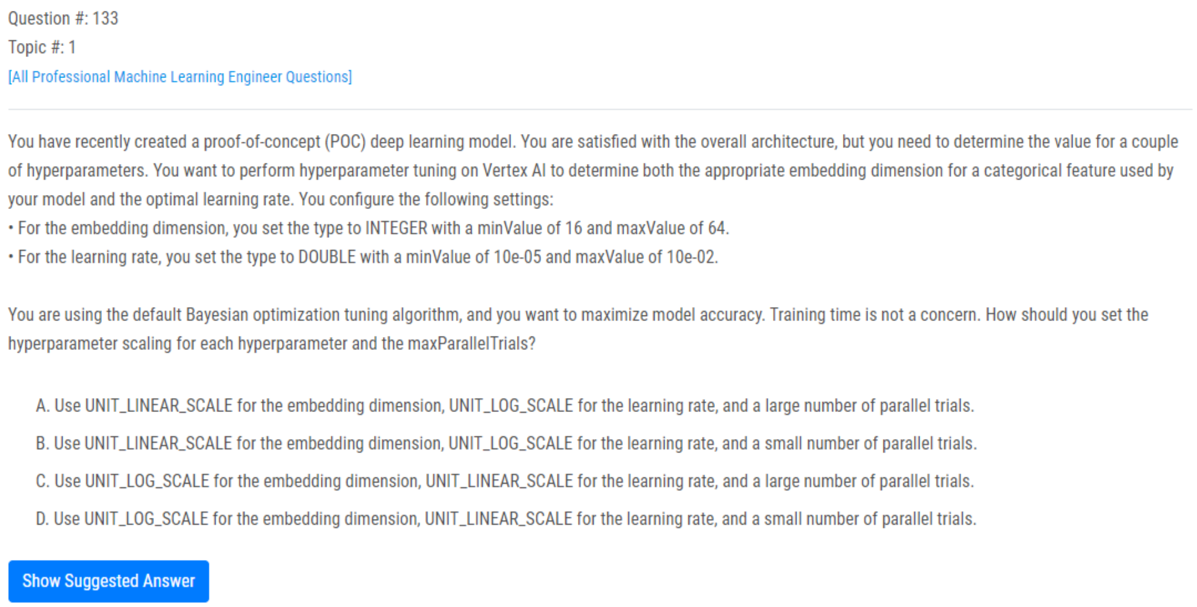
You work on a data science team at a bank and are creating an ML model to predict loan default risk. You have collected and cleaned hundreds of millions of records worth of training data in a BigQuery table, and you now want to develop and compare multiple models on this data using TensorFlow and Vertex AI. You want to minimize any bottlenecks during the data ingestion state while considering scalability. What should you do?

A. Use the BigQuery client library to load data into a dataframe, and use tf.data.Dataset.from_tensor_slices() to read it.

B. Export data to CSV files in Cloud Storage, and use tf.data.TextLineDataset() to read them.

C. Convert the data into TFRecords, and use tf.data.TFRecordDataset() to read them.

D. Use TensorFlow I/O's BigQuery Reader to directly read the data.



Question #: 133

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You have recently created a proof-of-concept (POC) deep learning model. You are satisfied with the overall architecture, but you need to determine the value for a couple of hyperparameters. You want to perform hyperparameter tuning on Vertex AI to determine both the appropriate embedding dimension for a categorical feature used by your model and the optimal learning rate. You configure the following settings:

- For the embedding dimension, you set the type to INTEGER with a minValue of 16 and maxValue of 64.
- For the learning rate, you set the type to DOUBLE with a minValue of 10e-05 and maxValue of 10e-02.

