CertyIQ

CertyIQ

Premium exam material

Get certification quickly with the CertyIQ Premium exam material.
Everything you need to prepare, learn & pass your certification exam easily. Lifetime free updates
First attempt guaranteed success.

https://www.CertyIQ.com

About CertyIQ

We here at CertyIQ eventually got enough of the industry's greedy exam paid for. Our team of IT professionals comes with years of experience in the IT industry Prior to training CertIQ we worked in test areas where we observed the horrors of the paywall exam preparation system.

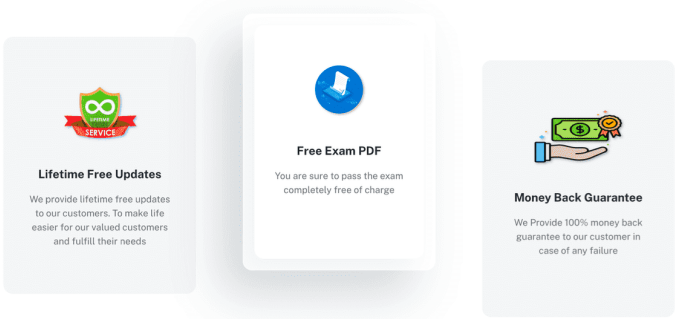
The misuse of the preparation system has left our team disillusioned. And for that reason, we decided it was time to make a difference. We had to make In this way, CertyIQ was created to provide quality materials without stealing from everyday people who are trying to make a living.

Doubt Support

We have developed a very scalable solution using which we are able to solve 400+ doubts every single day with an average rating of 4.8 out of 5.

https://www.certyiq.com

Mail us on - certyiqofficial@gmail.com



John

October 19, 2022

★★★★★

Thanks you so much for your help. I scored 972 in my exam today. More than 90% were from your PDFs!

Dana

September 04, 2022

★★★★★

Thanks a lot for this updated AZ-900 Q&A. I just passed my exam and got 974, I followed both of your Az-900 videos and the 6 PDF, the PDFs are very much valid, all answers are correct. Could you please create a similar video/PDF for DP900, your content/PDF's is really awesome. The team did a really good job. Thank You 😊.

Ahamed Shibly

2 months ago

★★★★★

Customer support is realy fast and helpful, I just finished my exam and this video along with the 6 PDF helped me pass! Definitely recommend getting the PDFs. Thank you!

October 22, 2022

★★★★★

Passed my exam today with 891 marks. Out of 52 questions, 51 were from certyiq PDFs including Contoso case study.
Thank You certyiq team!

Henry Rome

2 months ago

★★★★★

These questions are real and 100 % valid. Thank you so much for your efforts, also your 4 PDFs are awesome, I passed the DP900 exam on 1 Sept. With 968 marks. Thanks a lot, buddy!

Esmaria

2 months ago

★★★★★

Simple easy to understand explanations. To anyone out there wanting to write AZ900, I highly recommend 6 PDF's.Thank you so much, appreciate all your hard work in having such great content. Passed my exam Today - 3 September with 942 score.

Google

(Professional Machine Learning Engineer)

Professional Machine Learning Engineer

Total: **285 Questions**

Link: https://certyiq.com/papers/google/professional-machine-learning-engineer

Question: 1 CertyIQ

You are building an ML model to detect anomalies in real-time sensor data. 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 = AutoML, 3 = Cloud Functions

D.1 = BigQuery, 2 = AI Platform, 3 = Cloud Storage

Answer: A

Explanation:

Dataflow - data transformationVertex AI - model creationBig query - analytics and visualization

Question: 2 CertyIQ

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

Explanation:

No need to predict the presences since they are already confirmed, best thing we can do is optimize the route

Question: 3 CertyIQ

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

Explanation:

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

Question: 4 CertyIQ

You want to rebuild your ML pipeline for structured data on Google Cloud. You are using PySpark to conduct data transformations at scale, but your pipelines are taking over 12 hours to run. To speed up development and pipeline run time, you want to use a serverless tool and SQL syntax. You have already moved your raw data into Cloud Storage. How should you build the pipeline on Google Cloud while meeting the speed and processing requirements?

A.Use Data Fusion's GUI to build the transformation pipelines, and then write the data into BigQuery.

B.Convert your PySpark into SparkSQL queries to transform the data, and then run your pipeline on Dataproc to write the data into BigQuery.

C.Ingest your data into Cloud SQL, convert your PySpark commands into SQL queries to transform the data, and then use federated queries from BigQuery for machine learning.

D.Ingest your data into BigQuery using BigQuery Load, convert your PySpark commands into BigQuery SQL queries to transform the data, and then write the transformations to a new table.

Answer: D

Explanation:

Data Fusion is not in SQL syntax, so no A;Dataproc is not serverless, so no B;Passing through Cloud SQL is uselss, just go with BigQuery, so no C;D is correct

Question: 5 CertyIQ

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-learn, and custom libraries. What should you do?

A.Use the AI 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 TF Job.

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

Explanation:

Now it's Vertex AI (instead of AI Platform), but it's the best solution, no need to do anything more complicated

Question: 6 CertyIQ

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 dataset. What should you do?

A.Keep the original test dataset unchanged even if newer products are incorporated into retraining.

B.Extend your test dataset with images of the newer products when they are introduced to retraining.

C.Replace your test dataset with images of the newer products when they are introduced to retraining.

D.Update your test dataset with images of the newer products when your evaluation metrics drop below a pre-decided threshold.

Answer: B

Explanation:

you can't just replace the old product data with just new product, until you don't sell old product anymore

You need to correctly classify newer products, so you need the new training data ==> A is wrong;You need to keep doing a good job on older dataset, you can't just ignore it ==> C is wrong;You know when you are introducing new products, there is no need to wait for a drop in preformaces ==> D is wrong;B is correct

Question: 7 CertyIQ

You need to build classification workflows over several structured datasets currently stored in BigQuery. Because you will be performing the classification several times, you want to complete the following steps without writing code: exploratory data analysis, feature selection, model building, training, and hyperparameter tuning and serving. What should you do?

A.Configure AutoML Tables to perform the classification task.

B.Run a BigQuery ML task to perform logistic regression for the classification.

C.Use AI Platform Notebooks to run the classification model with pandas library.

D.Use AI Platform to run the classification model job configured for hyperparameter tuning.

Answer: A

Explanation:

Because BigQueryML doesn't have lots of steps that mentioned in question

"without writing code" ==> AutoMLA is correct

Question: 8 CertyIQ

You work for a public transportation company and need to build a model to estimate delay times for multiple transportation routes. Predictions are served directly to users in an app in real time. Because different seasons and population increases impact the data relevance, you will retrain the model every month. You want to follow Google-recommended best practices. How should you configure the end-to-end architecture of the predictive model?

A.Configure Kubeflow Pipelines to schedule your multi-step workflow from training to deploying your model.

B.Use a model trained and deployed on BigQuery ML, and trigger retraining with the scheduled query feature in BigQuery.

C.Write a Cloud Functions script that launches a training and deploying job on AI Platform that is triggered by

Cloud Scheduler.

D.Use Cloud Composer to programmatically schedule a Dataflow job that executes the workflow from training to deploying your model.

Answer: A

Explanation:

Answer: A A. Kubeflow Pipelines can form an end-to-end architecture (https://www.kubeflow.org/docs/components/pipelines/overview/pipelines-overview/) and deploy models.B. BigQuery ML can't offer an end-to-end architecture because it must use another tool, like AI Platform, for serving models at the end of the process (https://cloud.google.com/bigquery-ml/docs/export-model-tutorial#online_deployment_and_serving).C. Cloud Scheduler can trigger the first step in a pipeline, but then some orchestrator is needed to continue the remaining steps. Besides, having Cloud Scheduler alone can't ensure failure handling during pipeline execution.D. A Dataflow job can't deploy models, it must use AI Platform at the end instead.

Question: 9 CertyIQ

You are developing ML models with AI Platform for image segmentation on CT scans. You frequently update your model architectures based on the newest available research papers, and have to rerun training on the same dataset to benchmark their performance. You want to minimize computation costs and manual intervention while having version control for your code. What should you do?

A.Use Cloud Functions to identify changes to your code in Cloud Storage and trigger a retraining job.

B.Use the gcloud command-line tool to submit training jobs on AI Platform when you update your code.

C.Use Cloud Build linked with Cloud Source Repositories to trigger retraining when new code is pushed to the repository.

D.Create an automated workflow in Cloud Composer that runs daily and looks for changes in code in Cloud Storage using a sensor.

Answer: C

Explanation:

C follows a best practice, B is a manual step

C is the correct answer, it's the Google recommended approach;Checking for changes in code without using Cloud Source Repository is a bad choice, so no A and B;Cloud Composer is an overkill, so no D.

Question: 10 CertyIQ

Your team needs to build a model that predicts whether images contain a driver's license, passport, or credit card. The data engineering team already built the pipeline and generated a dataset composed of 10,000 images with driver's licenses, 1,000 images with passports, and 1,000 images with credit cards. You now have to train a model with the following label map: [`˜drivers_license', `˜passport', `˜credit_card']. Which loss function should you use?

A.Categorical hinge

B.Binary cross-entropy

C.Categorical cross-entropy

D.Sparse categorical cross-entropy

Answer: C

Explanation:

It's C. Sparse categorical cross-entropy is not required when the number of classes is only 3.

Answer is C https://stackoverflow.com/questions/58565394/what-is-the-difference-between-sparse-categorical-crossentropy-and-categorical-c

Question: 11 CertyIQ

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.

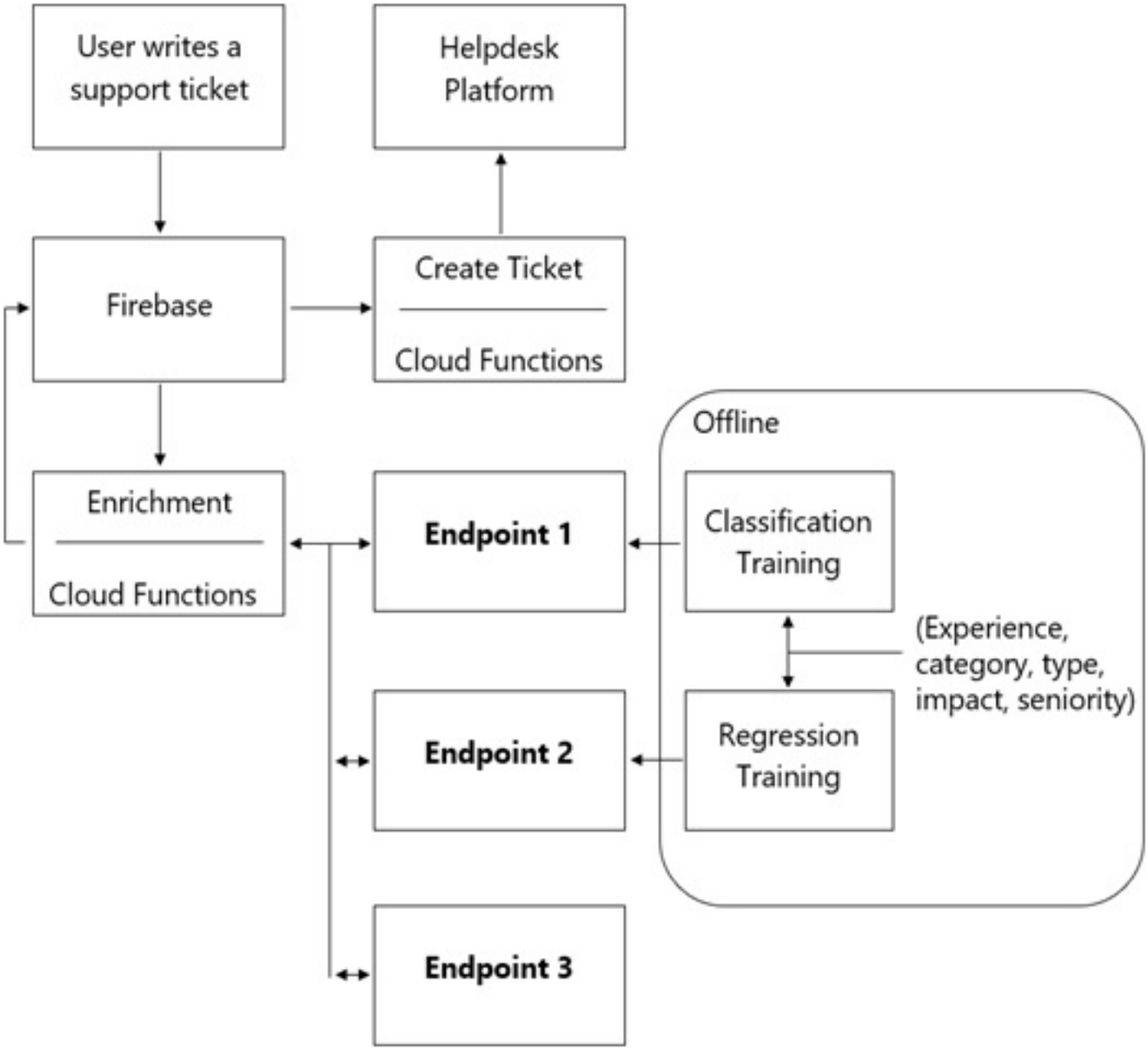
Answer: B

Explanation:
Frequently bought together' recommendations aim to up-sell and cross-sell customers by providing product.

Reference:
https://rejoiner.com/resources/amazon-recommendations-secret-selling-online/

Question: 12 CertyIQ

You are designing an architecture with a serverless ML system to enrich customer support tickets with informative metadata before they are routed to a support agent. You need a set of models to predict ticket priority, predict ticket resolution time, and perform sentiment analysis to help agents make strategic decisions when they process support requests. Tickets are not expected to have any domain-specific terms or jargon.
The proposed architecture has the following flow:



Which endpoints should the Enrichment Cloud Functions call?

A.1 = AI Platform, 2 = AI Platform, 3 = AutoML Vision
B.1 = AI Platform, 2 = AI Platform, 3 = AutoML Natural Language
C.1 = AI Platform, 2 = AI Platform, 3 = Cloud Natural Language API
D.1 = Cloud Natural Language API, 2 = AI Platform, 3 = Cloud Vision API

Answer: C

Explanation:

ANS: CThis is the exact solution by Google: https://web.archive.org/web/20210618072649/https://cloud.google.com/architecture/architecture-of-a-serverless-ml-model#architecture

AI Platform (now Vertex AI) for both the predictions and Natural Language API for sentiment analysis since there are no specific terms (so no need to custom build something with an AutoML), so C

Question: 13 CertyIQ

You have trained a deep neural network model on Google Cloud. The model has low loss on the training data, but is performing worse on the validation data. You want the model to be resilient to overfitting. Which strategy should you use when retraining the model?

A.Apply a dropout parameter of 0.2, and decrease the learning rate by a factor of 10.

B.Apply a L2 regularization parameter of 0.4, and decrease the learning rate by a factor of 10.

C.Run a hyperparameter tuning job on AI Platform to optimize for the L2 regularization and dropout parameters.

D.Run a hyperparameter tuning job on AI Platform to optimize for the learning rate, and increase the number of neurons by a factor of 2.

Answer: C

Explanation:

ANS: CA and B are random values, why they choose that values?D could increase even more overfitting since you're using a more complex model.

We don't know the optimum values for the parameters, so we need to run a hyperparameter tuning job; L2 regularization and dropout parameters are great ways to avoid overfitting.So C is the answer

Question: 14 CertyIQ

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

Explanation:

B is correct. Model needs to keep up with the market changes, implying that the underlying data distribution would be changing as well. Hence retrain the model.

The questions says the model is required to keep up with market changes, hence retraining needed.

Question: 15 CertyIQ

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

Explanation:

Converting your data into TFRecord has many advantages, such as: More efficient storage: the TFRecord data can take up less space than the original data; it can also be partitioned into multiple files. Fast I/O: the TFRecord format can be read with parallel I/O operations, which is useful for TPUs or multiple hosts

Question: 16 CertyIQ

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 model's 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

Explanation:

RNN are a fit tool to work with time-series as this one, so C

Question: 17 CertyIQ

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.Stream all files to Google Cloud, and then write the data to BigQuery. Periodically conduct a bulk scan of the table using the DLP API.

B.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.

C.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.

D.Create three buckets of data: Quarantine, Sensitive, and Non-sensitive. Write all data to the Quarantine bucket. Periodically conduct a bulk scan of that bucket using the DLP API, and move the data to either the Sensitive or Non-Sensitive bucket.

Answer: D

Explanation:

D is the right answer: you can temporarily store the sensitive data in a Quarantine bucket with restricted access, then move the data to the relative buckets once the PII have been protected.

Question: 18 CertyIQ

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 20 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 sets from 30 days after your validation set.

Answer: D

Explanation:

D. As time signal that is spread across multiple columns so manual split is required.

Question: 19 CertyIQ

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

A.Write a script that sequentially performs the push to your development branch and executes the unit tests on Cloud Run.

B.Using Cloud Build, set an automated trigger to execute the unit tests when changes are pushed to your development branch.

C.Set up a Cloud Logging sink to a Pub/Sub topic that captures interactions with Cloud Source Repositories. Configure a Pub/Sub trigger for Cloud Run, and execute the unit tests on Cloud Run.

D.Set up a Cloud Logging sink to a Pub/Sub topic that captures interactions with Cloud Source Repositories. Execute the unit tests using a Cloud Function that is triggered when messages are sent to the Pub/Sub topic.

Answer: B

Explanation:

B. GCP recommends to use Cloud Build when building KubeFlow Pipelines. It's possible to run unit tests in Cloud Build. And, the others seems overly complex/unnecessary

Question: 20 CertyIQ

You are training an LSTM-based model on AI Platform to summarize text using the following job submission script:
gcloud ai-platform jobs submit training $JOB_NAME \
--package-path $TRAINER_PACKAGE_PATH \
--module-name $MAIN_TRAINER_MODULE \
--job-dir $JOB_DIR \

--region $REGION \
--scale-tier basic \
-- \
--epochs 20 \
--batch_size=32 \
--learning_rate=0.001 \
You want to ensure that training time is minimized without significantly compromising the accuracy of your model. What should you do?

A.Modify the 'epochs' parameter.

B.Modify the 'scale-tier' parameter.

C.Modify the 'batch size' parameter.

D.Modify the 'learning rate' parameter.

Answer: B

Explanation:

A is incorrect, less training iteration will affect model performance.B is correct, cost is not a concern as it is not mentioned in the question, the scale tier can be upgraded to significantly minimize the training time.C is incorrect, wouldn't affect training time, but would affect model performance.D is incorrect, the model might converge faster with higher learning rate, but this would affect the training routine and might cause exploding gradients.

Question: 21 CertyIQ

You have deployed multiple versions of an image classification model on AI Platform. You want to monitor the performance of the model versions over time. How should you perform this comparison?

A.Compare the loss performance for each model on a held-out dataset.

B.Compare the loss performance for each model on the validation data.

C.Compare the receiver operating characteristic (ROC) curve for each model using the What-If Tool.

D.Compare the mean average precision across the models using the Continuous Evaluation feature.

Answer: D

Explanation:

If you have multiple model versions in a single model and have created an evaluation job for each one, you can view a chart comparing the mean average precision of the model versions over time

Question: 22 CertyIQ

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)
What is the correct way to write the predict request?

A.data = json.dumps(signature_name: seving_default, instances [['ab', 'bc', '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'], ['d', 'e', 'f']])

D.data = json.dumps(signature_name: serving_default, instances [['a', 'b'], ['c', 'd'], ['e', 'f']])

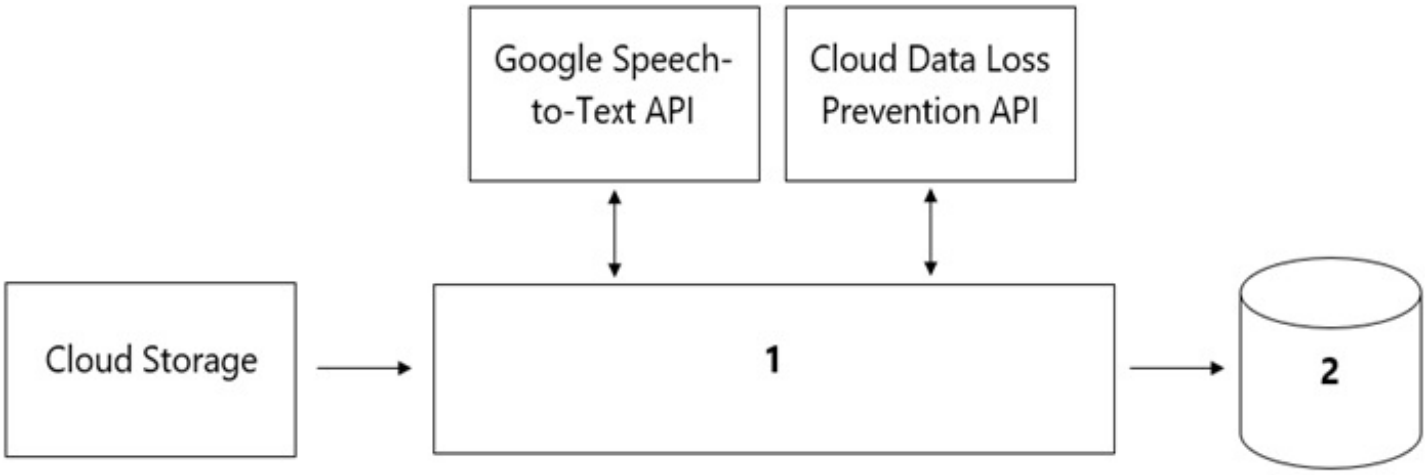
Answer: D

Explanation:

Having "shape=[-1,2]", the input can have as many rows as we want, but each row needs to be of 2 elements.
The only option satisfying this requirement is D.

Question: 23 CertyIQ

Your organization's call center has asked you to develop a model that analyzes customer sentiments in each call. The call center receives over one million calls daily, and data is stored in Cloud Storage. The data collected must not leave the region in which the call originated, and no Personally Identifiable Information (PII) can be stored or analyzed. The data science team has a third-party tool for visualization and access which requires a SQL ANSI-2011 compliant interface. You need to select components for data processing and for analytics. How should the data pipeline be designed?



A.1= Dataflow, 2= BigQuery

B.1 = Pub/Sub, 2= Datastore

C.1 = Dataflow, 2 = Cloud SQL

D.1 = Cloud Function, 2= Cloud SQL

Answer: A

Explanation:

we need a dataflow to process data from cloud storage and data is unstructured and if we want to perform analysis on unstructured with SQL interface BIgQuery is the only option

Question: 24 CertyIQ

You are an ML engineer at a global shoe store. You manage the ML models for the company's website. You are asked to build a model that will recommend new products to the user based on their purchase behavior and similarity with other users. What should you do?

A.Build a classification model

B.Build a knowledge-based filtering model

C.Build a collaborative-based filtering model

D.Build a regression model using the features as predictors

Answer: C

Explanation:

C. Collaborative filtering is about user similarity and product recommendations. Other models won't work

Reference:

https://cloud.google.com/solutions/recommendations-using-machine-learning-on-compute-engine

Question: 25 CertyIQ

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

A.Increase the recall.

B.Decrease the recall.

C.Increase the number of false positives.

D.Decrease the number of false negatives.

Answer: B

Explanation:

Option B is the best approach because decreasing the threshold will increase the precision by reducing the number of false positives.

A , C , D they are the same. So I go with B , it is threshold adjustment from 0.5 +-

Question: 26 CertyIQ

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

Explanation:

D.Datafusion is more designed for data ingestion from one source to another one, with few transformation. Dataprep is more designed for data preparation (as its name means), data cleaning, new column creation, splitting column. Dataprep also provide insight of the data for helping you in your recipes.

Question: 27 CertyIQ

You are an ML engineer at a regulated insurance company. You are asked to develop an insurance approval model that accepts or rejects insurance applications from potential customers. What factors should you consider before building the model?

A.Redaction, reproducibility, and explainability
B.Traceability, reproducibility, and explainability
C.Federated learning, reproducibility, and explainability
D.Differential privacy, federated learning, and explainability

Answer: B

Explanation:

B. Traceability, reproducibility, and explainabilityWhen developing an insurance approval model, it's crucial to consider several factors to ensure that the model is fair, accurate, and compliant with regulations. The factors to consider include:Traceability: It's important to be able to trace the data used to build the model and the decisions made by the model. This is important for transparency and accountability.Reproducibility: The model should be built in a way that allows for its reproducibility. This means that other researchers should be able to reproduce the same results using the same data and methods.Explainability: The model should be able to provide clear and understandable explanations for its decisions. This is important for building trust with customers and ensuring compliance with regulations.Other factors that may also be important to consider, depending on the specific context of the insurance company and its customers, include data privacy and security, fairness, and bias mitigation.

Question: 28 CertyIQ

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 two.)

A.Use the interleave option for reading data.

B.Reduce the value of the repeat parameter.

C.Increase the buffer size for the shuttle option.

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

E.Decrease the batch size argument in your transformation.

Answer: AD

Explanation:

AD - agree with danielp1By the way, this is handy to understand the significance of shuffle buffer_size: https://stackoverflow.com/a/48096625/1933315

Question: 29 CertyIQ

You have trained a model on a dataset that required computationally expensive preprocessing operations. You need to execute the same preprocessing at prediction time. You deployed the model on AI Platform for high-throughput online prediction. Which architecture should you use?

