



**Free Questions for Professional-Machine-Learning-Engineer
by vceexamstest**

Shared by Moran on 12-12-2023

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

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?

Options:

- 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.

Answer:

D

Explanation:

This approach would allow you to keep the critical columns of data while reducing the sensitivity of the dataset by removing the personally identifiable information (PII) before training the model. By creating an authorized view of the data, you can ensure that sensitive values cannot be accessed by unauthorized individuals.

Question 2

Question Type: MultipleChoice

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?

Options:

- 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

Explanation:

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.

Question 3

Question Type: MultipleChoice

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?

Options:

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

Answer:

C

Explanation:

BigQuery is a powerful tool for analyzing large datasets and can be used to quickly calculate descriptive statistics, such as mean, median, and mode, on large amounts of data. By using BigQuery, you can analyze the entire dataset and minimize the computational resources required for your analyses.

Once you have calculated the descriptive statistics, you can use Vertex AI Workbench user-managed notebooks to visualize the time plots and run the statistical analyses. Vertex AI Workbench allows you to interactively explore the data, create visualizations, and perform advanced statistical analysis. It's also possible to run these notebooks on a powerful GPU which will help to increase the speed of the analysis.

Question 4

Question Type: MultipleChoice

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?

Options:

- 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

Explanation:

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.

Question 5

Question Type: MultipleChoice

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?

Options:

- 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

Explanation:

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.

[Recall and Precision](#)

[Predictive Maintenance](#)

[Metrics for classification](#)

Question 6

Question Type: MultipleChoice

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)

Options:

- 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

Explanation:

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.

[Google Kubernetes Engine](#)

[TFJob operator](#)

[Vertex AI](#)

Question 7

Question Type: MultipleChoice

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?

Options:

- 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:

C

Explanation:

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.

Question 8

Question Type: MultipleChoice

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?

Options:

- 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%.

Answer:

B

Explanation:

In this case, the model has a high accuracy of 99% for identifying customers who cancel their subscriptions, but a lower accuracy of 82% for identifying customers who renew their subscriptions. However, this does not necessarily mean that the model is performing well, because 90% of the customers renew their subscription, so if the model always predicts that customers will renew, it will be correct 90% of the time. Therefore, the model's performance is worse than the baseline of always predicting that customers will renew their subscription.

https://en.wikipedia.org/wiki/Imbalanced_data

<https://machinelearningmastery.com/baseline-performance-machine-learning-algorithms/>

Question 9

Question Type: MultipleChoice

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?

Options:

- 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.

Answer:

A

Question 10

Question Type: MultipleChoice

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 50milliseconds). 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?

Options:

- 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.

Answer:

A

**To Get Premium Files for Professional-Machine-Learning-Engineer
Visit**

<https://www.p2pexams.com/products/professional-machine-learning-engineer>

For More Free Questions Visit

<https://www.p2pexams.com/google/pdf/professional-machine-learning-engineer>