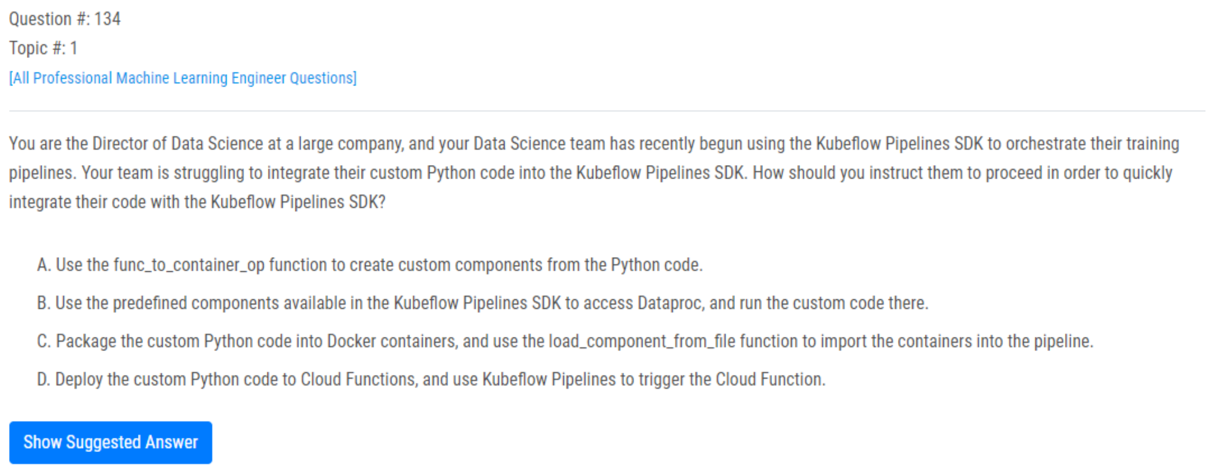
You are using the default Bayesian optimization tuning algorithm, and you want to maximize model accuracy. Training time is not a concern. How should you set the hyperparameter scaling for each hyperparameter and the maxParallelTrials?

A. Use UNIT_LINEAR_SCALE for the embedding dimension, UNIT_LOG_SCALE for the learning rate, and a large number of parallel trials.

B. Use UNIT_LINEAR_SCALE for the embedding dimension, UNIT_LOG_SCALE for the learning rate, and a small number of parallel trials.

C. Use UNIT_LOG_SCALE for the embedding dimension, UNIT_LINEAR_SCALE for the learning rate, and a large number of parallel trials.

D. Use UNIT_LOG_SCALE for the embedding dimension, UNIT_LINEAR_SCALE for the learning rate, and a small number of parallel trials.



Question #: 134
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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?

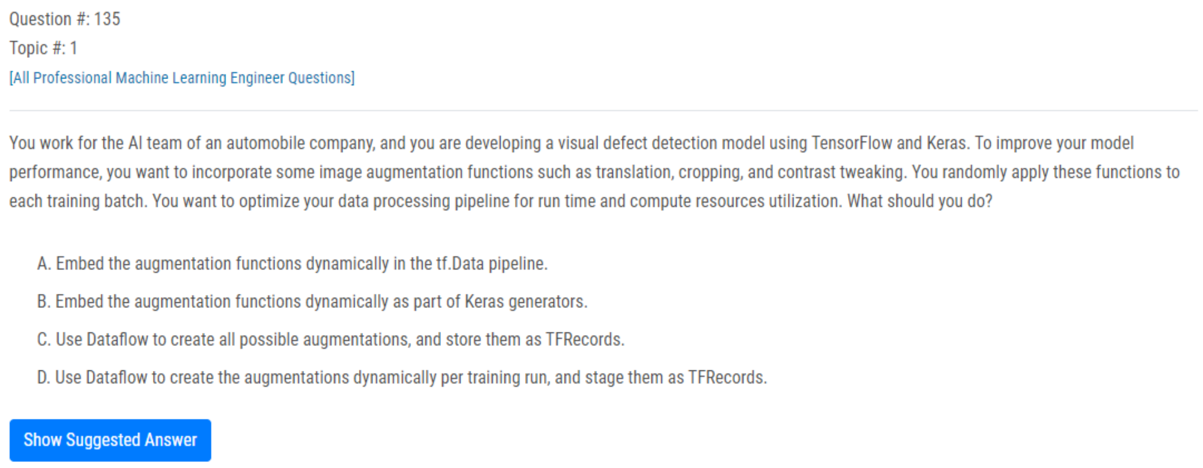
A. Use the func_to_container_op function to create custom components from the Python code.