A.Validate the accuracy of the model that you trained on preprocessed data. Create a new model that uses the raw data and is available in real time. Deploy the new model onto AI Platform for online prediction.

B.Send incoming prediction requests to a Pub/Sub topic. Transform the incoming data using a Dataflow job. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue.

C.Stream incoming prediction request data into Cloud Spanner. Create a view to abstract your preprocessing logic. Query the view every second for new records. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue.

D.Send incoming prediction requests to a Pub/Sub topic. Set up a Cloud Function that is triggered when messages are published to the Pub/Sub topic. Implement your preprocessing logic in the Cloud Function. Submit a prediction request to AI Platform using the transformed data. Write the predictions to an outbound Pub/Sub queue.

Answer: B

Explanation:

using options eliminatios , A totally wrong , D also not valid as cloud functions is not sutiable for heavy data workflows- answer between B,D will vote for B as dataflow is the best solution while dealing with heavy data workflows

Question: 30 CertyIQ

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 data. 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

Explanation:

Creating alerts to monitor for skew in the input data can help to detect when the distribution of the data has changed and the model's performance is affected. Once a skew is detected, retraining the model with the new data can improve its performance.

Question: 31 CertyIQ

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 = 224 — 224
- 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

Explanation:

to fix memory overflow you need to reduce batch size also reduce input resolution is validbut reducing image size can harm model performance , so answer is B

Reference:

https://github.com/tensorflow/tensorflow/issues/136

Question: 32 CertyIQ

You developed an ML model with AI Platform, and you want to move it to production. You serve a few thousand queries per second and are experiencing latency issues. Incoming requests are served by a load balancer that distributes them across multiple Kubeflow CPU-only pods running on Google Kubernetes Engine
(GKE). Your goal is to improve the serving latency without changing the underlying infrastructure. What should you do?

A.Significantly increase the max_batch_size TensorFlow Serving parameter.

B.Switch to the tensorflow-model-server-universal version of TensorFlow Serving.

C.Significantly increase the max_enqueued_batches TensorFlow Serving parameter.

D.Recompile TensorFlow Serving using the source to support CPU-specific optimizations. Instruct GKE to choose an appropriate baseline minimum CPU platform for serving nodes.

Answer: A

Explanation:

A is correct. max_batch_size TensorFlow Serving parameter

CPU-only: One ApproachIf your system is CPU-only (no GPU), then consider starting with the following values: num_batch_threads equal to the number of CPU cores; max_batch_size to a really high value; batch_timeout_micros to 0. Then experiment with batch_timeout_micros values in the 1-10 millisecond (1000-10000 microsecond) range, while keeping in mind that 0 may be the optimal value.https://github.com/tensorflow/serving/tree/master/tensorflow_serving/batching

Question: 33 CertyIQ

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.Normalize the data using Google Kubernetes Engine.

B.Translate the normalization algorithm into SQL for use with BigQuery.

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: B

Explanation:

BiqQuery definitely minimizes computational time for normalization. I think it would also minimize manual intervention. For data normalization in dataflow you'd have to pass in values of mean and standard deviation as a side-input. That seems more work than a simple SQL query

Question: 34 CertyIQ

You need to design a customized deep neural network in Keras that will predict customer purchases based on their purchase history. You want to explore model performance using multiple model architectures, store training data, and be able to compare the evaluation metrics in the same dashboard. What should you do?

A.Create multiple models using AutoML Tables.

B.Automate multiple training runs using Cloud Composer.

C.Run multiple training jobs on AI Platform with similar job names.

D.Create an experiment in Kubeflow Pipelines to organize multiple runs.

Answer: D

Explanation:

With Kubeflow Pipelines, you can create experiments that help you keep track of multiple training runs with different model architectures and hyperparameters.

https://www.kubeflow.org/docs/components/pipelines/concepts/experiment/https://www.kubeflow.org/docs/comp

Question: 35 CertyIQ

You are developing a Kubeflow pipeline on Google Kubernetes Engine. The first step in the pipeline is to issue a query against BigQuery. You plan to use the results of that query as the input to the next step in your pipeline. You want to achieve this in the easiest way possible. What should you do?

A.Use the BigQuery console to execute your query, and then save the query results into a new BigQuery table.

B.Write a Python script that uses the BigQuery API to execute queries against BigQuery. Execute this script as the first step in your Kubeflow pipeline.

C.Use the Kubeflow Pipelines domain-specific language to create a custom component that uses the Python BigQuery client library to execute queries.

D.Locate the Kubeflow Pipelines repository on GitHub. Find the BigQuery Query Component, copy that component's URL, and use it to load the component into your pipeline. Use the component to execute queries against BigQuery.

Answer: D

Explanation:

Not sure what is the reason behind putting A as it is manual and manual steps can not be part of automation. I would say Answer is D as it just require a clone of the component from github. Using a Python and import bigquery component may sounds good too, but ask was what is easiest. It depends how word "easy" is taken by individuals but definitely not A.

https://linuxtut.com/en/f4771efee37658c083cc/

Question: 36 CertyIQ

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: B

Explanation:

train accuracy 97% , production accuracy 66% ---> time series data ---> random split ---> cause leakage , answer is B

You don't split data randomly for time series prediction.

Question: 37 CertyIQ

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 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

Explanation:

Option A is the best choice as AI Platform provides a distributed training framework, enabling you to train large-scale models faster and with less effort

using options eliminations answer between A,D will vote for A as it is easier

Question: 38 CertyIQ

You have trained a text classification model in TensorFlow using AI Platform. You want to use the trained model for batch predictions on text data stored in
BigQuery while minimizing computational overhead. What should you do?

A.Export the model to BigQuery ML.

B.Deploy and version the model on AI Platform.

C.Use Dataflow with the SavedModel to read the data from BigQuery.

D.Submit a batch prediction job on AI Platform that points to the model location in Cloud Storage.

Answer: D

Explanation:

D is more straightforward

To perform batch predictions on text data stored in BigQuery using a trained TensorFlow model, you can submit a batch prediction job on AI Platform. The batch prediction job reads the input data from BigQuery and the model from Cloud Storage. This approach minimizes computational overhead since the job is handled by AI Platform, and it allows you to easily scale up or down depending on the size of the data.

Question: 39 CertyIQ

You work with a data engineering team that has developed a pipeline to clean your dataset and save it in a Cloud Storage bucket. You have created an ML model and want to use the data to refresh your model as soon as new data is available. As part of your CI/CD workflow, you want to automatically run a Kubeflow
Pipelines training job on Google Kubernetes Engine (GKE). How should you architect this workflow?

A.Configure your pipeline with Dataflow, which saves the files in Cloud Storage. After the file is saved, start the training job on a GKE cluster.

B.Use App Engine to create a lightweight python client that continuously polls Cloud Storage for new files. As soon as a file arrives, initiate the training job.

C.Configure a Cloud Storage trigger to send a message to a Pub/Sub topic when a new file is available in a storage bucket. Use a Pub/Sub-triggered Cloud Function to start the training job on a GKE cluster.

D.Use Cloud Scheduler to schedule jobs at a regular interval. For the first step of the job, check the timestamp of objects in your Cloud Storage bucket. If there are no new files since the last run, abort the job.

Answer: C

Explanation:

The scenario involves automatically running a Kubeflow Pipelines training job on GKE as soon as new data becomes available. To achieve this, we can use Cloud Storage to store the cleaned dataset, and then configure a Cloud Storage trigger that sends a message to a Pub/Sub topic whenever a new file is added to the storage bucket. We can then create a Pub/Sub-triggered Cloud Function that starts the training job on a GKE cluster.

https://cloud.google.com/architecture/architecture-for-mlops-using-tfx-kubeflow-pipelines-and-cloud-build#triggering-and-scheduling-kubeflow-pipelines

Question: 40 CertyIQ

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

A.Decrease the number of parallel trials.

B.Decrease the range of floating-point values.

C.Set the early stopping parameter to TRUE.

D.Change the search algorithm from Bayesian search to random search.

E.Decrease the maximum number of trials during subsequent training phases.

Answer: CE

Explanation:

Answer C,E A. Decrease the number of parallel trials : doing this will of course make Hypertuning take more time , we need to increase parallel trials not decreaseB.Decrease the range of floating-point values : theoretically this should speed up the computation but this is not the most correct answerC. Set the early stopping parameter to TRUE : this is very good optionD. Change the search algorithm from Bayesian search to random search : also searching the search algorithm will not have a great impactE. Decrease the maximum number of trials during subsequent training phases : very good option

Question: 41 CertyIQ

Your team is building an application for a global bank that will be used by millions of customers. You built a forecasting model that predicts customers' account balances 3 days in the future. Your team will use the results in a new feature that will notify users when their account balance is likely to drop below $25. How should you serve your predictions?

A.1. Create a Pub/Sub topic for each user. 2. Deploy a Cloud Function that sends a notification when your model predicts that a user's account balance will drop below the $25 threshold.

B.1. Create a Pub/Sub topic for each user. 2. Deploy an application on the App Engine standard environment that sends a notification when your model predicts that a user's account balance will drop below the $25 threshold.

C.1. Build a notification system on Firebase. 2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when the average of all account balance predictions drops below the $25 threshold.

D.1. Build a notification system on Firebase. 2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when your model predicts that a user's account balance will drop

below the $25 threshold.

Answer: D

Explanation:

D is correct. Firebase is designed for exactly this sort of scenario. Also, it would not be possible to create millions of pubsub topics due to GCP quotashttps://cloud.google.com/pubsub/quotas#quotashttps://firebase.google.com/docs/cloud-messaging

"Create a Pub/Sub topic for each user" xD

Question: 42 CertyIQ

You work for an advertising company and want to understand the effectiveness of your company's latest advertising campaign. You have streamed 500 MB of campaign data into BigQuery. You want to query the table, and then manipulate the results of that query with a pandas dataframe in an AI Platform notebook.
What should you do?

A.Use AI Platform Notebooks' BigQuery cell magic to query the data, and ingest the results as a pandas dataframe.

B.Export your table as a CSV file from BigQuery to Google Drive, and use the Google Drive API to ingest the file into your notebook instance.

C.Download your table from BigQuery as a local CSV file, and upload it to your AI Platform notebook instance. Use pandas.read_csv to ingest he file as a pandas dataframe.

D.From a bash cell in your AI Platform notebook, use the bq extract command to export the table as a CSV file to Cloud Storage, and then use gsutil cp to copy the data into the notebook. Use pandas.read_csv to ingest the file as a pandas dataframe.

Answer: A

Explanation:

A, Using the command %%bigquery df

%%bigquery dfSELECT name, SUM(number) as countFROM `bigquery-public-data.usa_names.usa_1910_current`GROUP BY nameORDER BY count DESCLIMIT 3print(df.head())

Question: 43 CertyIQ

You are an ML engineer at a global car manufacture. You need to build an ML model to predict car sales in different cities around the world. Which features or feature crosses should you use to train city-specific relationships between car type and number of sales?

A.Thee individual features: binned latitude, binned longitude, and one-hot encoded car type.

B.One feature obtained as an element-wise product between latitude, longitude, and car type.

C.One feature obtained as an element-wise product between binned latitude, binned longitude, and one-hot encoded car type.

D.Two feature crosses as an element-wise product: the first between binned latitude and one-hot encoded car type, and the second between binned longitude and one-hot encoded car type.

Answer: C

Explanation:

https://developers.google.com/machine-learning/crash-course/feature-crosses/check-your-understandingAnswer C: It needs a feature cross to obtain one feature.

Question: 44 CertyIQ

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.Use the AI Platform Training built-in algorithms to create a custom model.

B.Use AutoMlL Natural Language to extract custom entities for classification.

C.Use the Cloud Natural Language API to extract custom entities for classification.

D.Build a custom model to identify the product keywords from the transcribed calls, and then run the keywords through a classification algorithm.

Answer: B

Explanation:

AutoML is appropriate to classify incoming calls by product (Custom) to be routed to the correct support team.Cloud Natural Language API is for general case (not particular business)

"minimize data preprocessing and development time" answer will be limited to B,Cwill choose C as Natural Language API does not handle custom operation

Question: 45 CertyIQ

You are training a TensorFlow model on a structured dataset with 100 billion records stored in several CSV files. You need to improve the input/output execution performance. What should you do?

A.Load the data into BigQuery, and read the data from BigQuery.

B.Load the data into Cloud Bigtable, and read the data from Bigtable.

C.Convert the CSV files into shards of TFRecords, and store the data in Cloud Storage.

D.Convert the CSV files into shards of TFRecords, and store the data in the Hadoop Distributed File System (HDFS).

Answer: C

Explanation:

Cloud Bigtable is typically used to process unstructured data, such as time-series data, logs, or other types of data that do not conform to a fixed schema. However, Cloud Bigtable can also be used to store structured data if necessary, such as in the case of a key-value store or a database that does not require complex relational queries.

Option C, converting the CSV files into shards of TFRecords and storing the data in Cloud Storage, is the most appropriate solution for improving input/output execution performance in this scenario

Question: 46 CertyIQ

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: A

Explanation:

Because aggregated data can be sent at the end of the day for batch prediction and AI platform is managed so satisfy minimal intervention requirementNot B as violates minimal intervention requirementNot C and D as real-time or online inference is not needed since data is aggregated at the end of the day

Question: 47 CertyIQ

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

Answer: A

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)

Question: 48 CertyIQ

You started working on a classification problem with time series data and achieved an area under the receiver operating characteristic curve (AUC ROC) value of
99% for training data after just a few experiments. You haven't explored using any sophisticated algorithms or spent any time on hyperparameter tuning. What should your next step be to identify and fix the problem?

A.Address the model overfitting by using a less complex algorithm.

B.Address data leakage by applying nested cross-validation during model training.

C.Address data leakage by removing features highly correlated with the target value.

D.Address the model overfitting by tuning the hyperparameters to reduce the AUC ROC value.

Answer: B

Explanation:

Ans: B High correlation doesn't mean leakage. The question may suggest target leakage and the defining point of this leakage is the availability of data after the target is available.
(https://www.kaggle.com/dansbecker/data-leakage)

https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9)

Question: 49 CertyIQ

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 , 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

Explanation:

Security => not A.B: doesn't handle processing with banner inventory.D: deployment on GKE is less simple than on AI Platform. Besides, MemoryStore is in-memory while banners are stored persistently.Ans: C

Question: 50 CertyIQ

Your team is building a convolutional neural network (CNN)-based architecture from scratch. The preliminary experiments running on your on-premises CPU-only infrastructure were encouraging, but have slow convergence. You have been asked to speed up model training to reduce time-to-market. You want to experiment with virtual machines (VMs) on Google Cloud to leverage more powerful hardware. Your code does not include any manual device placement and has not been wrapped in Estimator model-level abstraction. Which environment should you train your model on?

A.AVM on Compute Engine and 1 TPU with all dependencies installed manually.

B.AVM on Compute Engine and 8 GPUs with all dependencies installed manually.

C.A Deep Learning VM with an n1-standard-2 machine and 1 GPU with all libraries pre-installed.

D.A Deep Learning VM with more powerful CPU e2-highcpu-16 machines with all libraries pre-installed.

Answer: C

Explanation:

ANS: Cto support CNN, you should use GPU. for preliminary experiment, pre-installed pkgs/libs are good choice.https://cloud.google.com/deep-learning-

vm/docs/cli#creating_an_instance_with_one_or_more_gpushttps://cloud.google.com/deep-learning-vm/docs/introduction#pre-installed_packages

Question: 51 CertyIQ

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

A.Set up restrictive IAM permissions on the AI Platform notebooks so that only a single user or group can access a given instance.

B.Separate each data scientist's work into a different project to ensure that the jobs, models, and versions created by each data scientist are accessible only to that user.

C.Use labels to organize resources into descriptive categories. Apply a label to each created resource so that users can filter the results by label when viewing or monitoring the resources.

D.Set up a BigQuery sink for Cloud Logging logs that is appropriately filtered to capture information about AI Platform resource usage. In BigQuery, create a SQL view that maps users to the resources they are using

Answer: C

Explanation:

Restricting access is not scalable and creates silos - better to document sharable resources through tagging, hence C.

CResource tagging/labeling is the best way to manage ML resources for medium/big data science teams.

Question: 52 CertyIQ

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 GPU quota in the selected region.

B.Ensure that the required GPU is available in the selected region.

C.Ensure that you have preemptible GPU quota in the selected region.

D.Ensure that the selected GPU has enough GPU memory for the workload.

Answer: B

Explanation:

The error says the resource was not found - hence B.If quota was the problem (A) then you'd see a different error message.

ANS: B https://cloud.google.com/deep-learning-vm/docs/troubleshooting#resource_not_found https://cloud.google.com/compute/docs/gpus/gpu-regions-zones Resource not foundSymptom: - The resource 'projects/deeplearning-platform/zones/europe-west4-c/acceleratorTypes/nvidia-tesla-k80' was not foundProblem: You are trying to create an instance with one or more GPUs in a region where GPUs are not available (for example, an instance with a K80 GPU in europe-west4-c).Solution: To determine which region has the required GPU, see GPUs on Compute Engine.