B. Use the predefined components available in the Kubeflow Pipelines SDK to access Dataproc, and run the custom code there.

C. Package the custom Python code into Docker containers, and use the load_component_from_file function to import the containers into the pipeline.

D. Deploy the custom Python code to Cloud Functions, and use Kubeflow Pipelines to trigger the Cloud Function.

Show Suggested Answer



Question #: 135
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work for the AI team of an automobile company, and you are developing a visual defect detection model using TensorFlow and Keras. To improve your model performance, you want to incorporate some image augmentation functions such as translation, cropping, and contrast tweaking. You randomly apply these functions to each training batch. You want to optimize your data processing pipeline for run time and compute resources utilization. What should you do?

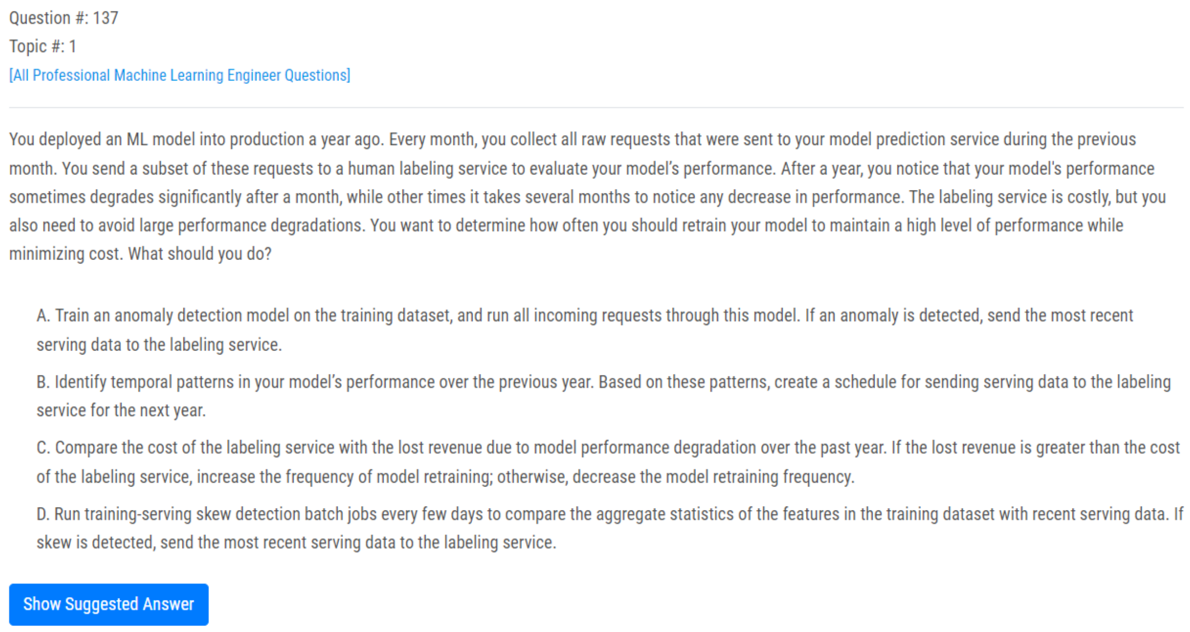
A. Embed the augmentation functions dynamically in the tf.Data pipeline.

B. Embed the augmentation functions dynamically as part of Keras generators.

C. Use Dataflow to create all possible augmentations, and store them as TFRecords.

D. Use Dataflow to create the augmentations dynamically per training run, and stage them as TFRecords.

Show Suggested Answer



Question #: 137
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You deployed an ML model into production a year ago. Every month, you collect all raw requests that were sent to your model prediction service during the previous month. You send a subset of these requests to a human labeling service to evaluate your model's performance. After a year, you notice that your model's performance sometimes degrades significantly after a month, while other times it takes several months to notice any decrease in performance. The labeling service is costly, but you also need to avoid large performance degradations. You want to determine how often you should retrain your model to maintain a high level of performance while minimizing cost. What should you do?

A. Train an anomaly detection model on the training dataset, and run all incoming requests through this model. If an anomaly is detected, send the most recent serving data to the labeling service.

B. Identify temporal patterns in your model's performance over the previous year. Based on these patterns, create a schedule for sending serving data to the labeling service for the next year.

C. Compare the cost of the labeling service with the lost revenue due to model performance degradation over the past year. If the lost revenue is greater than the cost of the labeling service, increase the frequency of model retraining; otherwise, decrease the model retraining frequency.

D. Run training-serving skew detection batch jobs every few days to compare the aggregate statistics of the features in the training dataset with recent serving data. If skew is detected, send the most recent serving data to the labeling service.

Show Suggested Answer

Question #: 138

Topic #: 1

[All Professional Machine Learning Engineer Questions]

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. Run batch inference with BigQuery ML every five minutes on each new set of tickets issued.

B. Export your model in TensorFlow format, and add a tfx_bsl.public.beam.RunInference step to the Dataflow pipeline.

C. Export your model in TensorFlow format, deploy it on Vertex AI, and query the prediction endpoint from your streaming pipeline.

D. 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.

Show Suggested Answer

Question #: 139

Topic #: 1

[All Professional Machine Learning Engineer Questions]

You work on a team in a data center that is responsible for server maintenance. Your management team wants you to build a predictive maintenance solution that uses monitoring data to detect potential server failures. Incident data has not been labeled yet. What should you do first?

A. Train a time-series model to predict the machines' performance values. Configure an alert if a machine's actual performance values significantly differ from the predicted performance values.

B. Develop a simple heuristic (e.g., based on z-score) to label the machines' historical performance data. Use this heuristic to monitor server performance in real time.

C. Develop a simple heuristic (e.g., based on z-score) to label the machines' historical performance data. Train a model to predict anomalies based on this labeled dataset.

D. Hire a team of qualified analysts to review and label the machines' historical performance data. Train a model based on this manually labeled dataset.

Show Suggested Answer

Question #: 140

Topic #: 1

[All Professional Machine Learning Engineer Questions]

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. Tokenize all of the fields using hashed dummy values to replace the real values.

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. Remove all sensitive data fields, and ask the data science team to build their models using non-sensitive data.

Show Suggested Answer

Question #: 141
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work for a magazine publisher and have been tasked with predicting whether customers will cancel their annual subscription. In your exploratory data analysis, you find that 90% of individuals renew their subscription every year, and only 10% of individuals cancel their subscription. After training a NN Classifier, your model predicts those who cancel their subscription with 99% accuracy and predicts those who renew their subscription with 82% accuracy. How should you interpret these results?

A. This is not a good result because the model should have a higher accuracy for those who renew their subscription than for those who cancel their subscription.

B. This is not a good result because the model is performing worse than predicting that people will always renew their subscription.

C. This is a good result because predicting those who cancel their subscription is more difficult, since there is less data for this group.

D. This is a good result because the accuracy across both groups is greater than 80%.

Show Suggested Answer

Question #: 143
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You have developed an ML model to detect the sentiment of users' posts on your company's social media page to identify outages or bugs. You are using Dataflow to provide real-time predictions on data ingested from Pub/Sub. You plan to have multiple training iterations for your model and keep the latest two versions live after every run. You want to split the traffic between the versions in an 80:20 ratio, with the newest model getting the majority of the traffic. You want to keep the pipeline as simple as possible, with minimal management required. What should you do?

A. Deploy the models to a Vertex AI endpoint using the traffic-split=0=80, PREVIOUS_MODEL_ID=20 configuration.

B. Wrap the models inside an App Engine application using the --splits PREVIOUS_VERSION=0.2, NEW_VERSION=0.8 configuration

C. Wrap the models inside a Cloud Run container using the REVISION1=20, REVISION2=80 revision configuration.

D. Implement random splitting in Dataflow using beam.Partition() with a partition function calling a Vertex AI endpoint.

Show Suggested Answer

Question #: 144
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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?

A. 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.

B. Create a Vertex AI Workbench user-managed notebooks instance with 4 V100 GPUs, and use it to train your model.

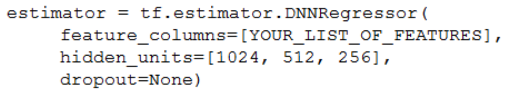
C. Package your code with Setuptools, and use a pre-built container. Train your model with Vertex AI using a custom tier that contains the required GPUs.