Question: 53 CertyIQ

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, SentenceDl3, ...]
    TextD2: [SentenceD21, SentenceD22, SentenceD23, ...]
    ...
...
```

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.Distribute texts randomly across the train-test-eval subsets: Train set: [TextA1, TextB2, ...] Test set: [TextA2, TextC1, TextD2, ...] Eval set: [TextB1, TextC2, TextD1, ...]

B.Distribute authors randomly across the train-test-eval subsets: (*) Train set: [TextA1, TextA2, TextD1, TextD2, ...] Test set: [TextB1, TextB2, ...] Eval set: [TexC1,TextC2 ...]

C.Distribute sentences randomly across the train-test-eval subsets: Train set: [SentenceA11, SentenceA21, SentenceB11, SentenceB21, SentenceC11, SentenceD21 ...] Test set: [SentenceA12, SentenceA22, SentenceB12, SentenceC22, SentenceC12, SentenceD22 ...] Eval set: [SentenceA13, SentenceA23, SentenceB13, SentenceC23, SentenceC13, SentenceD31 ...]

D.Distribute paragraphs of texts (i.e., chunks of consecutive sentences) across the train-test-eval subsets: Train set: [SentenceA11, SentenceA12, SentenceD11, SentenceD12 ...] Test set: [SentenceA13, SentenceB13, SentenceB21, SentenceD23, SentenceC12, SentenceD13 ...] Eval set: [SentenceA11, SentenceA22, SentenceB13, SentenceD22, SentenceC23, SentenceD11 ...]

Answer: B

Explanation:

https://cloud.google.com/automl-tables/docs/prepare#splithttps://developers.google.com/machine-learning/crash-course/18th-century-literature

Ans BThe model is to predict which political party the author belongs to, not which political party the text belongs to... You do not have the information of the political party of each text, you are assuming that the texts are associated with the political party of the author.

Question: 54 CertyIQ

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: C

Explanation:

Usage of Tensorflow, can build a simple model by using a sentence embedding and a single layer classifier.

"You analyzed the requirements and decided to use TensorFlow" this will make choices to reduce to C and D- " so that you have full control of the model's code " will make us choose C

Question: 55 CertyIQ

You recently joined a machine learning team that will soon release a new project. As a lead on the project, you are asked to determine the production readiness of the ML components. The team has already tested features and data, model development, and infrastructure. Which additional readiness check should you recommend to the team?

A.Ensure that training is reproducible.

B.Ensure that all hyperparameters are tuned.

C.Ensure that model performance is monitored.

D.Ensure that feature expectations are captured in the schema.

Answer: C

Explanation:

I'll go with C. Monitoring model performance is an important aspect of production readiness. It allows the team to detect and respond to changes in performance that may affect the quality of the model. The other options are also important, but they are more focused on the development phase of the project rather than the production phase.

Hey! all guysA+B+D=The team has already tested features and data, model development, and infrastructure. we are about to go live with production. Monitoring readiness is the last thing to account for.It will be very rediculous if you launch model as production regardless of how we will have about monitoring. you will lauch model as production for while and will make plan to model performance monitoring later ??? you are too reckless.Pls . Read it carefully https://developers.google.com/machine-learning/testing-debugging/pipeline/productionhttps://developers.google.com/machine-learning/testing-debugging/pipeline/overview#what-is-an-ml-pipeline.You Most guys prefer A : https://developers.google.com/machine-learning/testing-debugging/pipeline/deploying I think that it is all about model development prior to deploying .

Question: 56 CertyIQ

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

Explanation:

Detection of fraudulent transactions seems to be imbalanced data.https://cloud.google.com/automl-tables/docs/train#opt-objAUC ROC : Distinguish between classes. Default value for binary classification.AUC PROptimize results for predictions for the less common class.it is straightforward to answer, you just have to capture key word to get the right way. (Almost banlanced Or Imbalanced)https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python/When to Use ROC vs. Precision-Recall Curves?Generally, the use of ROC curves and precision-recall curves are as follows:ROC curves should be used when there are roughly equal numbers of observations for each class.Precision-Recall curves should be used when there is a moderate to large class imbalance.

Question: 57 CertyIQ

Your company manages a video sharing website where users can watch and upload videos. You need to create an ML model to predict which newly uploaded videos will be the most popular so that those videos can be prioritized on your company's website. Which result should you use to determine whether the model is successful?

A.The model predicts videos as popular if the user who uploads them has over 10,000 likes.

B.The model predicts 97.5% of the most popular clickbait videos measured by number of clicks.

C.The model predicts 95% of the most popular videos measured by watch time within 30 days of being uploaded.

D.The Pearson correlation coefficient between the log-transformed number of views after 7 days and 30 days after publication is equal to 0.

Answer: C

Explanation:

Ans: C it; though it's just an example.)(A) The absolute number of likes shouldn't be used because no information about subscribers or visits to the website is provided. The number may vary.(B) Clickbait videos are a subset of uploaded videos. Using them is an improper criterion.(D) The coefficient should reach 1. (Ref:https://arxiv.org/pdf/1510.06223.pdf)

Reference:

https://developers.google.com/machine-learning/problem-framing/framing#quantify

Question: 58 CertyIQ

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

Explanation:

B:"The presence of feature value X in the formula will affect the step size of the gradient descent. The difference in ranges of features will cause different step sizes for each feature. To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features, we scale the data before feeding it to the model."

Question: 59 CertyIQ

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

A.Use Kubeflow Pipelines to execute the experiments. Export the metrics file, and query the results using the Kubeflow Pipelines API.

B.Use AI Platform Training to execute the experiments. Write the accuracy metrics to BigQuery, and query the results using the BigQuery API.

C.Use AI Platform Training to execute the experiments. Write the accuracy metrics to Cloud Monitoring, and query the results using the Monitoring API.

D.Use AI Platform Notebooks to execute the experiments. Collect the results in a shared Google Sheets file, and query the results using the Google Sheets API.

Answer: A

Explanation:

ANS: Ahttps://codelabs.developers.google.com/codelabs/cloud-kubeflow-pipelines-gisKubeflow Pipelines (KFP) helps solve these issues by providing a way to deploy robust, repeatable machine learning pipelines along with monitoring, auditing, version tracking, and reproducibility. Cloud AI Pipelines makes it easy to set up a KFP installation.

Question: 60 CertyIQ

You work for a bank and are building a random forest model for fraud detection. You have a dataset that includes transactions, of which 1% are identified as fraudulent. Which data transformation strategy would likely improve the performance of your classifier?

A.Write your data in TFRecords.

B.Z-normalize all the numeric features.

C.Oversample the fraudulent transaction 10 times.

D.Use one-hot encoding on all categorical features.

Answer: C

Explanation:

Reference:
https://towardsdatascience.com/how-to-build-a-machine-learning-model-to-identify-credit-card-fraud-in-5-stepsa-hands-on-modeling-5140b3bd19f1

Question: 61 CertyIQ

You are using transfer learning to train an image classifier based on a pre-trained EfficientNet model. Your training dataset has 20,000 images. You plan to retrain the model once per day. You need to minimize the cost of infrastructure. What platform components and configuration environment should you use?

A.A Deep Learning VM with 4 V100 GPUs and local storage.

B.A Deep Learning VM with 4 V100 GPUs and Cloud Storage.

C.A Google Kubernetes Engine cluster with a V100 GPU Node Pool and an NFS Server

D.An AI Platform Training job using a custom scale tier with 4 V100 GPUs and Cloud Storage

Answer: D

Explanation:

ans: DA, C => local storage, NFS... discarded. Google encourages you to use Cloud Storage.B => could do the job, but here I would focus on the "daily training" thing, because Vertex AI Training jobs are better for this. Also I think that Google usually encourages to use Vertex AI over VMs.

Question: 62 CertyIQ

While conducting an exploratory analysis of a dataset, you discover that categorical feature A has substantial predictive power, but it is sometimes missing. What should you do?

A.Drop feature A if more than 15% of values are missing. Otherwise, use feature A as-is.

B.Compute the mode of feature A and then use it to replace the missing values in feature A.

C.Replace the missing values with the values of the feature with the highest Pearson correlation with feature A.

D.Add an additional class to categorical feature A for missing values. Create a new binary feature that indicates whether feature A is missing.

Answer: D

Explanation:

ans: DA => no, you don't want to drop a feature with high prediction power.B => i think this could confuse the model... a better solution could be to fill missing values using an algorithm like Expectation Maximization, but using the mode i think is a bad idea in this case, because if you have a significant number of missing values (for example >10%) this would modify the "predictive power". you don't want to lose predictive power of a feature, just guide the model to learn when to use that feature and when to ignore it.C => this doesn't make any sense for me. not sure what i would do that.D => i think this could be a really good approach, and i'm pretty sure it would work pretty well a lot of models. the model would learn that when "is_available_feat_A" == True, then it would use the feature A, but whenever it is missing then it would try to use other features.

If our objective was to produce a complete dataset then we might use some average value to fill in the gaps

(option B) but in this case we want to predict an outcome, so inventing our own data is not going to help in my view.Option D is the most sensible approach to let the model choose the best features.

Question: 63 CertyIQ

You work for a large retailer and have been asked to segment your customers by their purchasing habits. The purchase history of all customers has been uploaded to BigQuery. You suspect that there may be several distinct customer segments, however you are unsure of how many, and you don't yet understand the commonalities in their behavior. You want to find the most efficient solution. What should you do?

A.Create a k-means clustering model using BigQuery ML. Allow BigQuery to automatically optimize the number of clusters.

B.Create a new dataset in Dataprep that references your BigQuery table. Use Dataprep to identify similarities within each column.

C.Use the Data Labeling Service to label each customer record in BigQuery. Train a model on your labeled data using AutoML Tables. Review the evaluation metrics to understand whether there is an underlying pattern in the data.

D.Get a list of the customer segments from your company's Marketing team. Use the Data Labeling Service to label each customer record in BigQuery according to the list. Analyze the distribution of labels in your dataset using Data Studio.

Answer: A

Explanation:

I correct myself. It's A:According to the documentation, if you omit the num_clusters option, BigQuery ML will choose a reasonable default based on the total number of rows in the training data.

Ahttps://cloud.google.com/bigquery-ml/docs/kmeans-tutorialhttps://towardsdatascience.com/how-to-use-k-means-clustering-in-bigquery-ml-to-understand-and-describe-your-data-better-c972c6f5733b

Question: 64 CertyIQ

You recently designed and built a custom neural network that uses critical dependencies specific to your organization's framework. You need to train the model using a managed training service on Google Cloud. However, the ML framework and related dependencies are not supported by AI Platform Training. Also, both your model and your data are too large to fit in memory on a single machine. Your ML framework of choice uses the scheduler, workers, and servers distribution structure. What should you do?

A.Use a built-in model available on AI Platform Training.

B.Build your custom container to run jobs on AI Platform Training.

C.Build your custom containers to run distributed training jobs on AI Platform Training.

D.Reconfigure your code to a ML framework with dependencies that are supported by AI Platform Training.

Answer: C

Explanation:

Answer C. By running your machine learning (ML) training job in a custom container, you can use ML frameworks, non-ML dependencies, libraries, and binaries that are not otherwise supported on Vertex AI. Model and your data are too large to fit in memory on a single machine hence distributed training jobs.https://cloud.google.com/vertex-ai/docs/training/containers-overview

ans: CA, D => too much work.B => discarded because "model and your data are too large to fit in memory on a

single machine"

Question: 65 CertyIQ

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.Increase the CPU load

B.Add caching to the pipeline

C.Increase the network bandwidth

D.Add parallel interleave to the pipeline

Answer: D

Explanation:

It's Dhttps://www.tensorflow.org/guide/data_performance

Question: 66 CertyIQ

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.Convert the model to a Keras model, and run a Keras Tuner job.

B.Run a hyperparameter tuning job on AI Platform using custom containers.

C.Create a Kuberflow Pipelines instance, and run a hyperparameter tuning job on Katib.

D.Convert the model to a TensorFlow model, and run a hyperparameter tuning job on AI Platform.

Answer: B

Explanation:

1. B because Vertex AI supports custom models hyperparameter tuning
2. Went with B

Question: 67 CertyIQ

You have a large corpus of written support cases that can be classified into 3 separate categories: Technical Support, Billing Support, or Other Issues. You need to quickly build, test, and deploy a service that will automatically classify future written requests into one of the categories. How should you configure the pipeline?

A.Use the Cloud Natural Language API to obtain metadata to classify the incoming cases.

B.Use AutoML Natural Language to build and test a classifier. Deploy the model as a REST API.

C.Use BigQuery ML to build and test a logistic regression model to classify incoming requests. Use BigQuery ML to perform inference.

D.Create a TensorFlow model using Google's BERT pre-trained model. Build and test a classifier, and deploy the model using Vertex AI.

Answer: B

Explanation:

ans: BA => no, you need customization.C, B => more work and complexityB => AutoML is easier and faster and "you need to quickly build, test, and deploy". Also the REST API part fits our use case.

Question: 68 CertyIQ

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.AutoML Natural Language

B.Cloud Natural Language API

C.AI Hub pre-made Jupyter Notebooks

D.AI Platform Training built-in algorithms

Answer: A

Explanation:

ans: A B => "custom categories" so discarded.C, D => discarded because "without writing code" and "do not have enough data".A => AutoML can train with very little data ("The bare minimum required by AutoML Natural Language for training is 10 text examples per category/label"), as seifou says it will probably use transfer learning behind the scenes.

Question: 69 CertyIQ

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 AutoML to optimize the model's recall in order to minimize false negatives.

B.Use AutoML to optimize the model's F1 score in order to balance the accuracy of false positives and false negatives.

C.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.

D.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.

Answer: D

Explanation:

D becauseIf you increase the frequency of the negative class in your dataset, it will cause the classifier to become more cautious in its predictions and increase its precision for the negative class. This is because the classifier now has more negative examples to learn from and is less likely to make false positive predictions.

Question: 70 CertyIQ

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 with checkpoints.

B.Use AI Platform to run distributed training jobs without checkpoints.

C.Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs with checkpoints.

D.Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs without checkpoints.

Answer: C

Explanation:

https://cloud.google.com/blog/products/ai-machine-learning/reduce-the-costs-of-ml-workflows-with-preemptible-vms-and-gpus?hl=en

It's seem C- https://www.kubeflow.org/docs/distributions/gke/pipelines/preemptible/- https://cloud.google.com/optimization/docs/guide/checkpointing

Question: 71 CertyIQ

You need to train a regression model based on a dataset containing 50,000 records that is stored in BigQuery. The data includes a total of 20 categorical and numerical features with a target variable that can include negative values. You need to minimize effort and training time while maximizing model performance. What approach should you take to train this regression model?

A.Create a custom TensorFlow DNN model

B.Use BQML XGBoost regression to train the model.

C.Use AutoML Tables to train the model without early stopping.

D.Use AutoML Tables to train the model with RMSLE as the optimization objective.

Answer: B

Explanation:

Ans B.C --> No early stopping means longer training timeD --> RMSLE metric need non-negative Y values

B and C is the most likely because of regression approach, But RMSLE it not allow you to take negative label to train as https://cloud.google.com/automl-tables/docs/evaluate#evaluation_metrics_for_regression_models RMSLE: The root-mean-squared logarithmic error metric is similar to RMSE, except that it uses the natural logarithm of the predicted and actual values plus 1. RMSLE penalizes under-prediction more heavily than over-prediction. It can also be a good metric when you don't want to penalize differences for large prediction values more heavily than for small prediction values. This metric ranges from zero to infinity; a lower value indicates a higher quality model. The RMSLE evaluation metric is returned only if all label and predicted values are non-negative.

Question: 72 CertyIQ

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 (PCA) 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: B

Explanation:

L1 regularization it's good for feature selectionhttps://www.quora.com/How-does-the-L1-regularization-method-help-in-feature-selectionhttps://developers.google.com/machine-learning/crash-course/regularization-for-sparsity/l1-regularization

Question: 73 CertyIQ

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

A.Use then TFX ModelValidator tools to specify performance metrics for production readiness.

B.Use k-fold cross-validation as a validation strategy to ensure that your model is ready for production.

C.Use the last relevant week of data as a validation set to ensure that your model is performing accurately on current data.

D.Use the entire dataset and treat the area under the receiver operating characteristics curve (AUC ROC) as the main metric.

Answer: C

Explanation:

Use the last relevant week of data as a validation set to ensure that your model is performing accurately on current data.

Question: 74 CertyIQ

You have deployed a model on Vertex AI for real-time inference. During an online prediction request, you get an “Out of Memory” error. What should you do?

A.Use batch prediction mode instead of online mode.

B.Send the request again with a smaller batch of instances.

C.Use base64 to encode your data before using it for prediction.

D.Apply for a quota increase for the number of prediction requests.

Answer: B

Explanation:

B is the answer429 - Out of Memoryhttps://cloud.google.com/ai-platform/training/docs/troubleshooting

https://cloud.google.com/ai-platform/training/docs/troubleshooting

Question: 75 CertyIQ

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.Train local surrogate models to explain individual predictions.

B.Configure sampled Shapley explanations on Vertex Explainable AI.

C.Configure integrated gradients explanations on Vertex Explainable AI.

D.Measure the effect of each feature as the weight of the feature multiplied by the feature value.

Answer: B

Explanation:

Sampled Shapley works well for these models, which are meta-ensembles of trees and neural networks.https://cloud.google.com/vertex-ai/docs/explainable-ai/overview#sampled-shapley

B is optimal for tabular data Tree or DNNC integrated gradients explanations on Vertex Explainable AI.It is used for image.

Question: 76 CertyIQ

You are working on a classification problem with time series data. After conducting just a few experiments using random cross-validation, you achieved an Area Under the Receiver Operating Characteristic Curve (AUC ROC) value of 99% on the training data. You haven't explored using any sophisticated algorithms or spent any time on hyperparameter tuning. What should your next step be to identify and fix the problem?

A.Address the model overfitting by using a less complex algorithm and use k-fold cross-validation.

B.Address data leakage by applying nested cross-validation during model training.

C.Address data leakage by removing features highly correlated with the target value.

D.Address the model overfitting by tuning the hyperparameters to reduce the AUC ROC value.

Answer: B

Explanation:

Nested cross-validation to reduce data leakage - same as a previous question.

Question: 77 CertyIQ

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.

Answer: A

Explanation:

https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-inference-overviewPredict the label, either a numerical value for regression tasks or a categorical value for classification tasks on DNN regresion

Question: 78 CertyIQ

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.

Answer: B

Explanation:

B.The other options solves nada.

https://towardsdatascience.com/getting-deeper-into-categorical-encodings-for-machine-learning-2312acd347c8When you have millions uniques values try to do: Hash Encoding

Question: 79 CertyIQ

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.

Answer: B

Explanation:

https://developers.google.com/machine-learning/guides/text-classification/step-3#label_vectorization- https://developers.google.com/machine-learning/guides/text-classification/step-4- https://towardsai.net/p/deep-learning/text-classification-with-rnn - https://towardsdatascience.com/pre-trained-word-embedding-for-text-classification-end2end-approach-5fbf5cd8aead

Question: 80 CertyIQ

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 , 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, deploy the database on Firestore for writing and for reading the user's navigation context, 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

Explanation:

keywordsthe inventory is thousands of web banners -> BigtableYou want to Implement the simplest solution -> AI Platform Prediction

Question: 81 CertyIQ

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

Answer: B

Explanation:

The Cloud Compose may be good consideration if you are involved in getting Google Data Engineer CertApp enging is relevant to Dev-Op CertPls.if you know a bit about ML Google Cloud, we are preparing to take Google ML Cert, if there is no specifically particular requirement in the question.We must emphasize on use of Vertext AI as much as possible.

Question: 82 CertyIQ

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.

Answer: C

Explanation:

Option A, preprocess the input CSV file into a TFRecord file, is not as good because it requires additional processing time. Hence, I think C is the best choice.

split data it's best way in my opinion

Question: 83 CertyIQ

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.

Answer: B

Explanation:

B.Online prediction, and need decoupling with Pub/Sub to make it asynchronous. Option A is synchronous.

if you have sensors inyour architecture.. you need pub/sub...

Question: 84 CertyIQ

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.

Answer: B

Explanation:

https://cloud.google.com/blog/topics/developers-practitioners/meet-ais-multitool-vector-embeddingsAnswer B

Currently reading is the keyword here. Going to need B for that, A won't work since it would be based on e.g. all reading history and not the article currently being read.

Question: 85 CertyIQ

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

Answer: D

Explanation:

D- https://cloud.google.com/natural-language/automl/docs/beginners-guide-
https://cloud.google.com/vertex-ai/docs/text-data/classification/evaluate-model

we need to monitor the model, so D

Question: 86 CertyIQ

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.

Answer: D

Explanation:

D- https://cloud.google.com/vertex-ai/docs/ml-metadata/tracking

Question: 87 CertyIQ

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.

Answer: A

Explanation:

B,C,D requires coding. You only have some hours, A is the fastest.

Question: 88 CertyIQ

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.

Answer: A

Explanation:

To access local feature importance in AutoML Tables, you can use the "Explain" feature, which shows the contribution of each feature to the prediction for a specific example. This will help you identify the most important features that contributed to the loan request being rejected.Option B, using the correlation with target values in the data summary page, may not provide the most accurate explanation as it looks at the overall correlation between the features and target variable, rather than the contribution of each feature to a specific prediction.Option C, using the feature importance percentages in the model evaluation page, may not provide a sufficient explanation for the specific prediction, as it shows the importance of each feature across all predictions, rather than for a specific prediction.Option D, varying features independently to identify the threshold per feature that changes the classification, is not recommended as it can be time-consuming and does not provide a clear explanation for why the loan request was rejected

Local, not global since they asked about one specific prediction.Check out that section on this blog: https://cloud.google.com/blog/products/ai-machine-learning/explaining-model-predictions-structured-data/Cool stuff!

Question: 89 CertyIQ

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.

Answer: C

Explanation:

1. Went with C

2. https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/overviewAI Explanations helps you understand your model's outputs for classification and regression tasks. Whenever you request a prediction on AI Platform, AI Explanations tells you how much each feature in the data contributed to the predicted result.

Question: 90 CertyIQ

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

Answer: A

Explanation:

In this scenario, the dataset is highly imbalanced, where most of the examples do not have the company's logo. Therefore, accuracy could be misleading as the model can have high accuracy by simply predicting that all images do not have the logo. F1 score is a good metric to consider in such cases, as it takes both precision and recall into account. However, since the dataset is highly skewed, we should weigh recall more than precision to ensure that the model is correctly identifying the images that do have the logo. Therefore, F-score where recall is weighed more than precision is the best metric to evaluate the performance of the model in this scenario. Option B (RMSE) is not applicable to this classification problem, and option D (F-score where precision is weighed more than recall) is not suitable for highly skewed datasets.

Question: 91 CertyIQ

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.

Answer: C

Explanation:

Must be C. Always feature cross lat and lon on geographical problems. Also, D can not be right as we do not have revenue in the dataset.

Question: 92 CertyIQ

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).

Answer: A

Explanation:

https://cloud.google.com/vision/automl/docs/export-edgeCore ML -> iOS and macOSCoral -> Edge TPU-based deviceTensorFlow.js -> web

Question: 93 CertyIQ

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.

Answer: A

Explanation:

A. Use Vertex AI Workbench user-managed notebooks to generate the report.By using Vertex AI Workbench user-managed notebooks, you can create a one-time report that includes both informative visualizations and sophisticated statistical analyses. The notebooks provide maximum flexibility for data analysis, as they allow you to use a wide range of libraries and tools to create visualizations, perform statistical tests, and share your findings with your team. You can easily connect to the BigQuery table from the notebook and perform the necessary data exploration and analysis.

Question: 94 CertyIQ

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.

Answer: C

Explanation:

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

Reference:

https://www.geeksforgeeks.org/z-score-for-outlier-detection-python/

Question: 95 CertyIQ

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.

Answer: C

Explanation:

Vertex AI Pipelines with Kubeflow Pipelines SDK provides a high-level interface for building end-to-end machine learning pipelines. This approach allows for easy integration with Google Cloud services, including Cloud Storage for data ingestion and preprocessing, Vertex AI for training and hyperparameter tuning, and deployment to an endpoint. The Kubeflow Pipelines SDK also allows for easy orchestration of the entire pipeline, minimizing cluster management.

Answer C...https://cloud.google.com/architecture/ml-on-gcp-best-practices#use-vertex-pipelines

Question: 96 CertyIQ

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.

Answer: B

Explanation:

B. Early stopping would sacrifice model performance. We're also running on a single compute engine machine with, so distributed training is not available to us - right? Only have one single GPU/CPU/worker imo...So throw money at the problem and switch to TPUs, hence B?