D. 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.

Show Suggested Answer

Question #: 145
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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're 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. Switch from CPU to GPU serving.

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

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

D. Increase the dropout rate to 0.8 in _PREDICT mode by adjusting the TensorFlow Serving parameters.

Show Suggested Answer

Question #: 146
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You work on the data science team at a manufacturing company. You are reviewing the company's historical sales data, which has hundreds of millions of records. For your exploratory data analysis, you need to calculate descriptive statistics such as mean, median, and mode; conduct complex statistical tests for hypothesis testing; and plot variations of the features over time. You want to use as much of the sales data as possible in your analyses while minimizing computational resources. What should you do?

A. Visualize the time plots in Google Data Studio. Import the dataset into Vertex AI Workbench user-managed notebooks. Use this data to calculate the descriptive statistics and run the statistical analyses.

B. Spin up a Vertex AI Workbench user-managed notebooks instance and import the dataset. Use this data to create statistical and visual analyses.

C. Use BigQuery to calculate the descriptive statistics. Use Vertex AI Workbench user-managed notebooks to visualize the time plots and run the statistical analyses.

D. Use BigQuery to calculate the descriptive statistics, and use Google Data Studio to visualize the time plots. Use Vertex AI Workbench user-managed notebooks to run the statistical analyses.

Show Suggested Answer

Question #: 148
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You are training an ML model using data stored in BigQuery that contains several values that are considered Personally Identifiable Information (PII). You need to reduce the sensitivity of the dataset before training your model. Every column is critical to your model. How should you proceed?

A. Using Dataflow, ingest the columns with sensitive data from BigQuery, and then randomize the values in each sensitive column.

B. Use the Cloud Data Loss Prevention (DLP) API to scan for sensitive data, and use Dataflow with the DLP API to encrypt sensitive values with Format Preserving Encryption.

C. Use the Cloud Data Loss Prevention (DLP) API to scan for sensitive data, and use Dataflow to replace all sensitive data by using the encryption algorithm AES-256 with a salt.

D. Before training, use BigQuery to select only the columns that do not contain sensitive data. Create an authorized view of the data so that sensitive values cannot be accessed by unauthorized individuals.

Show Suggested Answer

Question #: 149
Topic #: 1
[All Professional Machine Learning Engineer Questions]

You recently deployed an ML model. Three months after deployment, you notice that your model is underperforming on certain subgroups, thus potentially leading to biased results. You suspect that the inequitable performance is due to class imbalances in the training data, but you cannot collect more data. What should you do? (Choose two.)

A. Remove training examples of high-performing subgroups, and retrain the model.

B. Add an additional objective to penalize the model more for errors made on the minority class, and retrain the model

C. Remove the features that have the highest correlations with the majority class.

D. Upsample or reweight your existing training data, and retrain the model

E. Redeploy the model, and provide a label explaining the model's behavior to users.

Show Suggested Answer

Question #: 151
Topic #: 1
[All Professional Machine Learning Engineer Questions]

While running a model training pipeline on Vertex AI, you discover that the evaluation step is failing because of an out-of-memory error. You are currently using TensorFlow Model Analysis (TFMA) with a standard Evaluator TensorFlow Extended (TFX) pipeline component for the evaluation step. You want to stabilize the pipeline without downgrading the evaluation quality while minimizing infrastructure overhead. What should you do?

A. Include the flag -runner=DataflowRunner in beam_pipeline_args to run the evaluation step on Dataflow.

B. Move the evaluation step out of your pipeline and run it on custom Compute Engine VMs with sufficient memory.

C. Migrate your pipeline to Kubeflow hosted on Google Kubernetes Engine, and specify the appropriate node parameters for the evaluation step.

D. Add tfma.MetricsSpec () to limit the number of metrics in the evaluation step.

Show Suggested Answer

Question #: 152
Topic #: 1
[All Professional Machine Learning Engineer Questions]

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. Use sparse representation in the test set.

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

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

D. Collect more data representing all categories

Show Suggested Answer

Question #: 164

Topic #: 1

[All Professional Data Engineer Questions]

You are working on 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?