We don't have money problems, and we need something that doesn't impair the performance of the model.So I think it's good to change GPU for TPU

Question: 97 CertyIQ

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.

Answer: D

Explanation:

The key is to understand the amount of data that needs to be used for training - the sensor collects tens of millions of records every day and the model needs to use all the data up to the current date. There is a limitation for AutoML is 100M rows -> https://cloud.google.com/vertex-ai/docs/tabular-data/classification-regression/prepare-data

Answer Dhttps://cloud.google.com/blog/products/data-analytics/automl-tables-now-generally-available-bigquery-mlThis legacy version of AutoML Tables is deprecated and will no longer be available on Google Cloud after January 23, 2024. All the functionality of legacy AutoML Tables and new features are available on the Vertex AI platform. See Migrate to Vertex AI to learn how to migrate your resources.

Question: 98 CertyIQ

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.

Answer: C

Explanation:

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

Question: 99 CertyIQ

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.

Answer: D

Explanation:

It is easy to compare via Vertex ML Metadata UI the performance statistics across the different slices and see how the model performance varies over time.

https://cloud.google.com/vertex-ai/docs/ml-metadata/introduction

Question: 100 CertyIQ

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)

Answer: D

Explanation:

DCNN is good for images processing-

https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks

Question: 101 CertyIQ

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.

Answer: D

Explanation:

https://cloud.google.com/tpu/docs/bfloat16

Question: 102 CertyIQ

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.

Answer: B

Explanation:

Continuous retraining (option A) is not necessarily the best solution for preventing prediction drift, as it can be time-consuming and expensive. Instead, monitoring the performance of the model in production is a better approach. Option B is a good choice because it samples a small percentage of incoming predictions and checks for any significant changes in the feature data distribution over a 24-hour period. This allows you to detect any drift and take appropriate action to address it before it affects the model's performance. Options C and D are less effective because they either sample too many or too few predictions and/or at too frequent interv

Question: 103 CertyIQ

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.

Answer: D

Explanation:

Ans D:https://www.tensorflow.org/guide/gpu_performance_analysis for details on how to Optimize the performance on the multi-GPU single host

If distributing the training across multiple GPUs did not result in a decrease in training time, the issue may be related to the batch size being too small. When using multiple GPUs, each GPU gets a smaller portion of the batch size, which can lead to slower training times due to increased communication overhead. Therefore, increasing the batch size can help utilize the GPUs more efficiently and speed up training.

Question: 104 CertyIQ

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.

Answer: B

Explanation:

Since the current model has significant differences in performance across the different languages, it is likely that the translations produced by the Cloud Translation API are not of uniform quality across all languages. Therefore, it would be best to train a classifier using the chat messages in their original language instead of relying on translations.This approach has several advantages. First, the model can directly learn the nuances of each language, leading to better performance across all languages. Second, it eliminates the need for translation, reducing the possibility of errors and improving the overall speed of the system. Finally, it is a relatively simple approach that can be implemented without changing the serving infrastructure.

Answer BSince the performance of the model varies significantly across different languages, it suggests that the translation process might have introduced some noise in the chat messages, making it difficult for the model to generalize across languages. One way to address this issue is to train a classifier using the chat messages in their original language.

Question: 105 CertyIQ

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.

Answer: A

Explanation:

https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-tensorflow

Question: 106 CertyIQ

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 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.

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

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

D.Use Dataprep to transform the state column using a one-hot encoding method, and make each city a column with binary values.

Answer: D

Explanation:

One-hot is a good way to use categorical variables in regressions problemshttps://academic.oup.com/rheumatology/article/54/7/1141/1849688https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-auto-preprocessing

Question: 107 CertyIQ

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

Answer: B

Explanation:

B. Federated learning would be the best learning strategy to train and deploy the ML model for biometric authentication in this scenario. Federated learning allows for training an ML model on distributed data without transferring the raw data to a centralized location.

Question: 108 CertyIQ

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: C

Explanation:

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.

Question: 109 CertyIQ

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.

Answer: B

Explanation:

Blarger learning rates can reduce training time but may lead to model oscillation and may miss the optimal model parameter values.

Question: 110 CertyIQ

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

Answer: B

Explanation:

B"reduce the amount of time spent by quality control inspectors checking for product defects."-> low latency

Question: 111 CertyIQ

You need to build classification workflows over several structured datasets currently stored in BigQuery. Because you will be performing the classification several times, you want to complete the following steps without writing code: exploratory data analysis, feature selection, model building, training, and hyperparameter tuning and serving. What should you do?

A.Train a TensorFlow model on Vertex AI.

B.Train a classification Vertex AutoML model.

C.Run a logistic regression job on BigQuery ML.

D.Use scikit-learn in Notebooks with pandas library.

Answer: B

Explanation:

BQML will need codingonly AutoML in Vertex AI is codeless from end to end

Question: 112 CertyIQ

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.

Answer: A

Explanation:

Syntactic Analysis is not for sentiment analysis

Question: 113 CertyIQ

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,

Answer: A

Explanation:

Previously Google pattern was Pub/Sub -> Dataflow -> BQbut now it looks as there is new Pub/Sub -> BQ

Reference:

https://cloud.google.com/blog/products/data-analytics/pub-sub-launches-direct-path-to-bigquery-for-streaming-analytics

Question: 114 CertyIQ

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

Answer: B

Explanation:

This is B, as it directly relates to our model's ability to predict player ability. There are many factors beyond our model which will impact user engagement (e.g. whether the game is actually enjoyable) so it's not a good measurement of the model performance

The template uses the 'ability' to create teams. For this, we can conclude that the system measures the player's skill. So, nothing better than comparing the predict ability with the actual ability to understand the performance of the model.

Question: 115 CertyIQ

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.

Answer: D

Explanation:

D. Normalize the data by scaling it to have values between 0 and 1.Standardization and normalization are common techniques to preprocess the data to be more suitable for machine learning models. Normalization scales the data to be within a specific range (commonly between 0 and 1 or -1 and 1), which can help prevent features with large magnitudes from dominating the model. This approach is especially useful when using models that are sensitive to the magnitude of features, such as distance-based models or neural networks.

Question: 116 CertyIQ

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

Answer: B

Explanation:

To determine the appropriate hardware for training the models, we need to calculate the required memory and processing power based on the size of the model and the size of the input data.Given that the batch size is 1024 and each example is 1 MB, the total size of each batch is 1024 * 1 MB = 1024 MB = 1 GB. Therefore, we need to load 1 GB of data into memory for each batch.The total size of the network is 20 GB, which means that it can fit in the memory of most modern GPUs.

Question: 117 CertyIQ

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.

Answer: C

Explanation:

This type of model is well-suited to predicting inventory levels because it can take into account trends and patterns in the data over time, such as seasonal fluctuations in demand or changes in customer behavior.

Question: 118 CertyIQ

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

Answer: D

Explanation:

Dyou have built frequent checkpointing into the training process / minimize cost -> preemptible

Preemptible v3-8 TPUs are the most cost-effective option for training large TensorFlow models. They are up to 80% cheaper than non-preemptible v3-8 TPUs, and they are only preempted if Google Cloud needs the resources for other workloads.In this case, the model is long-running and checkpointing is used. This means that the training process can be interrupted and resumed without losing any progress. Therefore, preemptible TPUs are a safe choice, as the training process will not be interrupted if the TPU is preempted.The other options are not as cost-effective.

Question: 119 CertyIQ

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.

Answer: B

Explanation:

B. Machine learning algorithms can learn to identify spam posts based on a wider range of factors, such as the content of the post, the user's behavior, and the context in which the post appears.

Question: 120 CertyIQ

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.

Answer: A

Explanation:

TensorFlow Data Validation (TFDV) is a library that can help you detect and flag anomalies in your dataset, such as changes in the schema or data types.

https://www.tensorflow.org/tfx/data_validation/get_started

Question: 121 CertyIQ

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.

Answer: B

Explanation:

Since you do not have any historical user event data, options C and D are not suitable. In this scenario, it is better to start with a simpler approach, so options A and B are the most suitable. However, option B is preferred because it uses some logic based on content metadata to provide recommendations, which may be more personalized and relevant than presenting videos in alphabetical order. Additionally, collecting user event data from the beginning will help improve the recommendation system in the future.

ans B, you need something easier to implement

Question: 122 CertyIQ

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.

Answer: A

Explanation:

1. Went with A

2. The most likely reason for this result is the model is overfitting in areas with less traffic and underfitting in areas with more traffic.Probably because the model was trained on a dataset that did not have enough examples of congested traffic. As a result, the model is not able to generalise well. When the model is validated on congested traffic, it makes mistakes because it has not seen this type of data before.

Question: 123 CertyIQ

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.

Answer: C

Explanation:

A no - Every row is importantB no - product of other feature values with no values makes no sense to meD no - zero value would bias the model as zero distance from school has the highest value to modelC yes - there is an approach using linear regression to predict missing values

Question: 124 CertyIQ

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.

Answer: B

Explanation:

B. Create the pipeline using TensorFlow Extended (TFX) and standard TFX components. Orchestrate the pipeline using Vertex AI Pipelines.TFX provides a set of standard components for building end-to-end ML pipelines, including data validation and model analysis. Vertex AI Pipelines is a fully managed service for building and orchestrating machine learning pipelines on Google Cloud.

Question: 125 CertyIQ

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-learn, and custom libraries. What should you do?

A.Use the Vertex AI Training to submit training jobs using any framework.

B.Configure Kubeflow to run on Google Kubernetes Engine and submit 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

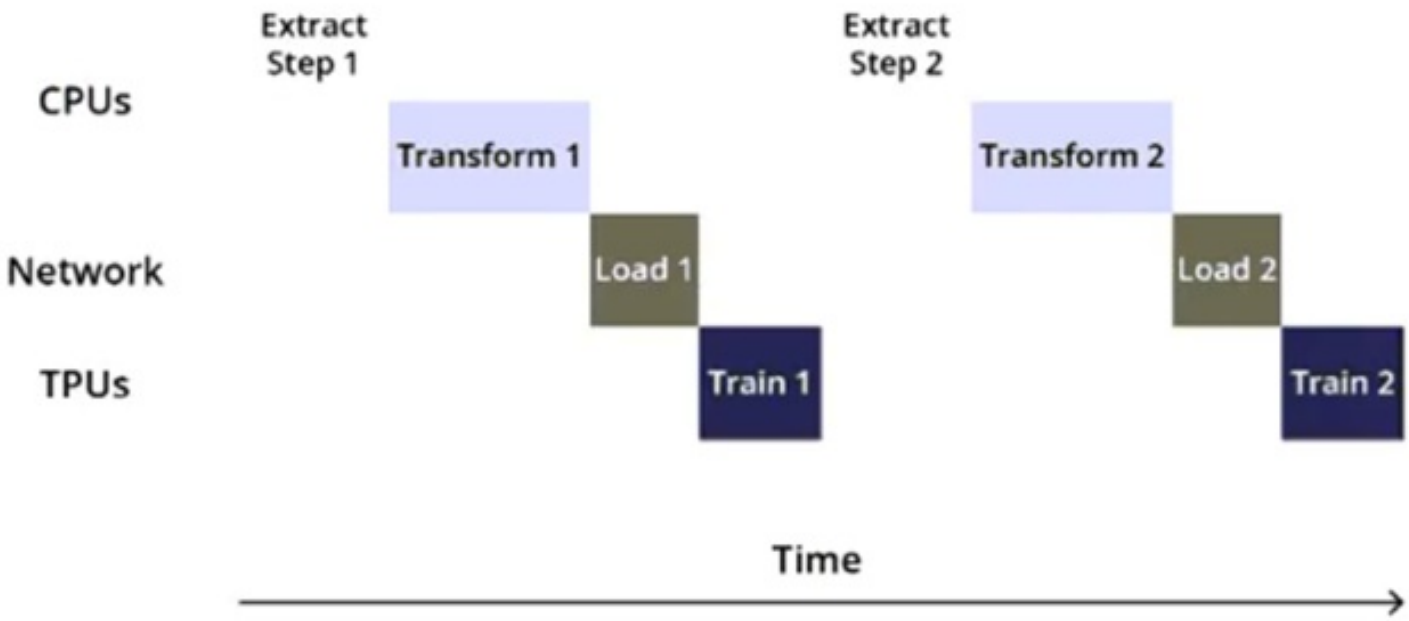
Explanation:

Answer A

Use the Vertex AI Training to submit training jobs using any framework.

Question: 126 CertyIQ

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.

Answer: D

Explanation:

Based on the profile, it appears that the Compute time is relatively low compared to the HostToDevice and DeviceToHost time. This suggests that the data transfer between the host (CPU) and the TPU device is a bottleneck. Therefore, the best action to decrease training time in a cost-efficient way would be to reduce the amount of data transferred between the host and the device.

D- https://www.tensorflow.org/guide/data_performance

Question: 127 CertyIQ

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.

Answer: C

Explanation:

C. Replace the missing values with a placeholder category indicating a missing value.This approach is often referred to as "imputing" missing values, and it is a common technique for dealing with missing data in categorical features. By using a placeholder category, you explicitly indicate that the value is missing, rather than assuming that the missing value is a particular category. This can help to minimize bias in downstream analyses, as it does not introduce any assumptions about the missing data that could bias your results.

When handling missing values in a categorical feature, replacing the missing values with a placeholder category indicating a missing value, as described in option C, is the most appropriate solution in order to minimize bias that could result from the missing values. This approach allows the algorithm to treat missing values as a separate category, avoiding the risk of any assumptions being made about the missing values.Option A, removing the rows with missing values and upsampling the dataset by 5%, can lead to a loss of valuable data and can also introduce bias into the data. This approach can lead to overrepresentation of certain classes and underrepresentation of others.Option B, replacing the missing values with the feature's mean, is not appropriate for categorical features as there is no meaningful average value for categorical features.Option D, moving the rows with missing values to the validation dataset, is not a good solution. This approach may introduce bias into the validation dataset and can lead to overfitting.

Question: 128 CertyIQ

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.

Answer: B

Explanation:

B. Develop an image segmentation ML model to locate the boundaries of the rust spots.An image segmentation model is well-suited for this task because it can identify the exact location and shape of the rust spots in the image, which is critical for determining the severity of the disease. Once the rust spots have been identified, other algorithms can be used to analyze the data and predict the severity of the disease. Object detection models are another option, but they may not be as accurate as image segmentation models when it comes to identifying the exact boundaries of the rust spots. Template matching algorithms using

traditional computer vision libraries are generally not as accurate as ML models when it comes to image analysis.

Question: 129 CertyIQ

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.

Answer: B

Explanation:

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.The reason for this choice is that TFX and Vertex AI Pipelines provide a scalable and cost-effective solution for productionizing machine learning models. TFX is an end-to-end ML platform for building scalable and repeatable ML workflows, while Vertex AI Pipelines provides a fully managed service for orchestrating ML workflows at scale. By using TFX and Vertex AI Pipelines, you can automate the execution of the steps contained in the Jupyter notebook, and schedule the pipeline for weekly retraining. This approach also takes advantage of managed services, which helps to minimize cost.

Question: 130 CertyIQ

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.

Answer: C

Explanation:

C when deploying the model to a Vertex AI endpoint it provides a dedicated prediction service optimised for real-time inference. Vertex AI endpoints are designed for high performance and low latency, making them ideal for real-time prediction use cases. Dataflow can easily invoke the Vertex AI endpoint to perform predictions, minimising serving latency.

Question: 131 CertyIQ

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

Answer: B

Explanation:

Without training a new model' --> B

B- https://www.tensorflow.org/lite/performance/post_training_quantization#dynamic_range_quantization

Question: 132 CertyIQ

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.

Answer: D

Explanation:

D- https://www.tensorflow.org/io/api_docs/python/tfio/bigquery

D. Use TensorFlow I/O's BigQuery Reader to directly read the data.The reason for this choice is that using TensorFlow I/O's BigQuery Reader is the most efficient and scalable option for reading data directly from BigQuery into TensorFlow models. It allows for distributed processing and avoids unnecessary data duplication, which can cause bottlenecks and consume large amounts of storage. Additionally, the BigQuery Reader is optimized for reading data in parallel from BigQuery tables and streaming them directly into TensorFlow. This eliminates the need for any intermediate file formats or data copies, reducing latency and increasing performance.

Question: 133 CertyIQ

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.

Answer: B

Explanation:

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

First we should choos an option with small trials:

"Before starting a job with a large number of trials, you may want to start with a small number of trials to gauge the effect your chosen hyperparameters have on your model's accuracy."

https://cloud.google.com/vertex-ai/docs/training/using-hyperparameter-tuning

https://cloud.google.com/blog/products/gcp/hyperparameter-tuning-on-google-cloud-platform-is-now-faster-and-smarter

Question: 134 CertyIQ

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.

Answer: A

Explanation:

The answer is A. because the Kubeflow Pipelines SDK provides a convenient way to create custom

components from existing Python code using the func_to_container_op function. This allows data science team to encapsulate the custom code as containerised components that can be easily integrated into the kubeflow pipeline. This approach allows for seamless integration of custom Python code into the Kubeflow Pipelines SDK without requiring additional dependencies or infrastructure setup.

A. Use the func_to_container_op function to create custom components from the Python code.The func_to_container_op function in the Kubeflow Pipelines SDK is specifically designed to convert Python functions into containerized components that can be executed in a Kubernetes cluster. By using this function, the Data Science team can easily integrate their custom Python code into the Kubeflow Pipelines SDK without having to learn the details of containerization or Kubernetes.

Question: 135 CertyIQ

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.

Answer: A

Explanation:

https://www.tensorflow.org/tutorials/load_data/images?hl=ja#tfdata_%E3%82%92%E4%BD%BF%E7%94%A8%E3%81%97%E3%81%A6%E3%82%88%E3%82%8A%E7%B2%BE%E5%AF%86%E3%81%AB%E5%88%B6%E5%BE%A1%E3%81%99%E3%82%8B

https://towardsdatascience.com/time-to-choose-tensorflow-data-over-imagedatagenerator-215e594f2435

Question: 136 CertyIQ

You work for an online publisher that delivers news articles to over 50 million readers. You have built an AI model that recommends content for the company's weekly newsletter. A recommendation is considered successful if the article is opened within two days of the newsletter's published date and the user remains on the page for at least one minute.

All the information needed to compute the success metric is available in BigQuery and is updated hourly. The model is trained on eight weeks of data, on average its performance degrades below the acceptable baseline after five weeks, and training time is 12 hours. You want to ensure that the model's performance is above the acceptable baseline while minimizing cost. How should you monitor the model to determine when retraining is necessary?

A.Use Vertex AI Model Monitoring to detect skew of the input features with a sample rate of 100% and a monitoring frequency of two days.

B.Schedule a cron job in Cloud Tasks to retrain the model every week before the newsletter is created.

C.Schedule a weekly query in BigQuery to compute the success metric.

D.Schedule a daily Dataflow job in Cloud Composer to compute the success metric.

Answer: C

Explanation:

Option C is the best answer. Since all the information needed to compute the success metric is available in BigQuery and is updated hourly, scheduling a weekly query in BigQuery to compute the success metric is the simplest and most cost-effective way to monitor the model's performance. By comparing the computed success metric against the acceptable baseline, you can determine when the model's performance has degraded below the threshold, and retrain the model accordingly. This approach avoids the cost of additional monitoring infrastructure and leverages existing data processing capabilities.

https://cloud.google.com/blog/topics/developers-practitioners/continuous-model-evaluation-bigquery-ml-stored-procedures-and-cloud-scheduler

Question: 137 CertyIQ

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.

Answer: D

Explanation:

Option D is the best approach to determine how often to retrain the model while minimizing cost. Running 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 is an effective way to detect when the model's performance has degraded. If skew is detected, the most recent serving data should be sent to the labeling service to evaluate the model's performance. This approach is more cost-effective than sending a subset of requests to the labeling service every month because it only sends data when there is a high probability that the model's performance has degraded. By doing this, the model can be retrained at the right time, and the cost of the labeling service can be minimized.

D https://cloud.google.com/blog/topics/developers-practitioners/monitor-models-training-serving-skew-vertex-aiew-vertex-ai&ved=2ahUKEwiRg_aoj9n8AhWb7TgGHcGCDREQFnoECAwQAQ&usg=AOvVaw197NneIJM0ra7fLq2zsOin

Question: 138 CertyIQ

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.

Answer: D

Explanation:

D as you want to do the prediction before the purchase

D- https://www.tensorflow.org/lite/guide

Question: 139 CertyIQ

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.

Answer: B

Explanation:

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.

https://developers.google.com/machine-learning/guides/rules-of-ml

Question: 140 CertyIQ

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.

Answer: A

Explanation:

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

Question: 141 CertyIQ

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%.

Answer: C

Explanation:

We can consider it as follows reasonably.A: it doesn't make any sense, given that cancel=99% but renew =82%, how did you make renew class (82%) beat the Cancel class(99%), it must be 100% accuracy (bullshit)B: Cancel class have more accuracy than renew (99%>82%)D: You can justify, both are good 80% if we have a balance class.So it left us with C. This model predicts well upon imbalanced class circumstances.target class =10 samles meanwhile the another =90 samples

Question: 142 CertyIQ

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.

Answer: C

Explanation:

This will allow to reuse the same pipeline for different datasets without the need to manually preprocess and transform the data each time.

C. Add a ContainerOp to your pipeline that spins a Dataproc cluster, runs a transformation, and then saves the transformed data in Cloud Storage.The recommended approach to parametrize the model training in Kubeflow Pipelines would be to add a ContainerOp to the pipeline that spins up a Dataproc cluster, runs the PySpark transformation step, and saves the transformed data in Cloud Storage. This approach allows for easy integration of PySpark transformations with Kubeflow Pipelines while taking advantage of the scalability and efficiency of Dataproc.

Question: 143 CertyIQ

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.

Answer: A

Explanation:

A. Deploy the models to a Vertex AI endpoint using the traffic-split=0=80, PREVIOUS_MODEL_ID=20 configuration.The recommended approach to achieve the desired outcome would be to deploy the ML models to a Vertex AI endpoint and configure the traffic splitting using the traffic-split parameter. The traffic-split parameter enables you to split traffic between multiple versions of a model based on a percentage split. In this case, the newest model should receive the majority of the traffic, which can be achieved by setting the traffic-split parameter to 0=80. The previous version of the model should receive the remaining 20% of the traffic, which can be achieved by setting the PREVIOUS_MODEL_ID parameter to 20.

A because traffic can be split across different versions when using Endpoints https://cloud.google.com/vertex-ai/docs/general/deployment#models-endpoint.The --trafic-split flag does exist, but in the question the syntax is incorrect, it should be "--traffic-split = [MODEL_ID_1=value, MODEL_ID_2=value]" as explained in https://cloud.google.com/sdk/gcloud/reference/ai/endpoints/deploy-model

Question: 144 CertyIQ

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.

Answer: C

Explanation:

Answer COption A involves using Google Kubernetes Engine, which is a platform for deploying, managing, and scaling containerized applications. However, it requires more setup time and knowledge of Kubernetes, which might not be ideal for quickly scaling up training workloads. Furthermore, the use of the TensorFlow Job operator seems inappropriate for a PyTorch-based model.

Custom trainer , don't overthink 1000%, this is google recommendation.you don't need Vertex AI Workbench user-managed notebooks,Google Kubernetes Engine, Compute Engine at at all , it is a waste of your efforthttps://cloud.google.com/vertex-ai/docs/training/configure-compute#specifying_gpusYou can choose as your want

Question: 145 CertyIQ

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.

Answer: A

Explanation:

A and B both work well here, but I prefer A since B would imply some minor tradeoff between latency and model accuracy, which isn't the case for A. So I would consider quantization after switching to GPU serving. Can anyone explain why B might be better than A?

Question: 146 CertyIQ

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 Al Workbench user-managed notebooks. Use this data to calculate the descriptive statistics and run the statistical analyses.

B.Spin up a Vertex Al 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 Al 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 Al Workbench user-managed notebooks to run the statistical analyses.

Answer: C

Explanation:

the key here is that it says the dataset would be imported into the notebook for B, therefore no longer utilising BigQuery for calculating the descriptive stats, otherwise I would pick B. Therefore I think C is better. Can anyone find any documentation where Google gives best practice on this? It seems quite subjective

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.BigQuery is a powerful data analysis tool that can handle massive datasets, making it an ideal solution for calculating descriptive statistics for hundreds of millions of records. It can also perform complex statistical tests for hypothesis testing. For time series analysis, using Vertex AI Workbench user-managed notebooks would be the best solution as it provides a flexible environment for data exploration, visualization, and statistical analysis. By using the two tools together, the data science team can efficiently analyze the sales data while minimizing computational resources.Its C not B

Question: 147 CertyIQ

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

A.Use Vertex Al Pipelines to execute the experiments. Query the results stored in MetadataStore using the Vertex Al API.

B.Use Vertex Al Training to execute the experiments. Write the accuracy metrics to BigQuery, and query the results using the BigQuery API.

C.Use Vertex Al Training to execute the experiments. Write the accuracy metrics to Cloud Monitoring, and query the results using the Monitoring API.

D.Use Vertex Al Workbench user-managed notebooks to execute the experiments. Collect the results in a shared Google Sheets file, and query the results using the Google Sheets API.

Answer: A

Explanation:

Vertex AI Pipelines covers everything."Vertex AI Pipelines helps you to automate, monitor, and govern your ML systems by orchestrating your ML workflow in a serverless manner, and storing your workflow's artifacts using Vertex ML Metadata. By storing the artifacts of your ML workflow in Vertex ML Metadata, you can analyze the lineage of your workflow's artifacts — for example, an ML model's lineage may include the training data, hyperparameters, and code that were used to create the model."

Question: 148 CertyIQ

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.

Answer: B

Explanation:

Format Preserving Encryption uses deidentify configuration in which you can specify the param wrapped_key (the encrypted ('wrapped') AES-256 key to use).Answer is B according to me.Ref: https://cloud.google.com/dlp/docs/samples/dlp-deidentify-fpe

Question: 149 CertyIQ

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.

Answer: BD

Explanation:

Option B and D could be good approaches to address the issue.B. Adding an additional objective to penalize the model more for errors made on the minority class can help the model to focus more on correctly

classifying the underrepresented class.D. Upsampling or reweighting the existing training data can help balance the class distribution and increase the model's sensitivity to the underrepresented class.

Question: 150 CertyIQ

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 metric would give you the most confidence in your model?

A.Precision

B.Recall

C.RMSE

D.F1 score

Answer: D

Explanation:

F1 score provides a comprehensive evaluation by penalizing models that excel in just one aspect at the expense of the other. By considering both precision and recall, it helps identify models that effectively balance true positive identification with minimal false positives, making it a more suitable metric for imbalanced data like your logo detection problem.

Question: 151 CertyIQ

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.

Answer: A

Explanation:

Surely removing evaluation metrics downgrades the quality of the evaluation

TFX 0.30 and above adds an interface, with_beam_pipeline_args, for extending the pipeline level beam args per componenttfma.MetricSpec() OOB has recommended metrics; reducing any further might not serve the purpose.

Question: 152 CertyIQ

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

Answer: B

Explanation:

“Rows are selected for a data split randomly, but deterministically. (...) Training a new model with the same training data results in the same data split.” https://cloud.google.com/vertex-ai/docs/tabular-data/data-splits#classification-random. “Randomly redistribute data” [Option B] with different fractions, will result in a different data split. Having a higher fraction split of 70% for the training set will additionally help the model to better generalize (compared to only 50%), thus perform better when testing, the ultimate goal.

Question: 153 CertyIQ

You work for a bank and are building a random forest model for fraud detection. You have a dataset that includes transactions, of which 1% are identified as fraudulent. Which data transformation strategy would likely improve the performance of your classifier?

A.Modify the target variable using the Box-Cox transformation.
B.Z-normalize all the numeric features.
C.Oversample the fraudulent transaction 10 times.
D.Log transform all numeric features.

Answer: C

Explanation:

The answer is C beacause it's the only way to improve model performance.Box-Cox transformation: transform feature values according to normal distributionZ-normalization: transform feature values according to x_new = (x – μ) / σ (so x_new have mean 0 and std dev 1)Log transform: just log transformation Also, the Random Forest algorithm is not a distance-based model but it is a tree-based model, there's no need of normalization process.

Question: 154 CertyIQ

You are developing a classification model to support predictions for your company’s various products. The dataset you were given for model development has class imbalance You need to minimize false positives and false negatives What evaluation metric should you use to properly train the model?

A.F1 score
B.Recall
C.Accuracy
D.Precision

Answer: A

Explanation:

if there wasn't a class imbalance that C. Accuracy would have been the right answer. There A. F1-score which is harmonic mean of precision and recall, that balances the trade-off between precision and recall. It is useful when both false positives and false negatives are important as per the question at hand, and you want to optimize for both.

Question: 155 CertyIQ

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 K80 GPU 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.

Answer: D

Explanation:

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

Question: 156 CertyIQ

You need to build classification workflows over several structured datasets currently stored in BigQuery. Because you will be performing the classification several times, you want to complete the following steps without writing code: exploratory data analysis, feature selection, model building, training, and hyperparameter tuning and serving. What should you do?

A.Train a TensorFlow model on Vertex AI.

B.Train a classification Vertex AutoML model.

C.Run a logistic regression job on BigQuery ML.

D.Use scikit-learn in Vertex AI Workbench user-managed notebooks with pandas library.

Answer: B

Explanation:

1. A and D would require writing code.C. would also imply some “code writing” in BigQuery MLI would go with B.

2. B"without writing code"

Question: 157 CertyIQ

You recently developed a deep learning model. To test your new model, you trained it for a few epochs on a large dataset. You observe that the training and validation losses barely changed during the training run. You want to quickly debug your model. What should you do first?

A.Verify that your model can obtain a low loss on a small subset of the dataset

B.Add handcrafted features to inject your domain knowledge into the model

C.Use the Vertex AI hyperparameter tuning service to identify a better learning rate

D.Use hardware accelerators and train your model for more epochs

Answer: A

Explanation:

the first step to quickly debug the deep learning model is to verify that it can obtain a low loss on a small subset of the dataset (Option A). If the model fails to achieve good results on the smaller subset, further investigation is required to identify and address potential issues with the model.

Question: 158 CertyIQ

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

B.Develop a regression model using BigQuery ML.

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

D.Develop a custom PyTorch regression model, and optimize it using Vertex AI Training.

Answer: B

Explanation:

Develop a regression model using BigQuery ML.

Question: 159 CertyIQ

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: D

Explanation:

1. "Your team has a limited budget and is already overextended"

2. By raising the threshold for comments to be considered toxic or harmful, you will decrease the number of false positives.B is wrong because we are taking a Google MLE exam :) A and C are wrong because both of them involve a good amount of additional work, either for extending the dataset or training/experimenting

with a new model. Considering your team is already over the budget and has too many tasks on their plate (overextended), these two options are not available for you.

Question: 160 CertyIQ

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 Vertex AI. You need to determine which customer attribute has the most predictive power for each prediction served by the model. What should you do?

A.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.

B.Use Vertex Explainable AI. Submit each prediction request with the explain' keyword to retrieve feature attributions using the sampled Shapley method.

C.Use Vertex AI Workbench user-managed notebooks to perform a Lasso regression analysis on your model, which will eliminate features that do not provide a strong signal.

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.

Answer: B

Explanation:

to determine which customer attribute has the most predictive power for each prediction served by the model, you should use Vertex Explainable AI (Option B) with the 'explain' keyword to retrieve feature attributions using the sampled Shapley method. This will give you insights into feature importance at the individual prediction level, allowing you to understand the model's behavior for specific customers.

Question: 161 CertyIQ

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

Explanation:

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

Question: 162 CertyIQ

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 using Vertex AI Training with CPUs.

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

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

Answer: B

Explanation:

B. Train your model using Vertex AI Training with CPUs.No GPUs for ScikitLearn, but parrallelize/distribute training is a good way to increase model building

Question: 163 CertyIQ

You are an ML engineer at a retail company. You have built a model that predicts a coupon to offer an ecommerce customer at checkout based on the items in their cart. When a customer goes to checkout, your serving pipeline, which is hosted on Google Cloud, joins the customer's existing cart with a row in a BigQuery table that contains the customers' historic purchase behavior and uses that as the model's input. The web team is reporting that your model is returning predictions too slowly to load the coupon offer with the rest of the web page. How should you speed up your model's predictions?

A.Attach an NVIDIA P100 GPU to your deployed model's instance.

B.Use a low latency database for the customers' historic purchase behavior.

C.Deploy your model to more instances behind a load balancer to distribute traffic.

D.Create a materialized view in BigQuery with the necessary data for predictions.

Answer: D

Explanation:

D. Create a materialized view in BigQuery with the necessary data for predictions.

Current bottleneck: Joining the cart data with the BigQuery table containing historic purchases likely creates the latency bottleneck. Fetching data from BigQuery on every prediction request can be slow.

Materialized view: A materialized view pre-computes and stores the join between the cart data and the relevant historic purchase information in BigQuery. This eliminates the need for real-time joins during prediction, significantly reducing latency.

Faster access: The pre-computed data in the materialized view is readily available within BigQuery, ensuring faster access for your serving pipeline when predicting the coupon offer.

Lower cost: Compared to additional instances or GPU resources, a materialized view can be a more cost-effective solution, especially if prediction requests are frequent.

Question: 164 CertyIQ

You work for a small company that has deployed an ML model with autoscaling on Vertex AI to serve online predictions in a production environment. The current model receives about 20 prediction requests per hour with an

average response time of one second. You have retrained the same model on a new batch of data, and now you are canary testing it, sending ~10% of production traffic to the new model. During this canary test, you notice that prediction requests for your new model are taking between 30 and 180 seconds to complete. What should you do?

A.Submit a request to raise your project quota to ensure that multiple prediction services can run concurrently.

B.Turn off auto-scaling for the online prediction service of your new model. Use manual scaling with one node always available.

C.Remove your new model from the production environment. Compare the new model and existing model codes to identify the cause of the performance bottleneck.

D.Remove your new model from the production environment. For a short trial period, send all incoming prediction requests to BigQuery. Request batch predictions from your new model, and then use the Data Labeling Service to validate your model's performance before promoting it to production.

Answer: C

Explanation:

Remove your new model from the production environment. Compare the new model and existing model codes to identify the cause of the performance bottleneck.

Question: 165 CertyIQ

You want to train an AutoML model to predict house prices by using a small public dataset stored in BigQuery. You need to prepare the data and want to use the simplest, most efficient approach. What should you do?

A.Write a query that preprocesses the data by using BigQuery and creates a new table. Create a Vertex AI managed dataset with the new table as the data source.

B.Use Dataflow to preprocess the data. Write the output in TFRecord format to a Cloud Storage bucket.

C.Write a query that preprocesses the data by using BigQuery. Export the query results as CSV files, and use those files to create a Vertex AI managed dataset.

D.Use a Vertex AI Workbench notebook instance to preprocess the data by using the pandas library. Export the data as CSV files, and use those files to create a Vertex AI managed dataset.

Answer: A

Explanation:

Write a query that preprocesses the data by using BigQuery and creates a new table. Create a Vertex AI managed dataset with the new table as the data source.

Question: 166 CertyIQ

You developed a Vertex AI ML pipeline that consists of preprocessing and training steps and each set of steps runs on a separate custom Docker image. Your organization uses GitHub and GitHub Actions as CI/CD to run unit and integration tests. You need to automate the model retraining workflow so that it can be initiated both manually and when a new version of the code is merged in the main branch. You want to minimize the steps required to build the workflow while also allowing for maximum flexibility. How should you configure the CI/CD workflow?

A.Trigger a Cloud Build workflow to run tests, build custom Docker images, push the images to Artifact Registry, and launch the pipeline in Vertex AI Pipelines.

B.Trigger GitHub Actions to run the tests, launch a job on Cloud Run to build custom Docker images, push the images to Artifact Registry, and launch the pipeline in Vertex AI Pipelines.

C.Trigger GitHub Actions to run the tests, build custom Docker images, push the images to Artifact Registry, and launch the pipeline in Vertex AI Pipelines.

D.Trigger GitHub Actions to run the tests, launch a Cloud Build workflow to build custom Docker images, push the images to Artifact Registry, and launch the pipeline in Vertex AI Pipelines.

Answer: C

Explanation:

Considering the goal of minimizing steps while allowing for flexibility, option C - "Trigger GitHub Actions to run the tests, build custom Docker images, push the images to Artifact Registry, and launch the pipeline in Vertex AI Pipelines" appears to be the most straightforward approach. It leverages GitHub Actions for testing and image building, then directly triggers the Vertex AI Pipelines, simplifying the workflow and reducing unnecessary services involved in the process.

Question: 167 CertyIQ

You are working with a dataset that contains customer transactions. You need to build an ML model to predict customer purchase behavior. You plan to develop the model in BigQuery ML, and export it to Cloud Storage for online prediction. You notice that the input data contains a few categorical features, including product category and payment method. You want to deploy the model as quickly as possible. What should you do?

A.Use the TRANSFORM clause with the ML.ONE_HOT_ENCODER function on the categorical features at model creation and select the categorical and non-categorical features.

B.Use the ML.ONE_HOT_ENCODER function on the categorical features and select the encoded categorical features and non-categorical features as inputs to create your model.

C.Use the CREATE MODEL statement and select the categorical and non-categorical features.

D.Use the ML.MULTI_HOT_ENCODER function on the categorical features, and select the encoded categorical features and non-categorical features as inputs to create your model.

Answer: B

Explanation:

When the TRANSFORM clause is present, only output columns from the TRANSFORM clause are used in training. Any results from query_statement that don't appear in the TRANSFORM clause are ignored. https://cloud.google.com/bigquery/docs/reference/standard-sql/bigqueryml-syntax-create#transformso if you want TRANSFORM then use TRANSFORM for both categorical and non-categorical features

option B - "Use the ML.ONE_HOT_ENCODER function on the categorical features and select the encoded categorical features and non-categorical features as inputs to create your model" seems to be the most appropriate. This approach directly handles the encoding of categorical features using one-hot encoding and selects the necessary features for model creation, ensuring efficient utilization of categorical data in the BigQuery ML model.

Question: 168 CertyIQ

You need to develop an image classification model by using a large dataset that contains labeled images in a Cloud Storage bucket. What should you do?

A.Use Vertex AI Pipelines with the Kubeflow Pipelines SDK to create a pipeline that reads the images from Cloud Storage and trains the model.

B.Use Vertex AI Pipelines with TensorFlow Extended (TFX) to create a pipeline that reads the images from Cloud Storage and trains the model.

C.Import the labeled images as a managed dataset in Vertex AI and use AutoML to train the model.

D.Convert the image dataset to a tabular format using Dataflow Load the data into BigQuery and use BigQuery ML to train the model.

Answer: C

Explanation:

Import the labeled images as a managed dataset in Vertex AI and use AutoML to train the model.

Question: 169 CertyIQ

You are developing a model to detect fraudulent credit card transactions. You need to prioritize detection, because missing even one fraudulent transaction could severely impact the credit card holder. You used AutoML to tram a model on users' profile information and credit card transaction data After training the initial model, you notice that the model is failing to detect many fraudulent transactions. How should you adjust the training parameters in AutoML to improve model performance? (Choose two.)

A.Increase the score threshold

B.Decrease the score threshold.

C.Add more positive examples to the training set

D.Add more negative examples to the training set

E.Reduce the maximum number of node hours for training

Answer: BC

Explanation:

B.Decrease the score threshold.

C.Add more positive examples to the training set.

Question: 170 CertyIQ

You need to deploy a scikit-leam classification model to production. The model must be able to serve requests 24/7, and you expect millions of requests per second to the production application from 8 am to 7 pm. You need to minimize the cost of deployment. What should you do?

A.Deploy an online Vertex AI prediction endpoint. Set the max replica count to 1

B.Deploy an online Vertex AI prediction endpoint. Set the max replica count to 100

C.Deploy an online Vertex AI prediction endpoint with one GPU per replica. Set the max replica count to 1

D.Deploy an online Vertex AI prediction endpoint with one GPU per replica. Set the max replica count to 100

Answer: B

Explanation:

B. Deploy an online Vertex AI prediction endpoint. Set the max replica count to 100:This option provides a higher number of replicas (100) to handle the expected high volume of requests during peak hours. While it might result in increased costs, it provides the necessary scalability to manage the incoming traffic efficiently. During non-peak hours, you can consider scaling down the replicas to reduce costs, as Vertex AI allows dynamic scaling based on demand.

Question: 171 CertyIQ

You work with a team of researchers to develop state-of-the-art algorithms for financial analysis. Your team develops and debugs complex models in TensorFlow. You want to maintain the ease of debugging while also reducing the model training time. How should you set up your training environment?

A.Configure a v3-8 TPU VM. SSH into the VM to train and debug the model.

B.Configure a v3-8 TPU node. Use Cloud Shell to SSH into the Host VM to train and debug the model.

C.Configure a n1 -standard-4 VM with 4 NVIDIA P100 GPUs. SSH into the VM and use ParameterServerStraregv to train the model.

D.Configure a n1-standard-4 VM with 4 NVIDIA P100 GPUs. SSH into the VM and use MultiWorkerMirroredStrategy to train the model.

Answer: D

Explanation:

Option D - "Configure an n1-standard-4 VM with 4 NVIDIA P100 GPUs. SSH into the VM and use MultiWorkerMirroredStrategy to train the model" appears to be more suitable. This setup utilizes NVIDIA P100 GPUs for computational power and employs MultiWorkerMirroredStrategy, which can distribute the workload across GPUs efficiently, potentially reducing training time while maintaining a relatively straightforward environment for debugging.

Question: 172 CertyIQ

You created an ML pipeline with multiple input parameters. You want to investigate the tradeoffs between different parameter combinations. The parameter options are
•Input dataset
•Max tree depth of the boosted tree regressor
•Optimizer learning rate

You need to compare the pipeline performance of the different parameter combinations measured in F1 score, time to train, and model complexity. You want your approach to be reproducible, and track all pipeline runs on the same platform. What should you do?

A.1. Use BigQueryML to create a boosted tree regressor, and use the hyperparameter tuning capability.
2. Configure the hyperparameter syntax to select different input datasets: max tree depths, and optimizer learning rates. Choose the grid search option.

B.1. Create a Vertex AI pipeline with a custom model training job as part of the pipeline. Configure the pipeline's parameters to include those you are investigating.
2. In the custom training step, use the Bayesian optimization method with F1 score as the target to maximize.

C.1. Create a Vertex AI Workbench notebook for each of the different input datasets.
2. In each notebook, run different local training jobs with different combinations of the max tree depth and optimizer learning rate parameters.
3. After each notebook finishes, append the results to a BigQuery table.

D.1. Create an experiment in Vertex AI Experiments.
2. Create a Vertex AI pipeline with a custom model training job as part of the pipeline. Configure the pipeline's parameters to include those you are investigating.
3. Submit multiple runs to the same experiment, using different values for the parameters.

Answer: D

Explanation:

Vertex AI Experiment was created to compare runs.A is incorrect because you can't create a boosted tree using BigQueryML Given the objective of investigating parameter tradeoffs while ensuring reproducibility and tracking, option D - "Create an experiment in Vertex AI Experiments and submit multiple runs to the same

experiment, using different values for the parameters" seems to be the most suitable. This approach provides a structured and trackable environment within Vertex AI Experiments, allowing multiple runs with varied parameters to be monitored for F1 score, training times, and potentially model complexity, enabling a comprehensive analysis of parameter combinations' tradeoffs.

Reference:

https://cloud.google.com/bigquery/docs/bqml-introduction#supported_models

Question: 173 CertyIQ

You received a training-serving skew alert from a Vertex AI Model Monitoring job running in production. You retrained the model with more recent training data, and deployed it back to the Vertex AI endpoint, but you are still receiving the same alert. What should you do?

A.Update the model monitoring job to use a lower sampling rate.

B.Update the model monitoring job to use the more recent training data that was used to retrain the model.

C.Temporarily disable the alert. Enable the alert again after a sufficient amount of new production traffic has passed through the Vertex AI endpoint.

D.Temporarily disable the alert until the model can be retrained again on newer training data. Retrain the model again after a sufficient amount of new production traffic has passed through the Vertex AI endpoint.

Answer: B

Explanation:

B. Update the model monitoring job to use the more recent training data that was used to retrain the model:This option directly aligns the model monitoring with the recently retrained model and ensures that the monitoring job reflects the characteristics of the latest training data.

Question: 174 CertyIQ

You developed a custom model by using Vertex AI to forecast the sales of your company's products based on historical transactional data. You anticipate changes in the feature distributions and the correlations between the features in the near future. You also expect to receive a large volume of prediction requests. You plan to use Vertex AI Model Monitoring for drift detection and you want to minimize the cost. What should you do?

A.Use the features for monitoring. Set a monitoring-frequency value that is higher than the default.

B.Use the features for monitoring. Set a prediction-sampling-rate value that is closer to 1 than 0.

C.Use the features and the feature attributions for monitoring. Set a monitoring-frequency value that is lower than the default.

D.Use the features and the feature attributions for monitoring. Set a prediction-sampling-rate value that is closer to 0 than 1.

Answer: D

Explanation:

Option D - "Use the features and the feature attributions for monitoring. Set a prediction-sampling-rate value that is closer to 0 than 1" seems to be a reasonable choice. This option allows monitoring both features and feature attributions, offering insights into changes in feature importance, while the lower prediction-sampling-rate helps manage costs by monitoring a subset of predictions. It's a trade-off between cost efficiency and the need for effective drift detection

Question: 175 CertyIQ

You have recently trained a scikit-learn model that you plan to deploy on Vertex AI. This model will support both online and batch prediction. You need to preprocess input data for model inference. You want to package the model for deployment while minimizing additional code. What should you do?

A.1. Upload your model to the Vertex AI Model Registry by using a prebuilt scikit-ieam prediction container.
2. Deploy your model to Vertex AI Endpoints, and create a Vertex AI batch prediction job that uses the instanceConfig.instanceType setting to transform your input data.

B.1. Wrap your model in a custom prediction routine (CPR). and build a container image from the CPR local model.
2. Upload your scikit learn model container to Vertex AI Model Registry.
3. Deploy your model to Vertex AI Endpoints, and create a Vertex AI batch prediction job

C.1. Create a custom container for your scikit learn model.
2. Define a custom serving function for your model.
3. Upload your model and custom container to Vertex AI Model Registry.
4. Deploy your model to Vertex AI Endpoints, and create a Vertex AI batch prediction job.

D.1. Create a custom container for your scikit learn model.
2. Upload your model and custom container to Vertex AI Model Registry.
3. Deploy your model to Vertex AI Endpoints, and create a Vertex AI batch prediction job that uses the instanceConfig.instanceType setting to transform your input data.

Answer: B

Explanation:

B - Creating a custom container without CPR adds additional complexity. i.e. write model server write dockerfile and also build and upload image. Where as using a CPR only requires writing a predictor and using vertex SDK to build image.

https://cloud.google.com/vertex-ai/docs/predictions/custom-prediction-routines

Question: 176 CertyIQ

You work for a food product company. Your company's historical sales data is stored in BigQuery.You need to use Vertex AI's custom training service to train multiple TensorFlow models that read the data from BigQuery and predict future sales. You plan to implement a data preprocessing algorithm that performs mm-max scaling and bucketing on a large number of features before you start experimenting with the models. You want to minimize preprocessing time, cost, and development effort. How should you configure this workflow?

A.Write the transformations into Spark that uses the spark-bigquery-connector, and use Dataproc to preprocess the data.

B.Write SQL queries to transform the data in-place in BigQuery.

C.Add the transformations as a preprocessing layer in the TensorFlow models.

D.Create a Dataflow pipeline that uses the BigQueryIO connector to ingest the data, process it, and write it back to BigQuery.

Answer: B

Explanation:

Big Query can do both transformations.

https://cloud.google.com/bigquery/docs/manual-preprocessing#numerical_functions

Question: 177 CertyIQ

You have created a Vertex AI pipeline that includes two steps. The first step preprocesses 10 TB data completes in about 1 hour, and saves the result in a Cloud Storage bucket. The second step uses the processed data to train a model. You need to update the model's code to allow you to test different algorithms. You want to reduce pipeline execution time and cost while also minimizing pipeline changes. What should you do?

A.Add a pipeline parameter and an additional pipeline step. Depending on the parameter value, the pipeline step conducts or skips data preprocessing, and starts model training.

B.Create another pipeline without the preprocessing step, and hardcode the preprocessed Cloud Storage file location for model training.

C.Configure a machine with more CPU and RAM from the compute-optimized machine family for the data preprocessing step.

D.Enable caching for the pipeline job, and disable caching for the model training step.

Answer: D

Explanation:

Not A. Adding a pipeline parameter and new pipeline steps does not minimise pipeline changes.Not C. The idea is not to re-run the preprocessing step at all.Not B. Creating a whole new pipeline implies a significant investment of effort.I opt for D: Enabling caching only for preprocessing job (although it says "pipeline job" in the option, I think that is a typo). Quoting Vertex AI docs: "If there is a matching execution in Vertex ML Metadata, the outputs of that execution are used and the step is skipped. This helps to reduce costs by skipping computations that were completed in a previous pipeline run.

" https://cloud.google.com/vertex-ai/docs/pipelines/configure-caching

Question: 178 CertyIQ

You work for a bank. You have created a custom model to predict whether a loan application should be flagged for human review. The input features are stored in a BigQuery table. The model is performing well, and you plan to deploy it to production. Due to compliance requirements the model must provide explanations for each prediction. You want to add this functionality to your model code with minimal effort and provide explanations that are as accurate as possible. What should you do?

A.Create an AutoML tabular model by using the BigQuery data with integrated Vertex Explainable AI.

B.Create a BigQuery ML deep neural network model and use the ML.EXPLAIN_PREDICT method with the num_integral_steps parameter.

C.Upload the custom model to Vertex AI Model Registry and configure feature-based attribution by using sampled Shapley with input baselines.