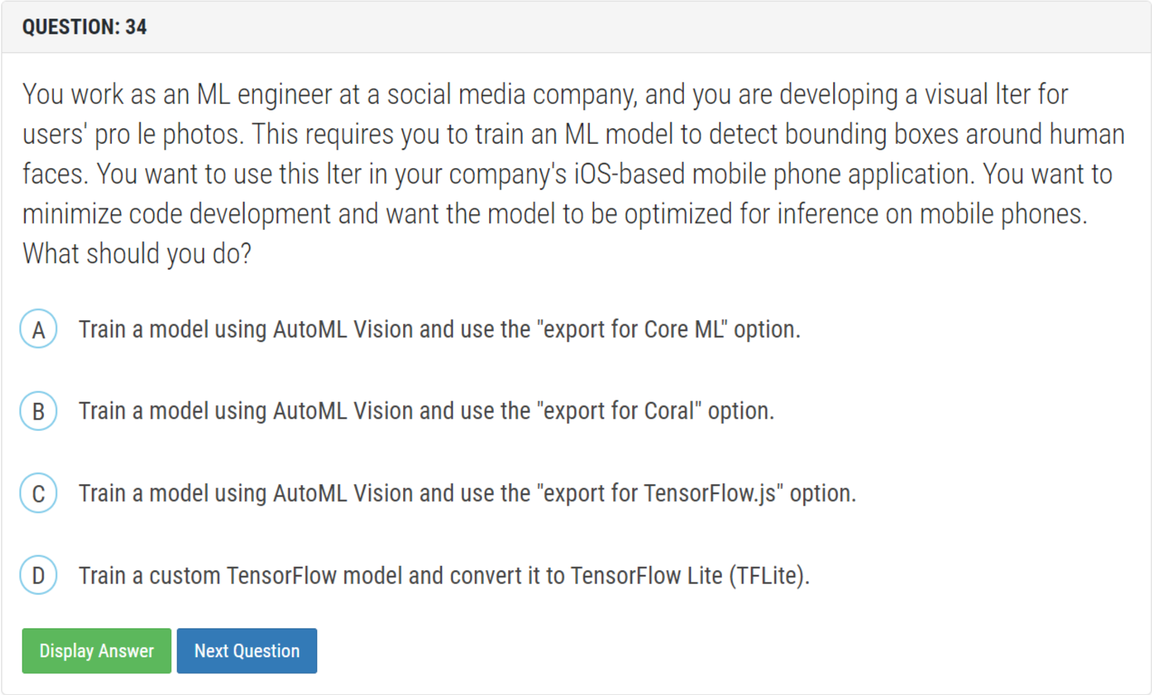
A. Create a new view with BigQuery that does not include a column with city information.

B. Use SQL in BigQuery to transform the state column using a one-hot encoding method, and make each city a column with binary values.

C. Use TensorFlow to create a categorical variable with a vocabulary list. Create the vocabulary file and upload that as part of your model to BigQuery ML.

D. Use Cloud Data Fusion to assign each city to a region that is labeled as 1, 2, 3, 4, or 5, and then use that number to represent the city in the model.

Show Suggested Answer



QUESTION: 34

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 Core ML" 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 TensorFlow.js" option.

D Train a custom TensorFlow model and convert it to TensorFlow Lite (TFLite).

QUESTION: 45

You have built a model that is trained on data stored in Parquet files. You access the data through a Hive table hosted on Google Cloud. You preprocessed these data with PySpark and exported it as a CSV file into Cloud Storage. After preprocessing, you execute additional steps to train and evaluate your model. You want to parametrize this model training in Kubeflow Pipelines. What should you do?

A Remove the data transformation step from your pipeline.

B Containerize the PySpark transformation step, and add it to your pipeline.

C Add a ContainerOp to your pipeline that spins a Dataproc cluster, runs a transformation, and then saves the transformed data in Cloud Storage.

D Deploy Apache Spark at a separate node pool in a Google Kubernetes Engine cluster. Add a ContainerOp to your pipeline that invokes a corresponding transformation job for this Spark instance.

Display Answer Next Question