D.Update the custom serving container to include sampled Shapley-based explanations in the prediction outputs.

Answer: C

Explanation:

Upload the custom model to Vertex AI Model Registry and configure feature-based attribution by using sampled Shapley with input baselines.

Question: 179 CertyIQ

You recently used XGBoost to train a model in Python that will be used for online serving. Your model prediction service will be called by a backend service implemented in Golang running on a Google Kubernetes Engine (GKE) cluster. Your model requires pre and postprocessing steps. You need to implement the processing steps so that they run at serving time. You want to minimize code changes and infrastructure maintenance, and deploy your model into production as quickly as possible. What should you do?

A.Use FastAPI to implement an HTTP server. Create a Docker image that runs your HTTP server, and deploy it on your organization's GKE cluster.

B.Use FastAPI to implement an HTTP server. Create a Docker image that runs your HTTP server, Upload the image to Vertex AI Model Registry and deploy it to a Vertex AI endpoint.

C.Use the Predictor interface to implement a custom prediction routine. Build the custom container, upload the container to Vertex AI Model Registry and deploy it to a Vertex AI endpoint.

D.Use the XGBoost prebuilt serving container when importing the trained model into Vertex AI. Deploy the model to a Vertex AI endpoint. Work with the backend engineers to implement the pre- and postprocessing steps in the Golang backend service.

Answer: D

Explanation:

Considering the goal of minimizing code changes, infrastructure maintenance, and quickly deploying the model into production, option D seems to be a pragmatic approach. It leverages the prebuilt XGBoost serving container in Vertex AI, providing a managed environment for serving. The pre- and postprocessing steps can be implemented in the Golang backend service, maintaining consistency with the existing Golang implementation and reducing the need for significant code changes.

Question: 180 CertyIQ

You recently deployed a pipeline in Vertex AI Pipelines that trains and pushes a model to a Vertex AI endpoint to serve real-time traffic. You need to continue experimenting and iterating on your pipeline to improve model performance. You plan to use Cloud Build for CI/CD You want to quickly and easily deploy new pipelines into production, and you want to minimize the chance that the new pipeline implementations will break in production. What should you do?

A.Set up a CI/CD pipeline that builds and tests your source code. If the tests are successful, use the Google. Cloud console to upload the built container to Artifact Registry and upload the compiled pipeline to Vertex AI Pipelines.

B.Set up a CI/CD pipeline that builds your source code and then deploys built artifacts into a pre-production environment. Run unit tests in the pre-production environment. If the tests are successful deploy the pipeline to production.

C.Set up a CI/CD pipeline that builds and tests your source code and then deploys built artifacts into a pre-production environment. After a successful pipeline run in the pre-production environment, deploy the pipeline to production.

D.Set up a CI/CD pipeline that builds and tests your source code and then deploys built artifacts into a pre-production environment. After a successful pipeline run in the pre-production environment, rebuild the source code and deploy the artifacts to production.

Answer: C

Explanation:

C. Set up a CI/CD pipeline that builds and tests your source code and then deploys built artifacts into a pre-production environment. After a successful pipeline run in the pre-production environment, deploy the pipeline to production.A - Does not have pre-production environment. B - Unit test is insufficient, there should be a pipeline run.D - (Uncertain) but there's shouldn't be a rebuilding as you have already built and tested

successfully, feels redundant to rebuild.

Question: 181 CertyIQ

You work for a bank with strict data governance requirements. You recently implemented a custom model to detect fraudulent transactions. You want your training code to download internal data by using an API endpoint hosted in your project's network. You need the data to be accessed in the most secure way, while mitigating the risk of data exfiltration. What should you do?

A.Enable VPC Service Controls for peerings, and add Vertex AI to a service perimeter.

B.Create a Cloud Run endpoint as a proxy to the data. Use Identity and Access Management (IAM) authentication to secure access to the endpoint from the training job.

C.Configure VPC Peering with Vertex AI, and specify the network of the training job.

D.Download the data to a Cloud Storage bucket before calling the training job.

Answer: A

Explanation:

Enable VPC Service Controls for peerings, and add Vertex AI to a service perimeter.

Question: 182 CertyIQ

You are deploying a new version of a model to a production Vertex AI endpoint that is serving traffic. You plan to direct all user traffic to the new model. You need to deploy the model with minimal disruption to your application. What should you do?

A.1. Create a new endpoint
2. Create a new model. Set it as the default version. Upload the model to Vertex AI Model Registry
3. Deploy the new model to the new endpoint
4. Update Cloud DNS to point to the new endpoint

B.1. Create a new endpoint
2. Create a new model. Set the parentModel parameter to the model ID of the currently deployed model and set it as the default version. Upload the model to Vertex AI Model Registry
3. Deploy the new model to the new endpoint, and set the new model to 100% of the traffic.

C.1. Create a new model. Set the parentModel parameter to the model ID of the currently deployed model. Upload the model to Vertex AI Model Registry.
2. Deploy the new model to the existing endpoint, and set the new model to 100% of the traffic

D.1. Create a new model. Set it as the default version. Upload the model to Vertex AI Model Registry
2. Deploy the new model to the existing endpoint

Answer: C

Explanation:

Leverages existing endpoint: Using the same endpoint maintains the same endpoint URL, avoiding DNS updates and potential service interruptions.Gradual traffic transition: Vertex AI allows you to gradually shift traffic between model versions, ensuring a smooth transition without impacting users.Clear versioning: Setting parentModel establishes a relationship between the new model and the existing one, aiding in organization and tracking model lineage.

Question: 183 CertyIQ

You are training an ML model on a large dataset. You are using a TPU to accelerate the training process. You notice that the training process is taking longer than expected. You discover that the TPU is not reaching its full capacity. What should you do?

A.Increase the learning rate

B.Increase the number of epochs

C.Decrease the learning rate

D.Increase the batch size

Answer: D

Explanation:

D, the bigger the batch size, the more resource is taken up.

Question: 184 CertyIQ

You work for a retail company. You have a managed tabular dataset in Vertex AI that contains sales data from three different stores. The dataset includes several features, such as store name and sale timestamp. You want to use the data to train a model that makes sales predictions for a new store that will open soon. You need to split the data between the training, validation, and test sets. What approach should you use to split the data?

A.Use Vertex AI manual split, using the store name feature to assign one store for each set

B.Use Vertex AI default data split

C.Use Vertex AI chronological split, and specify the sales timestamp feature as the time variable

D.Use Vertex AI random split, assigning 70% of the rows to the training set, 10% to the validation set, and 20% to the test set

Answer: C

Explanation:

Use Vertex AI chronological split, and specify the sales timestamp feature as the time variable.

Question: 185 CertyIQ

You have developed a BigQuery ML model that predicts customer churn, and deployed the model to Vertex AI Endpoints. You want to automate the retraining of your model by using minimal additional code when model feature values change. You also want to minimize the number of times that your model is retrained to reduce training costs. What should you do?

A.1 Enable request-response logging on Vertex AI Endpoints
2. Schedule a TensorFlow Data Validation job to monitor prediction drift
3. Execute model retraining if there is significant distance between the distributions

B.1. Enable request-response logging on Vertex AI Endpoints
2. Schedule a TensorFlow Data Validation job to monitor training/serving skew
3. Execute model retraining if there is significant distance between the distributions

C.1. Create a Vertex AI Model Monitoring job configured to monitor prediction drift
2. Configure alert monitoring to publish a message to a Pub/Sub queue when a monitoring alert is detected
3. Use a Cloud Function to monitor the Pub/Sub queue, and trigger retraining in BigQuery

D.1. Create a Vertex AI Model Monitoring job configured to monitor training/serving skew
2. Configure alert monitoring to publish a message to a Pub/Sub queue when a monitoring alert is detected

3. Use a Cloud Function to monitor the Pub/Sub queue, and trigger retraining in BigQuery

Answer: D

Explanation:

I would avoid using TensorFlow validation to minimize code written. That leaves us with options C and D. Now, since it is the values of the features that we want to flag and not the value of the predictions, this sounds more like training-serving skew situation than prediction drift. D.

https://www.evidentlyai.com/blog/machine-learning-monitoring-data-and-concept-drift

Question: 186 CertyIQ

You have been tasked with deploying prototype code to production. The feature engineering code is in PySpark and runs on Dataproc Serverless. The model training is executed by using a Vertex AI custom training job. The two steps are not connected, and the model training must currently be run manually after the feature engineering step finishes. You need to create a scalable and maintainable production process that runs end-to-end and tracks the connections between steps. What should you do?

A.Create a Vertex AI Workbench notebook. Use the notebook to submit the Dataproc Serverless feature engineering job. Use the same notebook to submit the custom model training job. Run the notebook cells sequentially to tie the steps together end-to-end.

B.Create a Vertex AI Workbench notebook. Initiate an Apache Spark context in the notebook and run the PySpark feature engineering code. Use the same notebook to run the custom model training job in TensorFlow. Run the notebook cells sequentially to tie the steps together end-to-end.

C.Use the Kubeflow pipelines SDK to write code that specifies two components:
- The first is a Dataproc Serverless component that launches the feature engineering job
- The second is a custom component wrapped in the create_custom_training_job_from_component utility that launches the custom model training job
Create a Vertex AI Pipelines job to link and run both components

D.Use the Kubeflow pipelines SDK to write code that specifies two components
- The first component initiates an Apache Spark context that runs the PySpark feature engineering code
- The second component runs the TensorFlow custom model training code
Create a Vertex AI Pipelines job to link and run both components.

Answer: C

Explanation:

By using Kubeflow Pipelines, you establish a structured, scalable, and maintainable production process for end-to-end model development and deployment, ensuring proper orchestration, tracking, and integration with the chosen services.

Question: 187 CertyIQ

You recently deployed a scikit-learn model to a Vertex AI endpoint. You are now testing the model on live production traffic. While monitoring the endpoint, you discover twice as many requests per hour than expected throughout the day. You want the endpoint to efficiently scale when the demand increases in the future to prevent users from experiencing high latency. What should you do?

A.Deploy two models to the same endpoint, and distribute requests among them evenly

B.Configure an appropriate minReplicaCount value based on expected baseline traffic

C.Set the target utilization percentage in the autoscailngMetricSpecs configuration to a higher value

D.Change the model's machine type to one that utilizes GPUs

Answer: B

Explanation:

Configure an appropriate minReplicaCount value based on expected baseline traffic.

Question: 188 CertyIQ

You work at a bank. You have a custom tabular ML model that was provided by the bank's vendor. The training data is not available due to its sensitivity. The model is packaged as a Vertex AI Model serving container, which accepts a string as input for each prediction instance. In each string, the feature values are separated by commas. You want to deploy this model to production for online predictions and monitor the feature distribution over time with minimal effort. What should you do?

A.1. Upload the model to Vertex AI Model Registry, and deploy the model to a Vertex AI endpoint
2. Create a Vertex AI Model Monitoring job with feature drift detection as the monitoring objective, and provide an instance schema

B.1. Upload the model to Vertex AI Model Registry, and deploy the model to a Vertex AI endpoint
2. Create a Vertex AI Model Monitoring job with feature skew detection as the monitoring objective, and provide an instance schema

C.1. Refactor the serving container to accept key-value pairs as input format
2. Upload the model to Vertex AI Model Registry, and deploy the model to a Vertex AI endpoint
3. Create a Vertex AI Model Monitoring job with feature drift detection as the monitoring objective.

D.1. Refactor the serving container to accept key-value pairs as input format
2. Upload the model to Vertex AI Model Registry, and deploy the model to a Vertex AI endpoint
3. Create a Vertex AI Model Monitoring job with feature skew detection as the monitoring objective

Answer: A

Explanation:

Handles string input format: Vertex AI Model Monitoring can parse comma-separated feature values, avoiding the need to refactor the serving container. It directly monitors feature distribution over time, aligning with the goal of detecting potential drifts.

Question: 189 CertyIQ

You are implementing a batch inference ML pipeline in Google Cloud. The model was developed using TensorFlow and is stored in SavedModel format in Cloud Storage. You need to apply the model to a historical dataset containing 10 TB of data that is stored in a BigQuery table. How should you perform the inference?

A.Export the historical data to Cloud Storage in Avro format. Configure a Vertex AI batch prediction job to generate predictions for the exported data

B.Import the TensorFlow model by using the CREATE MODEL statement in BigQuery ML. Apply the historical data to the TensorFlow model

C.Export the historical data to Cloud Storage in CSV format. Configure a Vertex AI batch prediction job to generate predictions for the exported data

D.Configure a Vertex AI batch prediction job to apply the model to the historical data in BigQuery

Answer: D

Explanation:

Limitations of other options:A and C. Exporting data: Exporting 10 TB of data to Cloud Storage incurs additional storage costs, transfer time, and potential data management complexities.B. BigQuery ML: While BigQuery ML supports some TensorFlow models, it might have limitations with certain model architectures or features. Additionally, it might not be as optimized for large-scale batch inference as Vertex AI.

Question: 190 CertyIQ

You recently deployed a model to a Vertex AI endpoint. Your data drifts frequently, so you have enabled request-response logging and created a Vertex AI Model Monitoring job. You have observed that your model is receiving higher traffic than expected. You need to reduce the model monitoring cost while continuing to quickly detect drift. What should you do?

A.Replace the monitoring job with a DataFlow pipeline that uses TensorFlow Data Validation (TFDV)

B.Replace the monitoring job with a custom SQL script to calculate statistics on the features and predictions in BigQuery

C.Decrease the sample_rate parameter in the RandomSampleConfig of the monitoring job

D.Increase the monitor_interval parameter in the ScheduleConfig of the monitoring job

Answer: C

Explanation:

Decrease the sample_rate parameter in the RandomSampleConfig of the monitoring job.

Reference:

https://cloud.google.com/vertex-ai/docs/model-monitoring/overview#considerations

Question: 191 CertyIQ

You work for a retail company. You have created a Vertex AI forecast model that produces monthly item sales predictions. You want to quickly create a report that will help to explain how the model calculates the predictions. You have one month of recent actual sales data that was not included in the training dataset. How should you generate data for your report?

A.Create a batch prediction job by using the actual sales data. Compare the predictions to the actuals in the report.

B.Create a batch prediction job by using the actual sales data, and configure the job settings to generate feature attributions. Compare the results in the report.

C.Generate counterfactual examples by using the actual sales data. Create a batch prediction job using the actual sales data and the counterfactual examples. Compare the results in the report.

D.Train another model by using the same training dataset as the original, and exclude some columns. Using the actual sales data create one batch prediction job by using the new model and another one with the original model. Compare the two sets of predictions in the report.

Answer: B

Explanation:

Feature attributions explicitly measure how much each input feature contributed to each prediction, providing the most relevant insights for understanding model behavior.

Question: 192 CertyIQ

Your team has a model deployed to a Vertex AI endpoint. You have created a Vertex AI pipeline that automates the model training process and is triggered by a Cloud Function. You need to prioritize keeping the model up-to-date, but also minimize retraining costs. How should you configure retraining?

A.Configure Pub/Sub to call the Cloud Function when a sufficient amount of new data becomes available

B.Configure a Cloud Scheduler job that calls the Cloud Function at a predetermined frequency that fits your team's budget

C.Enable model monitoring on the Vertex AI endpoint. Configure Pub/Sub to call the Cloud Function when anomalies are detected

D.Enable model monitoring on the Vertex AI endpoint. Configure Pub/Sub to call the Cloud Function when feature drift is detected

Answer: D

Explanation:

Enable model monitoring on the Vertex AI endpoint. Configure Pub/Sub to call the Cloud Function when feature drift is detected.

Question: 193 CertyIQ

Your company stores a large number of audio files of phone calls made to your customer call center in an on-premises database. Each audio file is in wav format and is approximately 5 minutes long. You need to analyze these audio files for customer sentiment. You plan to use the Speech-to-Text API You want to use the most efficient approach. What should you do?

A.1. Upload the audio files to Cloud Storage
2. Call the speech:longrunningrecognize API endpoint to generate transcriptions
3. Call the predict method of an AutoML sentiment analysis model to analyze the transcriptions.

B.1. Upload the audio files to Cloud Storage.
2. Call the speech:longrunningrecognize API endpoint to generate transcriptions
3. Create a Cloud Function that calls the Natural Language API by using the analyzeSentiment method

C.1. Iterate over your local files in Python
2. Use the Speech-to-Text Python library to create a speech.RecognitionAudio object, and set the content to the audio file data
3. Call the speech:recognize API endpoint to generate transcriptions
4. Call the predict method of an AutoML sentiment analysis model to analyze the transcriptions.

D.1. Iterate over your local files in Python
2. Use the Speech-to-Text Python Library to create a speech.RecognitionAudio object and set the content to the audio file data
3. Call the speech:longrunningrecognize API endpoint to generate transcriptions.
4. Call the Natural Language API by using the analyzeSentiment method

Answer: A

Explanation:

Efficient audio processing: speech:longrunningrecognize is specifically designed for handling large audio files, offering asynchronous processing and optimized performance.Scalability: Cloud Storage and Vertex AI AutoML scale seamlessly to handle large volumes of data and model inferences.Cost-effectiveness: Separating transcription and sentiment analysis allows for potential cost optimization by using different pricing models for each service.

Question: 194 CertyIQ

You work for a social media company. You want to create a no-code image classification model for an iOS mobile application to identify fashion accessories. You have a labeled dataset in Cloud Storage. You need to configure a training workflow that minimizes cost and serves predictions with the lowest possible latency. What should you do?

A.Train the model by using AutoML, and register the model in Vertex AI Model Registry. Configure your mobile application to send batch requests during prediction.

B.Train the model by using AutoML Edge, and export it as a Core ML model. Configure your mobile application to use the .mlmodel file directly.

C.Train the model by using AutoML Edge, and export the model as a TFLite model. Configure your mobile application to use the .tflite file directly.

D.Train the model by using AutoML, and expose the model as a Vertex AI endpoint. Configure your mobile application to invoke the endpoint during prediction.

Answer: B

Explanation:

Train the model by using AutoML Edge, and export it as a Core ML model. Configure your mobile application to use the .mlmodel file directly.

Reference:

https://cloud.google.com/vertex-ai/docs/export/export-edge-model#classification

Question: 195 CertyIQ

You work for a retail company. You have been asked to develop a model to predict whether a customer will purchase a product on a given day. Your team has processed the company's sales data, and created a table with the following rows:
•Customer_id
•Product_id
•Date
•Days_since_last_purchase (measured in days)
•Average_purchase_frequency (measured in 1/days)
•Purchase (binary class, if customer purchased product on the Date)

You need to interpret your model's results for each individual prediction. What should you do?

A.Create a BigQuery table. Use BigQuery ML to build a boosted tree classifier. Inspect the partition rules of the trees to understand how each prediction flows through the trees.

B.Create a Vertex AI tabular dataset. Train an AutoML model to predict customer purchases. Deploy the model to a Vertex AI endpoint and enable feature attributions. Use the "explain" method to get feature attribution values for each individual prediction.

C.Create a BigQuery table. Use BigQuery ML to build a logistic regression classification model. Use the values of the coefficients of the model to interpret the feature importance, with higher values corresponding to more importance

D.Create a Vertex AI tabular dataset. Train an AutoML model to predict customer purchases. Deploy the model to a Vertex AI endpoint. At each prediction, enable L1 regularization to detect non-informative features.

Answer: B

Explanation:

Individual prediction interpretability: Feature attributions specifically address the need to understand how features contribute to individual predictions, providing fine-grained insights.Vertex AI integration: Vertex AI offers seamless integration of feature attributions with AutoML models, simplifying the process.Model flexibility: AutoML can explore various model architectures, potentially finding the most suitable one for this task, while still providing interpretability.

Question: 196 CertyIQ

You work for a company that captures live video footage of checkout areas in their retail stores. You need to use the live video footage to build a model to detect the number of customers waiting for service in near real time. You want to implement a solution quickly and with minimal effort. How should you build the model?

A.Use the Vertex AI Vision Occupancy Analytics model.

B.Use the Vertex AI Vision Person/vehicle detector model.

C.Train an AutoML object detection model on an annotated dataset by using Vertex AutoML.

D.Train a Seq2Seq+ object detection model on an annotated dataset by using Vertex AutoML.

Answer: A

Explanation:

A. Use the Vertex AI Vision Occupancy Analytics model.

Pre-trained and optimized: Occupancy Analytics is a pre-trained and optimized model specifically designed for counting people in video footage, aligning perfectly with your task. This eliminates the need for extensive data collection, annotation, and training, saving time and effort.

Near real-time performance: The model is designed for low latency and near real-time inference, providing results quickly with minimal delay, important for live video analysis.

Minimal configuration: Compared to training your own model, this option requires minimal configuration within the Vertex AI console, allowing for a quicker setup and deployment.

Reference:

https://codelabs.developers.google.com/vertex-ai-vision-queue-detection#0

Question: 197 CertyIQ

You work as an analyst at a large banking firm. You are developing a robust scalable ML pipeline to tram several regression and classification models. Your primary focus for the pipeline is model interpretability. You want to productionize the pipeline as quickly as possible. What should you do?

A.Use Tabular Workflow for Wide & Deep through Vertex AI Pipelines to jointly train wide linear models and deep neural networks

B.Use Google Kubernetes Engine to build a custom training pipeline for XGBoost-based models

C.Use Tabular Workflow for TabNet through Vertex AI Pipelines to train attention-based models

D.Use Cloud Composer to build the training pipelines for custom deep learning-based models

Answer: C

Explanation:

Use Tabular Workflow for TabNet through Vertex AI Pipelines to train attention-based models.

Question: 198 CertyIQ

You developed a Transformer model in TensorFlow to translate text. Your training data includes millions of documents in a Cloud Storage bucket. You plan to use distributed training to reduce training time. You need to configure the training job while minimizing the effort required to modify code and to manage the cluster's configuration. What should you do?

A.Create a Vertex AI custom training job with GPU accelerators for the second worker pool. Use tf.distribute.MultiWorkerMirroredStrategy for distribution.

B.Create a Vertex AI custom distributed training job with Reduction Server. Use N1 high-memory machine type instances for the first and second pools, and use N1 high-CPU machine type instances for the third worker pool.

C.Create a training job that uses Cloud TPU VMs. Use tf.distribute.TPUStrategy for distribution.

D.Create a Vertex AI custom training job with a single worker pool of A2 GPU machine type instances. Use tf.distribute.MirroredStrategv for distribution.

Answer: A

Explanation:

Minimizes code modification: MultiWorkerMirroredStrategy often requires minimal code changes to distribute training across multiple workers, aligning with the goal of minimizing effort.Simplifies cluster management: Vertex AI handles cluster configuration and scaling for custom training jobs, reducing the need for manual management.Effective distributed training: MultiWorkerMirroredStrategy is well-suited for large models and datasets, efficiently distributing training across GPUs.

Question: 199 CertyIQ

You are developing a process for training and running your custom model in production. You need to be able to show lineage for your model and predictions. What should you do?

A.1. Create a Vertex AI managed dataset.
2. Use a Vertex AI training pipeline to train your model.
3. Generate batch predictions in Vertex AI.

B.1. Use a Vertex AI Pipelines custom training job component to tram your model.
2. Generate predictions by using a Vertex AI Pipelines model batch predict component.

C.1. Upload your dataset to BigQuery.
2. Use a Vertex AI custom training job to train your model.
3. Generate predictions by using Vertex Al SDK custom prediction routines.

D.1. Use Vertex AI Experiments to train your model.
2. Register your model in Vertex AI Model Registry.
3. Generate batch predictions in Vertex AI.

Answer: D

Explanation:

1. Use Vertex AI Experiments to train your model.

2. Register your model in Vertex AI Model Registry.

3. Generate batch predictions in Vertex AI.

Question: 200 CertyIQ

You work for a hotel and have a dataset that contains customers' written comments scanned from paper-based customer feedback forms, which are stored as PDF files. Every form has the same layout. You need to quickly predict an overall satisfaction score from the customer comments on each form. How should you accomplish this task?

A.Use the Vision API to parse the text from each PDF file. Use the Natural Language API analyzeSentiment feature to infer overall satisfaction scores.

B.Use the Vision API to parse the text from each PDF file. Use the Natural Language API analyzeEntitySentiment feature to infer overall satisfaction scores.

C.Uptrain a Document AI custom extractor to parse the text in the comments section of each PDF file. Use the Natural Language API analyzeSentiment feature to infer overall satisfaction scores.

D.Uptrain a Document AI custom extractor to parse the text in the comments section of each PDF file. Use the Natural Language API analyzeEntitySentiment feature to infer overall satisfaction scores.

Answer: C

Explanation:

Precision in text extraction: Document AI is specifically designed for extracting text from structured documents like forms, ensuring accurate extraction of comments, even with varying handwriting styles.

Custom model for form layout: Training a custom extractor tailored to the hotel's feedback form layout further enhances accuracy and targets the relevant comments section effectively.

Sentiment analysis: Natural Language API's analyzeSentiment feature analyzes overall sentiment in a text block, aligning with the goal of deriving overall satisfaction scores.

Question: 201 CertyIQ

You developed a Vertex AI pipeline that trains a classification model on data stored in a large BigQuery table. The pipeline has four steps, where each step is created by a Python function that uses the KubeFlow v2 API. The components have the following names:

```
dt=datetime.now().strftime("%Y%m%d%H%M%S")
f"export-{dt}.yaml", f"preprocess-{dt}.yaml", f"train-{dt}.yaml",
f"calibrate-{dt}.yaml"
```

You launch your Vertex AI pipeline as the following:

```
job = aip.PipelineJob(
    display_name="my-awesome-pipeline",
    template_path="pipeline.json",
    job_id=f"my-awesome-pipeline-{dt}",
    parameter_values=params,
    enable_caching=True,
    location="europe-west1"
)
```

You perform many model iterations by adjusting the code and parameters of the training step. You observe high costs associated with the development, particularly the data export and preprocessing steps. You need to reduce model development costs. What should you do?

A.Change the components' YAML filenames to export.yaml, preprocess,yaml, f "train- dt .yaml", f"calibrate- dt).vaml".

B.Add the "kubeflow.v1.caching": True parameter to the set of params provided to your PipelineJob.

C.Move the first step of your pipeline to a separate step, and provide a cached path to Cloud Storage as an input to the main pipeline.

D.Change the name of the pipeline to f"my-awesome-pipeline- dt ".

Answer: A

Explanation:

Change the components' YAML filenames to export.yaml, preprocess,yaml, f "train- dt .yaml", f"calibrate- dt).vaml".

Question: 202 CertyIQ

You work for a startup that has multiple data science workloads. Your compute infrastructure is currently on-premises, and the data science workloads are native to PySpark. Your team plans to migrate their data science workloads to Google Cloud. You need to build a proof of concept to migrate one data science job to Google Cloud. You want to propose a migration process that requires minimal cost and effort. What should you do first?

A.Create a n2-standard-4 VM instance and install Java, Scala, and Apache Spark dependencies on it.

B.Create a Google Kubernetes Engine cluster with a basic node pool configuration, install Java, Scala, and Apache Spark dependencies on it.

C.Create a Standard (1 master, 3 workers) Dataproc cluster, and run a Vertex AI Workbench notebook instance on it.

D.Create a Vertex AI Workbench notebook with instance type n2-standard-4.

Answer: D

Explanation:

Minimal setup: Vertex AI Workbench notebooks come pre-configured with PySpark and other data science tools, eliminating the need for manual installation and setup.Cost-effectiveness: Vertex AI Workbench offers managed notebooks with pay-as-you-go pricing, making it a cost-efficient option for proof-of-concept testing.Ease of use: Data scientists can directly run PySpark code in the notebook without managing

infrastructure, streamlining the migration process.Scalability: Vertex AI Workbench can easily scale to handle larger workloads or multiple users if the proof-of-concept is successful.

Question: 203 CertyIQ

You work for a bank. You have been asked to develop an ML model that will support loan application decisions. You need to determine which Vertex AI services to include in the workflow. You want to track the model's training parameters and the metrics per training epoch. You plan to compare the performance of each version of the model to determine the best model based on your chosen metrics. Which Vertex AI services should you use?

A.Vertex ML Metadata, Vertex AI Feature Store, and Vertex AI Vizier

B.Vertex AI Pipelines, Vertex AI Experiments, and Vertex AI Vizier

C.Vertex ML Metadata, Vertex AI Experiments, and Vertex AI TensorBoard

D.Vertex AI Pipelines, Vertex AI Feature Store, and Vertex AI TensorBoard

Answer: C

Explanation:

Vertex ML Metadata:Tracks model training parameters, hyperparameters, metrics, and lineage information.Stores metadata in a central repository for easy access and comparison.Integrates seamlessly with Vertex AI Experiments and TensorBoard.Vertex AI Experiments:Organizes and manages model training runs as experiments.Visualizes experiment results, including metrics and parameter comparisons.Facilitates tracking of the best performing model versions.Vertex AI TensorBoard:Provides detailed visualizations of training metrics and model performance.Enables analysis of model behavior at each training epoch.Integrates with Vertex AI Experiments for seamless access to experiment data.

Question: 204 CertyIQ

You work for an auto insurance company. You are preparing a proof-of-concept ML application that uses images of damaged vehicles to infer damaged parts. Your team has assembled a set of annotated images from damage claim documents in the company's database. The annotations associated with each image consist of a bounding box for each identified damaged part and the part name. You have been given a sufficient budget to train models on Google Cloud. You need to quickly create an initial model. What should you do?

A.Download a pre-trained object detection model from TensorFlow Hub. Fine-tune the model in Vertex AI Workbench by using the annotated image data.

B.Train an object detection model in AutoML by using the annotated image data.

C.Create a pipeline in Vertex AI Pipelines and configure the AutoMLTrainingJobRunOp component to train a custom object detection model by using the annotated image data.

D.Train an object detection model in Vertex AI custom training by using the annotated image data.

Answer: B

Explanation:

Speed: AutoML excels in creating high-quality models with minimal code and setup, significantly accelerating model development.Ease of use: It provides a user-friendly interface and automates many aspects of model training, making it accessible even for those without extensive ML expertise.Automatic optimization: AutoML automatically handles hyperparameter tuning, feature engineering, and architecture selection, reducing manual effort and expertise required.Custom object detection: It supports custom object detection tasks, directly addressing the need to identify damaged parts in images.

Question: 205 CertyIQ

You are analyzing customer data for a healthcare organization that is stored in Cloud Storage. The data contains personally identifiable information (PII). You need to perform data exploration and preprocessing while ensuring the security and privacy of sensitive fields. What should you do?

A.Use the Cloud Data Loss Prevention (DLP) API to de-identify the PII before performing data exploration and preprocessing.

B.Use customer-managed encryption keys (CMEK) to encrypt the PII data at rest, and decrypt the PII data during data exploration and preprocessing.

C.Use a VM inside a VPC Service Controls security perimeter to perform data exploration and preprocessing.

D.Use Google-managed encryption keys to encrypt the PII data at rest, and decrypt the PII data during data exploration and preprocessing.

Answer: A

Explanation:

Minimizes exposure of sensitive data: De-identification replaces or removes sensitive information, reducing the risk of accidental exposure or unauthorized access during analysis.Preserves data utility: DLP can de-identify data while maintaining its usefulness for exploration and preprocessing, ensuring meaningful analysis without compromising privacy.Flexibility in de-identification: You can choose appropriate de-identification techniques (e.g., masking, pseudonymization, generalization) based on specific privacy requirements and analysis needs.

Question: 206 CertyIQ

You are building a predictive maintenance model to preemptively detect part defects in bridges. You plan to use high definition images of the bridges as model inputs. You need to explain the output of the model to the relevant stakeholders so they can take appropriate action. How should you build the model?

A.Use scikit-learn to build a tree-based model, and use SHAP values to explain the model output.

B.Use scikit-learn to build a tree-based model, and use partial dependence plots (PDP) to explain the model output.

C.Use TensorFlow to create a deep learning-based model, and use Integrated Gradients to explain the model output.

D.Use TensorFlow to create a deep learning-based model, and use the sampled Shapley method to explain the model output.

Answer: C

Explanation:

Handling image input: Deep learning models excel in processing complex visual data like high-definition images, making them ideal for extracting relevant features from bridge images for defect detection.Explainability with Integrated Gradients: Integrated Gradients is a powerful technique specifically designed to explain the predictions of deep learning models. It attributes model output to specific input features, providing insights into how the model makes decisions.Visualization: Integrated Gradients can generate visual explanations, such as heatmaps, that highlight image regions most influential to predictions, aiding in understanding and trust for stakeholders.

Reference:

https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/overview#compare-methods

Question: 207 CertyIQ

You work for a hospital that wants to optimize how it schedules operations. You need to create a model that uses the relationship between the number of surgeries scheduled and beds used. You want to predict how many beds will be needed for patients each day in advance based on the scheduled surgeries. You have one year of data for the hospital organized in 365 rows.

The data includes the following variables for each day:
•Number of scheduled surgeries
•Number of beds occupied
•Date

You want to maximize the speed of model development and testing. What should you do?

A.Create a BigQuery table. Use BigQuery ML to build a regression model, with number of beds as the target variable, and number of scheduled surgeries and date features (such as day of week) as the predictors.

B.Create a BigQuery table. Use BigQuery ML to build an ARIMA model, with number of beds as the target variable, and date as the time variable.

C.Create a Vertex AI tabular dataset. Train an AutoML regression model, with number of beds as the target variable, and number of scheduled minor surgeries and date features (such as day of the week) as the predictors.

D.Create a Vertex AI tabular dataset. Train a Vertex AI AutoML Forecasting model, with number of beds as the target variable, number of scheduled surgeries as a covariate and date as the time variable.

Answer: D

Explanation:

Create a Vertex AI tabular dataset. Train a Vertex AI AutoML Forecasting model, with number of beds as the target variable, number of scheduled surgeries as a covariate and date as the time variable.

Question: 208 CertyIQ

You recently developed a wide and deep model in TensorFlow. You generated training datasets using a SQL script that preprocessed raw data in BigQuery by performing instance-level transformations of the data. You need to create a training pipeline to retrain the model on a weekly basis. The trained model will be used to generate daily recommendations. You want to minimize model development and training time. How should you develop the training pipeline?

A.Use the Kubeflow Pipelines SDK to implement the pipeline. Use the BigQueryJobOp component to run the preprocessing script and the CustomTrainingJobOp component to launch a Vertex AI training job.

B.Use the Kubeflow Pipelines SDK to implement the pipeline. Use the DataflowPythonJobOp component to preprocess the data and the CustomTrainingJobOp component to launch a Vertex AI training job.

C.Use the TensorFlow Extended SDK to implement the pipeline Use the ExampleGen component with the BigQuery executor to ingest the data the Transform component to preprocess the data, and the Trainer component to launch a Vertex AI training job.

D.Use the TensorFlow Extended SDK to implement the pipeline Implement the preprocessing steps as part of the input_fn of the model. Use the ExampleGen component with the BigQuery executor to ingest the data and the Trainer component to launch a Vertex AI training job.

Answer: D

Explanation:

Addressing Limitations of Other Options:Kubeflow Pipelines (A and B): While Kubeflow offers flexibility, it might require more setup and configuration, potentially increasing development time compared to TFX's integrated approach.Separate Preprocessing (C): Using a separate Transform component for preprocessing can add complexity and potential overheads, especially for instance-level transformations that can often be directly integrated within the model's input pipeline.

Question: 209 CertyIQ

You are training a custom language model for your company using a large dataset. You plan to use the Reduction Server strategy on Vertex AI. You need to configure the worker pools of the distributed training job. What should you do?

A.Configure the machines of the first two worker pools to have GPUs, and to use a container image where your training code runs. Configure the third worker pool to have GPUs, and use the reductionserver container image.

B.Configure the machines of the first two worker pools to have GPUs and to use a container image where your training code runs. Configure the third worker pool to use the reductionserver container image without accelerators, and choose a machine type that prioritizes bandwidth.

C.Configure the machines of the first two worker pools to have TPUs and to use a container image where your training code runs. Configure the third worker pool without accelerators, and use the reductionserver container image without accelerators, and choose a machine type that prioritizes bandwidth.

D.Configure the machines of the first two pools to have TPUs, and to use a container image where your training code runs. Configure the third pool to have TPUs, and use the reductionserver container image.

Answer: B

Explanation:

Configure the machines of the first two worker pools to have GPUs and to use a container image where your training code runs. Configure the third worker pool to use the reduction server container image without accelerators, and choose a machine type that prioritizes bandwidth.

Question: 210 CertyIQ

You have trained a model by using data that was preprocessed in a batch Dataflow pipeline. Your use case requires real-time inference. You want to ensure that the data preprocessing logic is applied consistently between training and serving. What should you do?

A.Perform data validation to ensure that the input data to the pipeline is the same format as the input data to the endpoint.

B.Refactor the transformation code in the batch data pipeline so that it can be used outside of the pipeline. Use the same code in the endpoint.

C.Refactor the transformation code in the batch data pipeline so that it can be used outside of the pipeline. Share this code with the end users of the endpoint.

D.Batch the real-time requests by using a time window and then use the Dataflow pipeline to preprocess the batched requests. Send the preprocessed requests to the endpoint.

Answer: B

Explanation:

A. Data validation: While essential, it doesn't guarantee consistency if the preprocessing logic itself differs between pipeline and endpoint.C. Sharing code with end users: This shifts the preprocessing burden to end users, potentially leading to inconsistencies and errors, and isn't feasible for real-time inference.D. Batching

real-time requests: This introduces latency and might not align with real-time requirements, as users expect immediate responses.

Question: 211 CertyIQ

You need to develop a custom TensorFlow model that will be used for online predictions. The training data is stored in BigQuery You need to apply instance-level data transformations to the data for model training and serving. You want to use the same preprocessing routine during model training and serving. How should you configure the preprocessing routine?

A.Create a BigQuery script to preprocess the data, and write the result to another BigQuery table.

B.Create a pipeline in Vertex AI Pipelines to read the data from BigQuery and preprocess it using a custom preprocessing component.

C.Create a preprocessing function that reads and transforms the data from BigQuery. Create a Vertex AI custom prediction routine that calls the preprocessing function at serving time.

D.Create an Apache Beam pipeline to read the data from BigQuery and preprocess it by using TensorFlow Transform and Dataflow.

Answer: C

Explanation:

Addressing limitations of other options:A. Data validation: While essential, it doesn't guarantee consistency if the preprocessing logic itself differs between pipeline and endpoint.C. Sharing code with end users: This shifts the preprocessing burden to end users, potentially leading to inconsistencies and errors, and isn't feasible for real-time inference.D. Batching real-time requests: This introduces latency and might not align with real-time requirements, as users expect immediate responses.

Question: 212 CertyIQ

You are pre-training a large language model on Google Cloud. This model includes custom TensorFlow operations in the training loop. Model training will use a large batch size, and you expect training to take several weeks. You need to configure a training architecture that minimizes both training time and compute costs. What should you do?

A.Implement 8 workers of a2-megagpu-16g machines by using tf.distribute.MultiWorkerMirroredStrategy.

B.Implement a TPU Pod slice with -accelerator-type=v4-l28 by using tf.distribute.TPUStrategy.

C.Implement 16 workers of c2d-highcpu-32 machines by using tf.distribute.MirroredStrategy.

D.Implement 16 workers of a2-highgpu-8g machines by using tf.distribute.MultiWorkerMirroredStrategy.

Answer: B

Explanation:

TPU Advantages:Highly Specialized: TPUs (Tensor Processing Units) are custom-designed hardware accelerators specifically optimized for machine learning workloads, particularly those involving large batch sizes and matrix-heavy computations, common in large language models.Exceptional Performance: TPUs can significantly outperform CPUs and GPUs in terms of speed and efficiency for these types of tasks.Cost-Effective: While TPUs might have a higher hourly cost, their exceptional performance often leads to lower overall costs due to faster training times and reduced resource usage.TPU Pod Slice:Scalability: TPU Pod slices allow you to distribute training across multiple TPUv4 chips for even greater performance and scalability.Custom Operations: The tf.distribute.TPUStrategy ensures compatibility with custom TensorFlow

operations,

Question: 213 CertyIQ

You are building a TensorFlow text-to-image generative model by using a dataset that contains billions of images with their respective captions. You want to create a low maintenance, automated workflow that reads the data from a Cloud Storage bucket collects statistics, splits the dataset into training/validation/test datasets performs data transformations trains the model using the training/validation datasets, and validates the model by using the test dataset. What should you do?

A.Use the Apache Airflow SDK to create multiple operators that use Dataflow and Vertex AI services. Deploy the workflow on Cloud Composer.

B.Use the MLFlow SDK and deploy it on a Google Kubernetes Engine cluster. Create multiple components that use Dataflow and Vertex AI services.

C.Use the Kubeflow Pipelines (KFP) SDK to create multiple components that use Dataflow and Vertex AI services. Deploy the workflow on Vertex AI Pipelines.

D.Use the TensorFlow Extended (TFX) SDK to create multiple components that use Dataflow and Vertex AI services. Deploy the workflow on Vertex AI Pipelines.

Answer: D

Explanation:

Airflow (A): While versatile, Airflow often requires more manual configuration and integration with ML services, potentially increasing maintenance effort.MLFlow (B): MLFlow focuses on experiment tracking and model management, lacking built-in pipeline components for data processing and model training.Kubeflow Pipelines (C): KFP is flexible but requires more setup and infrastructure management compared to TFX's managed services.

Question: 214 CertyIQ

You are developing an ML pipeline using Vertex AI Pipelines. You want your pipeline to upload a new version of the XGBoost model to Vertex AI Model Registry and deploy it to Vertex AI Endpoints for online inference. You want to use the simplest approach. What should you do?

A.Use the Vertex AI REST API within a custom component based on a vertex-ai/prediction/xgboost-cpu image

B.Use the Vertex AI ModelEvaluationOp component to evaluate the model

C.Use the Vertex AI SDK for Python within a custom component based on a python:3.10 image

D.Chain the Vertex AI ModelUploadOp and ModelDeployOp components together

Answer: D

Explanation:

A. Custom Component with REST API: This involves more manual coding and understanding of REST API endpoints, potentially increasing complexity and maintenance.B. ModelEvaluationOp: This component is primarily for model evaluation, not model upload and deployment.C. Custom Component with SDK: While feasible, it involves more setup and dependency management compared to using built-in components.

Question: 215 CertyIQ

You work for an online retailer. Your company has a few thousand short lifecycle products. Your company has five years of sales data stored in BigQuery. You have been asked to build a model that will make monthly sales predictions for each product. You want to use a solution that can be implemented quickly with minimal effort. What should you do?

A.Use Prophet on Vertex AI Training to build a custom model.

B.Use Vertex AI Forecast to build a NN-based model.

C.Use BigQuery ML to build a statistical ARIMA_PLUS model.

D.Use TensorFlow on Vertex AI Training to build a custom model.

Answer: C

Explanation:

Ease of Use: BigQuery ML integrates seamlessly with BigQuery, allowing you to create and train models directly within SQL queries, eliminating the need for separate environments or coding.Statistical ARIMA_PLUS Strengths: This model is well-suited for time series forecasting, automatically handling seasonality, trends, and holidays, making it appropriate for monthly sales predictions.Minimal Effort: BigQuery ML handles model training and tuning, reducing the need for manual configuration or hyperparameter tuning.Fast Implementation: Model creation and training can be done in a few lines of SQL, enabling rapid deployment.

Question: 216 CertyIQ

You are creating a model training pipeline to predict sentiment scores from text-based product reviews. You want to have control over how the model parameters are tuned, and you will deploy the model to an endpoint after it has been trained. You will use Vertex AI Pipelines to run the pipeline. You need to decide which Google Cloud pipeline components to use. What components should you choose?

A.TabularDatasetCreateOp, CustomTrainingJobOp, and EndpointCreateOp

B.TextDatasetCreateOp, AutoMLTextTrainingOp, and EndpointCreateOp

C.TabularDatasetCreateOp. AutoMLTextTrainingOp, and ModelDeployOp

D.TextDatasetCreateOp, CustomTrainingJobOp, and ModelDeployOp

Answer: D

Explanation:

TextDatasetCreateOp: This component is specifically designed to create datasets from text-based data, essential for handling product reviews.

CustomTrainingJobOp: This component provides full control over the training process, allowing you to specify model architecture, hyperparameter tuning strategies, and other training parameters, aligning with the requirement for control over model tuning.

ModelDeployOp: This component streamlines model deployment to a Vertex AI endpoint for real-time or batch inference, enabling the trained model to serve predictions.

Question: 217 CertyIQ

Your team frequently creates new ML models and runs experiments. Your team pushes code to a single repository hosted on Cloud Source Repositories. You want to create a continuous integration pipeline that automatically retrains the models whenever there is any modification of the code. What should be your first step to set up the CI

pipeline?

A.Configure a Cloud Build trigger with the event set as "Pull Request"
B.Configure a Cloud Build trigger with the event set as "Push to a branch"
C.Configure a Cloud Function that builds the repository each time there is a code change
D.Configure a Cloud Function that builds the repository each time a new branch is created

Answer: B

Explanation:

Cloud Build Integration: Cloud Build is Google Cloud's fully managed CI/CD platform, designed to automate builds and deployments, making it ideal for this task.Trigger on Code Pushes: Setting the trigger event to "Push to a branch" ensures that the pipeline automatically activates whenever new code is pushed to any branch of the repository, aligning with the goal of retraining models on code modifications.

Question: 218 CertyIQ

You have built a custom model that performs several memory-intensive preprocessing tasks before it makes a prediction. You deployed the model to a Vertex AI endpoint, and validated that results were received in a reasonable amount of time. After routing user traffic to the endpoint, you discover that the endpoint does not autoscale as expected when receiving multiple requests. What should you do?

A.Use a machine type with more memory
B.Decrease the number of workers per machine
C.Increase the CPU utilization target in the autoscaling configurations.
D.Decrease the CPU utilization target in the autoscaling configurations

Answer: D

Explanation:

The idea behind this question is getting autoscaling to handle well the fluctuating input of requests. Changing the machine (A) is not related to autoscaling, and you might not be using the full potential of the machine during the whole time, bur rather only during instances of peak traffic. You need to lower the autoscaling threshold (the target utilization metric mentioned in the options is CPU, so we will go with this) so you make use of more resources whenever too many memory-intensive requests are happening.

although memory-intensive is not directly related to CPU, for me the key is "the model does not autoscale as expected". To me this is addressing directly the settings of autoscaling, which won't change by changing the machine.

https://cloud.google.com/compute/docs/autoscaler/scaling-cpu#scaling_based_on_cpu_utilization

https://cloud.google.com/compute/docs/autoscaler#autoscaling_policy

Question: 219 CertyIQ

Your company manages an ecommerce website. You developed an ML model that recommends additional products to users in near real time based on items currently in the user's cart. The workflow will include the following processes:

1. The website will send a Pub/Sub message with the relevant data and then receive a message with the prediction

from Pub/Sub
2. Predictions will be stored in BigQuery
3. The model will be stored in a Cloud Storage bucket and will be updated frequently

You want to minimize prediction latency and the effort required to update the model. How should you reconfigure the architecture?

A.Write a Cloud Function that loads the model into memory for prediction. Configure the function to be triggered when messages are sent to Pub/Sub.

B.Create a pipeline in Vertex AI Pipelines that performs preprocessing, prediction, and postprocessing. Configure the pipeline to be triggered by a Cloud Function when messages are sent to Pub/Sub.

C.Expose the model as a Vertex AI endpoint. Write a custom DoFn in a Dataflow job that calls the endpoint for prediction.

D.Use the RunInference API with WatchFilePattern in a Dataflow job that wraps around the model and serves predictions.

Answer: D

Explanation:

Use the RunInference API with WatchFilePattern in a Dataflow job that wraps around the model and serves predictions.

Question: 220 CertyIQ

You are collaborating on a model prototype with your team. You need to create a Vertex AI Workbench environment for the members of your team and also limit access to other employees in your project. What should you do?

A.1. Create a new service account and grant it the Notebook Viewer role
2. Grant the Service Account User role to each team member on the service account
3. Grant the Vertex AI User role to each team member
4. Provision a Vertex AI Workbench user-managed notebook instance that uses the new service account

B.1. Grant the Vertex AI User role to the default Compute Engine service account
2. Grant the Service Account User role to each team member on the default Compute Engine service account
3. Provision a Vertex AI Workbench user-managed notebook instance that uses the default Compute Engine service account.

C.1. Create a new service account and grant it the Vertex AI User role
2. Grant the Service Account User role to each team member on the service account
3. Grant the Notebook Viewer role to each team member.
4. Provision a Vertex AI Workbench user-managed notebook instance that uses the new service account

D.1. Grant the Vertex AI User role to the primary team member
2. Grant the Notebook Viewer role to the other team members
3. Provision a Vertex AI Workbench user-managed notebook instance that uses the primary user's account

Answer: C

Explanation:

Vertex AI User Role: Granting this role to the service account provides it with necessary permissions to interact with Vertex AI services.

Service Account User Role: Assigning this role to team members allows them to impersonate the service account, enabling them to use its permissions.

Notebook Viewer Role: This role grants team members access to the notebook instance, but not direct Vertex AI resource management.

User-Managed Notebook Instance: This type of instance uses a specific service account, ensuring access control is aligned with the designated service account's permissions.

Question: 221 CertyIQ

You work at a leading healthcare firm developing state-of-the-art algorithms for various use cases. You have unstructured textual data with custom labels. You need to extract and classify various medical phrases with these labels. What should you do?

A.Use the Healthcare Natural Language API to extract medical entities

B.Use a BERT-based model to fine-tune a medical entity extraction model

C.Use AutoML Entity Extraction to train a medical entity extraction model

D.Use TensorFlow to build a custom medical entity extraction model

Answer: C

Explanation:

C. "AutoML Entity Extraction for Healthcare allows you to create a custom entity extraction model trained using your own annotated medical text and using your own categories." https://cloud.google.com/healthcare-api/docs/concepts/nlp#choosing_between_the_and

Question: 222 CertyIQ

You developed a custom model by using Vertex AI to predict your application's user churn rate. You are using Vertex AI Model Monitoring for skew detection. The training data stored in BigQuery contains two sets of features - demographic and behavioral. You later discover that two separate models trained on each set perform better than the original model. You need to configure a new model monitoring pipeline that splits traffic among the two models. You want to use the same prediction-sampling-rate and monitoring-frequency for each model. You also want to minimize management effort. What should you do?

A.Keep the training dataset as is. Deploy the models to two separate endpoints, and submit two Vertex AI Model Monitoring jobs with appropriately selected feature-thresholds parameters.

B.Keep the training dataset as is. Deploy both models to the same endpoint and submit a Vertex AI Model Monitoring job with a monitoring-config-from-file parameter that accounts for the model IDs and feature selections.

C.Separate the training dataset into two tables based on demographic and behavioral features. Deploy the models to two separate endpoints, and submit two Vertex AI Model Monitoring jobs.

D.Separate the training dataset into two tables based on demographic and behavioral features. Deploy both models to the same endpoint, and submit a Vertex AI Model Monitoring job with a monitoring-config-from-file parameter that accounts for the model IDs and training datasets.

Answer: D

Explanation:

D. You need to split the training dataset for each respective model. Furthermore, you only need to control for 2 differences between models in monitoring-config-from-file: model ID, and training set. Feature selection should be the same in both models.

D - makes more sense two models to be trained seperately and more accuarately also submits a Vertex AI

Model Monitoring job with a monitoring-config-from parameter which would enable the skew detection to work for each model.

Question: 223 CertyIQ

You work for a pharmaceutical company based in Canada. Your team developed a BigQuery ML model to predict the number of flu infections for the next month in Canada. Weather data is published weekly, and flu infection statistics are published monthly. You need to configure a model retraining policy that minimizes cost. What should you do?

A.Download the weather and flu data each week. Configure Cloud Scheduler to execute a Vertex AI pipeline to retrain the model weekly.

B.Download the weather and flu data each month. Configure Cloud Scheduler to execute a Vertex AI pipeline to retrain the model monthly.

C.Download the weather and flu data each week. Configure Cloud Scheduler to execute a Vertex AI pipeline to retrain the model every month.

D.Download the weather data each week, and download the flu data each month. Deploy the model to a Vertex AI endpoint with feature drift monitoring, and retrain the model if a monitoring alert is detected.

Answer: D

Explanation:

Selective Retraining: Retraining occurs only when necessary, triggered by feature drift alerts, reducing cloud resource usage and associated costs.Efficient Data Utilization: Weather data is downloaded weekly to capture potential changes, but model retraining waits for monthly flu data, ensuring model relevance without excessive updates.Early Drift Detection: Vertex AI's feature drift monitoring proactively identifies model performance degradation, prompting timely retraining to maintain accuracy.

Question: 224 CertyIQ

You are building a MLOps platform to automate your company's ML experiments and model retraining. You need to organize the artifacts for dozens of pipelines. How should you store the pipelines' artifacts?

A.Store parameters in Cloud SQL, and store the models' source code and binaries in GitHub.

B.Store parameters in Cloud SQL, store the models' source code in GitHub, and store the models' binaries in Cloud Storage.

C.Store parameters in Vertex ML Metadata, store the models' source code in GitHub, and store the models' binaries in Cloud Storage.

D.Store parameters in Vertex ML Metadata and store the models' source code and binaries in GitHub.

Answer: C

Explanation:

A. Cloud SQL and GitHub: Cloud SQL isn't designed for ML metadata management, potentially leading to challenges in tracking experiment details and lineage.B. Cloud SQL, GitHub, and Cloud Storage: While viable, this approach misses the benefits of Vertex ML Metadata for organized ML artifact management.D. Vertex ML Metadata and GitHub: Storing model binaries in GitHub can be inefficient for large files and might incur higher storage costs.

Question: 225 CertyIQ

You work for a telecommunications company. You're building a model to predict which customers may fail to pay their next phone bill. The purpose of this model is to proactively offer at-risk customers assistance such as service discounts and bill deadline extensions. The data is stored in BigQuery and the predictive features that are available for model training include:

- Customer_id
- Age
- Salary (measured in local currency)
- Sex
- Average bill value (measured in local currency)
- Number of phone calls in the last month (integer)
- Average duration of phone calls (measured in minutes)

You need to investigate and mitigate potential bias against disadvantaged groups, while preserving model accuracy.

What should you do?

A.Determine whether there is a meaningful correlation between the sensitive features and the other features. Train a BigQuery ML boosted trees classification model and exclude the sensitive features and any meaningfully correlated features.

B.Train a BigQuery ML boosted trees classification model with all features. Use the ML.GLOBAL_EXPLAIN method to calculate the global attribution values for each feature of the model. If the feature importance value for any of the sensitive features exceeds a threshold, discard the model and tram without this feature.

C.Train a BigQuery ML boosted trees classification model with all features. Use the ML.EXPLAIN_PREDICT method to calculate the attribution values for each feature for each customer in a test set. If for any individual customer, the importance value for any feature exceeds a predefined threshold, discard the model and train the model again without this feature.

D.Define a fairness metric that is represented by accuracy across the sensitive features. Train a BigQuery ML boosted trees classification model with all features. Use the trained model to make predictions on a test set. Join the data back with the sensitive features, and calculate a fairness metric to investigate whether it meets your requirements.

Answer: D

Explanation:

Direct Bias Assessment: It directly measures model fairness using a relevant metric, providing clear insights into potential issues.Preserving Information: It avoids prematurely removing features, potentially capturing valuable predictive signals while mitigating bias.Aligning with Goals: It allows tailoring the fairness metric to specific ethical and business objectives.

Question: 226 CertyIQ

You recently trained a XGBoost model that you plan to deploy to production for online inference. Before sending a predict request to your model's binary, you need to perform a simple data preprocessing step. This step exposes a REST API that accepts requests in your internal VPC Service Controls and returns predictions. You want to configure this preprocessing step while minimizing cost and effort. What should you do?

A.Store a pickled model in Cloud Storage. Build a Flask-based app, package the app in a custom container image, and deploy the model to Vertex AI Endpoints.

B.Build a Flask-based app, package the app and a pickled model in a custom container image, and deploy the model to Vertex AI Endpoints.

C.Build a custom predictor class based on XGBoost Predictor from the Vertex AI SDK, package it and a pickled model in a custom container image based on a Vertex built-in image, and deploy the model to Vertex AI Endpoints.

D.Build a custom predictor class based on XGBoost Predictor from the Vertex AI SDK, and package the handler in a custom container image based on a Vertex built-in container image. Store a pickled model in Cloud Storage, and deploy the model to Vertex AI Endpoints.

Answer: D

Explanation:

Minimal Custom Code: Leverages the pre-built XGBoost Predictor class for core model prediction, reducing development effort and potential errors.

Optimized Container Image: Utilizes a Vertex built-in container image, pre-configured for efficient model serving and compatibility with Vertex AI Endpoints.

Separated Model Storage: Stores the model in Cloud Storage, reducing container image size and simplifying model updates independently of the container.

VPC Service Controls: Vertex AI Endpoints support VPC Service Controls, ensuring adherence to internal traffic restrictions.

Question: 227 CertyIQ

You work at a bank. You need to develop a credit risk model to support loan application decisions. You decide to implement the model by using a neural network in TensorFlow. Due to regulatory requirements, you need to be able to explain the model's predictions based on its features. When the model is deployed, you also want to monitor the model's performance over time. You decided to use Vertex AI for both model development and deployment. What should you do?

A.Use Vertex Explainable AI with the sampled Shapley method, and enable Vertex AI Model Monitoring to check for feature distribution drift.

B.Use Vertex Explainable AI with the sampled Shapley method, and enable Vertex AI Model Monitoring to check for feature distribution skew.

C.Use Vertex Explainable AI with the XRAI method, and enable Vertex AI Model Monitoring to check for feature distribution drift.

D.Use Vertex Explainable AI with the XRAI method, and enable Vertex AI Model Monitoring to check for feature distribution skew.

Answer: A

Explanation:

Explainable AI with the XRAI method is for unstructured, image region analysis, in this case we use structured data for loan approval analysis.

Question: 228 CertyIQ

You are investigating the root cause of a misclassification error made by one of your models. You used Vertex AI Pipelines to train and deploy the model. The pipeline reads data from BigQuery. creates a copy of the data in Cloud Storage in TFRecord format, trains the model in Vertex AI Training on that copy, and deploys the model to a Vertex AI endpoint. You have identified the specific version of that model that misclassified, and you need to recover the data this model was trained on. How should you find that copy of the data?

A.Use Vertex AI Feature Store. Modify the pipeline to use the feature store, and ensure that all training data is stored in it. Search the feature store for the data used for the training.

B.Use the lineage feature of Vertex AI Metadata to find the model artifact. Determine the version of the model

and identify the step that creates the data copy and search in the metadata for its location.

C.Use the logging features in the Vertex AI endpoint to determine the timestamp of the model's deployment. Find the pipeline run at that timestamp. Identify the step that creates the data copy, and search in the logs for its location.

D.Find the job ID in Vertex AI Training corresponding to the training for the model. Search in the logs of that job for the data used for the training.

Answer: B

Explanation:

A. Feature Store: While useful for managing features, it might not store complete training datasets, and modifying the pipeline would not help recover historical data.C. Endpoint Logs and Pipeline Run: This approach involves more manual searching and might be less precise for identifying the exact data copy.D. Training Job Logs: Training job logs might not reliably contain complete data paths or might be purged after a certain period.

Question: 229 CertyIQ

You work for a manufacturing company. You need to train a custom image classification model to detect product defects at the end of an assembly line. Although your model is performing well, some images in your holdout set are consistently mislabeled with high confidence. You want to use Vertex AI to understand your model's results. What should you do?

A.Configure feature-based explanations by using Integrated Gradients. Set visualization type to PIXELS, and set clip_percent_upperbound to 95.

B.Create an index by using Vertex AI Matching Engine. Query the index with your mislabeled images.

C.Configure feature-based explanations by using XRAI. Set visualization type to OUTLINES, and set polarity to positive.

D.Configure example-based explanations. Specify the embedding output layer to be used for the latent space representation.

Answer: A

Explanation:

Configure feature-based explanations by using Integrated Gradients. Set visualization type to PIXELS, and set clip_percent_upperbound to 95.

Reference:

https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/visualizing-explanations

Question: 230 CertyIQ

You are training models in Vertex AI by using data that spans across multiple Google Cloud projects. You need to find, track, and compare the performance of the different versions of your models. Which Google Cloud services should you include in your ML workflow?

A.Dataplex, Vertex AI Feature Store, and Vertex AI TensorBoard

B.Vertex AI Pipelines, Vertex AI Feature Store, and Vertex AI Experiments

C.Dataplex, Vertex AI Experiments, and Vertex AI ML Metadata

D.Vertex AI Pipelines, Vertex AI Experiments, and Vertex AI Metadata

Answer: C

Explanation:

Dataplex, Vertex AI Experiments, and Vertex AI ML Metadata.

Question: 231 CertyIQ

You are using Keras and TensorFlow to develop a fraud detection model. Records of customer transactions are stored in a large table in BigQuery. You need to preprocess these records in a cost-effective and efficient way before you use them to train the model. The trained model will be used to perform batch inference in BigQuery. How should you implement the preprocessing workflow?

A.Implement a preprocessing pipeline by using Apache Spark, and run the pipeline on Dataproc. Save the preprocessed data as CSV files in a Cloud Storage bucket.

B.Load the data into a pandas DataFrame. Implement the preprocessing steps using pandas transformations, and train the model directly on the DataFrame.

C.Perform preprocessing in BigQuery by using SQL. Use the BigQueryClient in TensorFlow to read the data directly from BigQuery.

D.Implement a preprocessing pipeline by using Apache Beam, and run the pipeline on Dataflow. Save the preprocessed data as CSV files in a Cloud Storage bucket.

Answer: C

Explanation:

A. Spark on Dataproc: While powerful, it incurs additional cluster setup and management costs, potentially less cost-effective for this specific use case.

B. pandas DataFrame: Loading large datasets into memory might lead to resource constraints and performance issues, especially for large-scale preprocessing.

D. Apache Beam on Dataflow: While scalable, it introduces extra complexity for managing a separate pipeline and storage for preprocessed data.

Question: 232 CertyIQ

You need to use TensorFlow to train an image classification model. Your dataset is located in a Cloud Storage directory and contains millions of labeled images. Before training the model, you need to prepare the data. You want the data preprocessing and model training workflow to be as efficient, scalable, and low maintenance as possible. What should you do?

A.1. Create a Dataflow job that creates sharded TFRecord files in a Cloud Storage directory.
2. Reference tf.data.TFRecordDataset in the training script.
3. Train the model by using Vertex AI Training with a V100 GPU.

B.1. Create a Dataflow job that moves the images into multiple Cloud Storage directories, where each directory is named according to the corresponding label
2. Reference tfds.folder_dataset:ImageFolder in the training script.
3. Train the model by using Vertex AI Training with a V100 GPU.

C.1. Create a Jupyter notebook that uses an nt-standard-64 V100 GPU Vertex AI Workbench instance.
2. Write a Python script that creates sharded TFRecord files in a directory inside the instance.
3. Reference tf.data.TFRecordDataset in the training script.

4. Train the model by using the Workbench instance.

D.1. Create a Jupyter notebook that uses an n1-standard-64, V100 GPU Vertex AI Workbench instance.
2. Write a Python script that copies the images into multiple Cloud Storage directories, where each. directory is named according to the corresponding label.
3. Reference tfds.foladr_dataset.ImageFolder in the training script.
4. Train the model by using the Workbench instance.

Answer: A

Explanation:

B. Folder-Based Structure: While viable, it's less efficient for large datasets compared to TFRecord files, potentially leading to slower I/O during training.

C. Workbench Processing: Local preprocessing on a single instance can be less scalable and efficient for millions of images, potentially introducing bottlenecks.

D. Workbench Training: While Workbench offers a Jupyter environment, Vertex AI Training is specifically designed for scalable model training, providing optimized hardware and infrastructure.

Question: 233 CertyIQ

You are building a custom image classification model and plan to use Vertex AI Pipelines to implement the end-to-end training. Your dataset consists of images that need to be preprocessed before they can be used to train the model. The preprocessing steps include resizing the images, converting them to grayscale, and extracting features. You have already implemented some Python functions for the preprocessing tasks. Which components should you use in your pipeline?

A.DataprocSparkBatchOp and CustomTrainingJobOp

B.DataflowPythonJobOp, WaitGcpResourcesOp, and CustomTrainingJobOp

C.dsl.ParallelFor, dsl.component, and CustomTrainingJobOp

D.ImageDatasetImportDataOp, dsl.component, and AutoMLImageTrainingJobRunOp

Answer: B

Explanation:

A. DataprocSparkBatchOp: While capable of data processing, it's less well-suited for image-specific tasks like resizing and grayscale conversion compared to DataflowPythonJob.

OptionC. dsl.ParallelFor, dsl.component: While offering flexibility, they require more manual orchestration and potentially less efficient for image preprocessing compared to DataflowPythonJob

Option D. ImageDatasetImportDataOp, AutoMLImageTrainingJobRunOp: These components are designed for AutoML Image training, not directly compatible with custom preprocessing and training tasks.

Question: 234 CertyIQ

You work for a retail company that is using a regression model built with BigQuery ML to predict product sales. This model is being used to serve online predictions. Recently you developed a new version of the model that uses a different architecture (custom model). Initial analysis revealed that both models are performing as expected. You want to deploy the new version of the model to production and monitor the performance over the next two months. You need to minimize the impact to the existing and future model users. How should you deploy the model?

A.Import the new model to the same Vertex AI Model Registry as a different version of the existing model.

Deploy the new model to the same Vertex AI endpoint as the existing model, and use traffic splitting to route 95% of production traffic to the BigQuery ML model and 5% of production traffic to the new model.

B.Import the new model to the same Vertex AI Model Registry as the existing model. Deploy the models to one Vertex AI endpoint. Route 95% of production traffic to the BigQuery ML model and 5% of production traffic to the new model.

C.Import the new model to the same Vertex AI Model Registry as the existing model. Deploy each model to a separate Vertex AI endpoint.

D.Deploy the new model to a separate Vertex AI endpoint. Create a Cloud Run service that routes the prediction requests to the corresponding endpoints based on the input feature values.

Answer: A

Explanation:

B. Doesn't Specify Traffic Splitting: Deploying models to a single endpoint without explicit traffic splitting might lead to unpredictable model selection behavior, hindering controlled evaluation.

C. Separate Endpoints: While isolating models, it introduces complexity in managing multiple endpoints and routing logic, increasing operational overhead.

D. Cloud Run Routing: Adds complexity by requiring a separate service to manage routing, potentially increasing latency and maintenance overhead compared to Vertex AI's built-in traffic splitting.

Question: 235 CertyIQ

You are using Vertex AI and TensorFlow to develop a custom image classification model. You need the model's decisions and the rationale to be understandable to your company's stakeholders. You also want to explore the results to identify any issues or potential biases. What should you do?

A.1. Use TensorFlow to generate and visualize features and statistics.
2. Analyze the results together with the standard model evaluation metrics.

B.1. Use TensorFlow Profiler to visualize the model execution.
2. Analyze the relationship between incorrect predictions and execution bottlenecks.

C.1. Use Vertex Explainable AI to generate example-based explanations.
2. Visualize the results of sample inputs from the entire dataset together with the standard model evaluation metrics.

D.1. Use Vertex Explainable AI to generate feature attributions. Aggregate feature attributions over the entire dataset.
2. Analyze the aggregation result together with the standard model evaluation metrics.

Answer: D

Explanation:

Feature-Level Insights: Feature attributions pinpoint which image regions contribute most to predictions, offering granular understanding of model reasoning.Bias Detection: Aggregating feature attributions over the entire dataset can reveal systematic biases or patterns of model behavior, helping identify potential fairness issues.Complementary to Evaluation Metrics: Combining attributions with standard metrics (e.g., accuracy, precision, recall) provides a comprehensive view of model performance and fairness.

Question: 236 CertyIQ

You work for a large retailer, and you need to build a model to predict customer chum. The company has a dataset of historical customer data, including customer demographics purchase history, and website activity. You need to

create the model in BigQuery ML and thoroughly evaluate its performance. What should you do?

A.Create a linear regression model in BigQuery ML, and register the model in Vertex AI Model Registry. Evaluate the model performance in Vertex AI .

B.Create a logistic regression model in BigQuery ML and register the model in Vertex AI Model Registry. Evaluate the model performance in Vertex AI .

C.Create a linear regression model in BigQuery ML. Use the ML.EVALUATE function to evaluate the model performance.

D.Create a logistic regression model in BigQuery ML. Use the ML.CONFUSION_MATRIX function to evaluate the model performance.

Answer: B

Explanation:

Create a logistic regression model in BigQuery ML and register the model in Vertex AI Model Registry. Evaluate the model performance in Vertex AI .

Reference:

https://cloud.google.com/vertex-ai/docs/evaluation/introduction#classification_1

Question: 237 CertyIQ

You are developing a model to identify traffic signs in images extracted from videos taken from the dashboard of a vehicle. You have a dataset of 100,000 images that were cropped to show one out of ten different traffic signs. The images have been labeled accordingly for model training, and are stored in a Cloud Storage bucket. You need to be able to tune the model during each training run. How should you train the model?

A.Train a model for object detection by using Vertex AI AutoML.

B.Train a model for image classification by using Vertex AI AutoML.

C.Develop the model training code for object detection, and train a model by using Vertex AI custom training.

D.Develop the model training code for image classification, and train a model by using Vertex AI custom training.

Answer: D

Explanation:

Develop the model training code for image classification, and train a model by using Vertex AI custom training.

Question: 238 CertyIQ

You have deployed a scikit-team model to a Vertex AI endpoint using a custom model server. You enabled autoscaling: however, the deployed model fails to scale beyond one replica, which led to dropped requests. You notice that CPU utilization remains low even during periods of high load. What should you do?

A.Attach a GPU to the prediction nodes

B.Increase the number of workers in your model server

C.Schedule scaling of the nodes to match expected demand

D.Increase the minReplicaCount in your DeployedModel configuration

Answer: A

Explanation:

Attach a GPU to the prediction nodes.

Question: 239 CertyIQ

You work for a pet food company that manages an online forum. Customers upload photos of their pets on the forum to share with others. About 20 photos are uploaded daily. You want to automatically and in near real time detect whether each uploaded photo has an animal. You want to prioritize time and minimize cost of your application development and deployment. What should you do?

A.Send user-submitted images to the Cloud Vision API. Use object localization to identify all objects in the image and compare the results against a list of animals.

B.Download an object detection model from TensorFlow Hub. Deploy the model to a Vertex AI endpoint. Send new user-submitted images to the model endpoint to classify whether each photo has an animal.

C.Manually label previously submitted images with bounding boxes around any animals. Build an AutoML object detection model by using Vertex AI. Deploy the model to a Vertex AI endpoint Send new user-submitted images to your model endpoint to detect whether each photo has an animal.

D.Manually label previously submitted images as having animals or not. Create an image dataset on Vertex AI. Train a classification model by using Vertex AutoML to distinguish the two classes. Deploy the model to a Vertex AI endpoint. Send new user-submitted images to your model endpoint to classify whether each photo has an animal.

Answer: A

Explanation:

Send user-submitted images to the Cloud Vision API. Use object localization to identify all objects in the image and compare the results against a list of animals.

Question: 240 CertyIQ

You work at a mobile gaming startup that creates online multiplayer games. Recently, your company observed an increase in players cheating in the games, leading to a loss of revenue and a poor user experience You built a binary classification model to determine whether a player cheated after a completed game session, and then send a message to other downstream systems to ban the player that cheated. Your model has performed well during testing, and you now need to deploy the model to production. You want your serving solution to provide immediate classifications after a completed game session to avoid further loss of revenue. What should you do?

A.Import the model into Vertex AI Model Registry. Use the Vertex Batch Prediction service to run batch inference jobs.

B.Save the model files in a Cloud Storage bucket. Create a Cloud Function to read the model files and make online inference requests on the Cloud Function.

C.Save the model files in a VM. Load the model files each time there is a prediction request, and run an inference job on the VM

D.Import the model into Vertex AI Model Registry. Create a Vertex AI endpoint that hosts the model, and make online inference requests.

Answer: D

Explanation:

Option A: Batch prediction is too slow for your needs.Option B: Cloud Functions are ideal for short-lived tasks,

not for continuously serving models. Loading the model on every request would be inefficient.Option C: VMs offer less scalability and management overhead compared to Vertex AI.

Question: 241 CertyIQ

You have created a Vertex AI pipeline that automates custom model training. You want to add a pipeline component that enables your team to most easily collaborate when running different executions and comparing metrics both visually and programmatically. What should you do?

A.Add a component to the Vertex AI pipeline that logs metrics to a BigQuery table. Query the table to compare different executions of the pipeline. Connect BigQuery to Looker Studio to visualize metrics.

B.Add a component to the Vertex AI pipeline that logs metrics to a BigQuery table. Load the table into a pandas DataFrame to compare different executions of the pipeline. Use Matplotlib to visualize metrics.

C.Add a component to the Vertex AI pipeline that logs metrics to Vertex ML Metadata. Use Vertex AI Experiments to compare different executions of the pipeline. Use Vertex AI TensorBoard to visualize metrics.

D.Add a component to the Vertex AI pipeline that logs metrics to Vertex ML Metadata. Load the Vertex ML Metadata into a pandas DataFrame to compare different executions of the pipeline. Use Matplotlib to visualize metrics.

Answer: C

Explanation:

Add a component to the Vertex AI pipeline that logs metrics to Vertex ML Metadata. Use Vertex AI Experiments to compare different executions of the pipeline. Use Vertex AI TensorBoard to visualize metrics.

Question: 242 CertyIQ

Your team is training a large number of ML models that use different algorithms, parameters, and datasets. Some models are trained in Vertex AI Pipelines, and some are trained on Vertex AI Workbench notebook instances. Your team wants to compare the performance of the models across both services. You want to minimize the effort required to store the parameters and metrics. What should you do?

A.Implement an additional step for all the models running in pipelines and notebooks to export parameters and metrics to BigQuery.

B.Create a Vertex AI experiment. Submit all the pipelines as experiment runs. For models trained on notebooks log parameters and metrics by using the Vertex AI SDK.

C.Implement all models in Vertex AI Pipelines Create a Vertex AI experiment, and associate all pipeline runs with that experiment.

D.Store all model parameters and metrics as model metadata by using the Vertex AI Metadata API.

Answer: B

Explanation:

Vertex AI experiments - provides a unified way to store and compare model runs.

pipeline runs - It provides a unified way to store and compare model runs.

notebook instances - models trained on Vertex AI Workbench notebook instances, logging parameters and metrics using the Vertex AI SDK provides a consistent way to record the necessary information.

Question: 243 CertyIQ

You work on a team that builds state-of-the-art deep learning models by using the TensorFlow framework. Your team runs multiple ML experiments each week, which makes it difficult to track the experiment runs. You want a simple approach to effectively track, visualize, and debug ML experiment runs on Google Cloud while minimizing any overhead code. How should you proceed?

A.Set up Vertex AI Experiments to track metrics and parameters. Configure Vertex AI TensorBoard for visualization.

B.Set up a Cloud Function to write and save metrics files to a Cloud Storage bucket. Configure a Google Cloud VM to host TensorBoard locally for visualization.

C.Set up a Vertex AI Workbench notebook instance. Use the instance to save metrics data in a Cloud Storage bucket and to host TensorBoard locally for visualization.

D.Set up a Cloud Function to write and save metrics files to a BigQuery table. Configure a Google Cloud VM to host TensorBoard locally for visualization.

Answer: A

Explanation:

Options B and D: These options involve more setup and maintenance overhead, as they require managing Cloud Functions, VMs, and storage resources.

Option C: Vertex AI Workbench is excellent for interactive experimentation, but it's not optimized for long-term experiment tracking and visualization.

Question: 244 CertyIQ

Your work for a textile manufacturing company. Your company has hundreds of machines, and each machine has many sensors. Your team used the sensory data to build hundreds of ML models that detect machine anomalies. Models are retrained daily, and you need to deploy these models in a cost-effective way. The models must operate 24/7 without downtime and make sub millisecond predictions. What should you do?

A.Deploy a Dataflow batch pipeline and a Vertex AI Prediction endpoint.

B.Deploy a Dataflow batch pipeline with the RunInference API, and use model refresh.

C.Deploy a Dataflow streaming pipeline and a Vertex AI Prediction endpoint with autoscaling.

D.Deploy a Dataflow streaming pipeline with the RunInference API, and use automatic model refresh.

Answer: D

Explanation:

Deploy a Dataflow streaming pipeline with the RunInference API, and use automatic model refresh.

Reference:

https://beam.apache.org/documentation/ml/about-ml/

Question: 245 CertyIQ

You are developing an ML model that predicts the cost of used automobiles based on data such as location, condition, model type, color, and engine/battery efficiency. The data is updated every night. Car dealerships will use the model to determine appropriate car prices. You created a Vertex AI pipeline that reads the data splits the data into training/evaluation/test sets performs feature engineering trains the model by using the training dataset and validates the model by using the evaluation dataset. You need to configure a retraining workflow that

minimizes cost. What should you do?

A.Compare the training and evaluation losses of the current run. If the losses are similar, deploy the model to a Vertex AI endpoint. Configure a cron job to redeploy the pipeline every night.

B.Compare the training and evaluation losses of the current run. If the losses are similar, deploy the model to a Vertex AI endpoint with training/serving skew threshold model monitoring. When the model monitoring threshold is triggered redeploy the pipeline.

C.Compare the results to the evaluation results from a previous run. If the performance improved deploy the model to a Vertex AI endpoint. Configure a cron job to redeploy the pipeline every night.

D. Compare the results to the evaluation results from a previous run. If the performance improved deploy the model to a Vertex AI endpoint with training/serving skew threshold model monitoring. When the model monitoring threshold is triggered redeploy the pipeline.

Answer: B

Explanation:

Option A: Redeploying the pipeline every night without checking for degradation wastes resources if model performance is stable.

Option C: Comparing results to a previous run doesn't guarantee model degradation detection in the current run.

Option D: Comparing to a previous run and using model monitoring is redundant; model monitoring alone is sufficient.

Question: 246 CertyIQ

You recently used BigQuery ML to train an AutoML regression model. You shared results with your team and received positive feedback. You need to deploy your model for online prediction as quickly as possible. What should you do?

A.Retrain the model by using BigQuery ML, and specify Vertex AI as the model registry. Deploy the model from Vertex AI Model Registry to a Vertex AI endpoint,

B.Retrain the model by using Vertex Al Deploy the model from Vertex AI Model. Registry to a Vertex AI endpoint.

C.Alter the model by using BigQuery ML, and specify Vertex AI as the model registry. Deploy the model from Vertex AI Model Registry to a Vertex AI endpoint.

D.Export the model from BigQuery ML to Cloud Storage. Import the model into Vertex AI Model Registry. Deploy the model to a Vertex AI endpoint.

Answer: D

Explanation:

Export the model from Big Query ML to Cloud Storage. Import the model into Vertex AI Model Registry. Deploy the model to a Vertex AI endpoint.

Question: 247 CertyIQ

You built a deep learning-based image classification model by using on-premises data. You want to use Vertex AI to deploy the model to production. Due to security concerns, you cannot move your data to the cloud. You are aware that the input data distribution might change over time. You need to detect model performance changes in production. What should you do?

A.Use Vertex Explainable AI for model explainability. Configure feature-based explanations.

B.Use Vertex Explainable AI for model explainability. Configure example-based explanations.

C.Create a Vertex AI Model Monitoring job. Enable training-serving skew detection for your model.

D.Create a Vertex AI Model Monitoring job. Enable feature attribution skew and drift detection for your model.

Answer: C

Explanation:

Option A and B: Vertex Explainable AI provides insights into model behavior but doesn't directly detect performance changes or concept drift. It's more suitable for understanding model decisions, not monitoring production performance.

Option D: Feature attribution skew and drift detection requires feature attributions calculated during training, which might not be feasible without cloud access to the data.

Question: 248 CertyIQ

You trained a model packaged it with a custom Docker container for serving, and deployed it to Vertex AI Model Registry. When you submit a batch prediction job, it fails with this error: "Error model server never became ready. Please validate that your model file or container configuration are valid. " There are no additional errors in the logs. What should you do?

A.Add a logging configuration to your application to emit logs to Cloud Logging

B.Change the HTTP port in your model's configuration to the default value of 8080

C.Change the healthRoute value in your model's configuration to /healthcheck

D.Pull the Docker image locally, and use the docker run command to launch it locally. Use the docker logs command to explore the error logs

Answer: D

Explanation:

Option A: Adding logging to Cloud Logging is useful for long-term monitoring but might not provide immediate insights for this specific error.

Options B and C: Changing port and health check configuration might be necessary if incorrect, but local debugging often reveals the root cause more effectively.

Question: 249 CertyIQ

You are developing an ML model to identify your company's products in images. You have access to over one million images in a Cloud Storage bucket. You plan to experiment with different TensorFlow models by using Vertex AI Training. You need to read images at scale during training while minimizing data I/O bottlenecks. What should you do?

A.Load the images directly into the Vertex AI compute nodes by using Cloud Storage FUSE. Read the images by using the tf.data.Dataset.from_tensor_slices function

B.Create a Vertex AI managed dataset from your image data. Access the AIP_TRAINING_DATA_URI environment variable to read the images by using the tf.data.Dataset.list_files function.

C.Convert the images to TFRecords and store them in a Cloud Storage bucket. Read the TFRecords by using the tf.data.TFRecordDataset function.

D.Store the URLs of the images in a CSV file. Read the file by using the tf.data.experimental.CsvDataset

function.

Answer: C

Explanation:

Option A: Cloud Storage FUSE can be slower for large datasets and adds complexity.

Option B: Vertex AI managed datasets offer convenience but might not match TFRecord performance for large-scale image training.

Option D: CSV files require manual loading and parsing, increasing overhead.

Question: 250 CertyIQ

You work at an ecommerce startup. You need to create a customer churn prediction model. Your company's recent sales records are stored in a BigQuery table. You want to understand how your initial model is making predictions. You also want to iterate on the model as quickly as possible while minimizing cost. How should you build your first model?

A.Export the data to a Cloud Storage bucket. Load the data into a pandas DataFrame on Vertex AI Workbench and train a logistic regression model with scikit-learn.

B.Create a tf.data.Dataset by using the TensorFlow BigQueryClient. Implement a deep neural network in TensorFlow.

C.Prepare the data in BigQuery and associate the data with a Vertex AI dataset. Create an AutoMLTabularTrainingJob to tram a classification model.

D.Export the data to a Cloud Storage bucket. Create a tf.data.Dataset to read the data from Cloud Storage. Implement a deep neural network in TensorFlow.

Answer: C

Explanation:

Option A: While logistic regression is interpretable, manual training in Vertex AI Workbench adds time and complexity.

Options B and D: Deep neural networks can be powerful but often lack interpretability, making it challenging to understand model decisions. They also require more hands-on model development and infrastructure management.

Question: 251 CertyIQ

You are developing a training pipeline for a new XGBoost classification model based on tabular data. The data is stored in a BigQuery table. You need to complete the following steps:

1. Randomly split the data into training and evaluation datasets in a 65/35 ratio
2. Conduct feature engineering
3. Obtain metrics for the evaluation dataset
4. Compare models trained in different pipeline executions

How should you execute these steps?

A.1. Using Vertex AI Pipelines, add a component to divide the data into training and evaluation sets, and add another component for feature engineering.
2. Enable autologging of metrics in the training component.
3. Compare pipeline runs in Vertex AI Experiments.

B.1. Using Vertex AI Pipelines, add a component to divide the data into training and evaluation sets, and add another component for feature engineering.
2. Enable autologging of metrics in the training component.
3. Compare models using the artifacts' lineage in Vertex ML Metadata.

C.1. In BigQuery ML, use the CREATE MODEL statement with BOOSTED_TREE_CLASSIFIER as the model type and use BigQuery to handle the data splits.
2. Use a SQL view to apply feature engineering and train the model using the data in that view.
3. Compare the evaluation metrics of the models by using a SQL query with the ML.TRAINING_INFO statement.

D.1. In BigQuery ML, use the CREATE MODEL statement with BOOSTED_TREE_CLASSIFIER as the model type and use BigQuery to handle the data splits.
2. Use ML TRANSFORM to specify the feature engineering transformations and tram the model using the data in the table.
3. Compare the evaluation metrics of the models by using a SQL query with the ML.TRAINING_INFO statement.

Answer: A

Explanation:

Option B: While Vertex ML Metadata provides artifact lineage, it's less comprehensive for model comparison than Experiments.Options C and D: BigQuery ML is powerful for in-database model training, but it has limitations in pipeline orchestration, complex feature engineering, and detailed model comparison features, making it less suitable for this scenario.

Question: 252 CertyIQ

You work for a company that sells corporate electronic products to thousands of businesses worldwide. Your company stores historical customer data in BigQuery. You need to build a model that predicts customer lifetime value over the next three years. You want to use the simplest approach to build the model and you want to have access to visualization tools. What should you do?

A.Create a Vertex AI Workbench notebook to perform exploratory data analysis. Use IPython magics to create a new BigQuery table with input features. Use the BigQuery console to run the CREATE MODEL statement. Validate the results by using the ML.EVALUATE and ML.PREDICT statements.

B.Run the CREATE MODEL statement from the BigQuery console to create an AutoML model. Validate the results by using the ML.EVALUATE and ML.PREDICT statements.

C.Create a Vertex AI Workbench notebook to perform exploratory data analysis and create input features. Save the features as a CSV file in Cloud Storage. Import the CSV file as a new BigQuery table. Use the BigQuery console to run the CREATE MODEL statement. Validate the results by using the ML.EVALUATE and ML.PREDICT statements.

D.Create a Vertex AI Workbench notebook to perform exploratory data analysis. Use IPython magics to create a new BigQuery table with input features, create the model, and validate the results by using the CREATE MODEL, ML.EVALUATE, and ML.PREDICT statements.

Answer: A

Explanation:

Option B: While AutoML simplifies model selection and training, it lacks the flexibility and visualization capabilities of Vertex AI Workbench.Option C: Manually saving features as CSV files and importing them back into BigQuery involves unnecessary data movement and complexity.Option D: Completing all steps within the notebook is possible but requires more coding and might not be as intuitive for those less familiar with BigQuery ML syntax.

Question: 253 CertyIQ

You work for a delivery company. You need to design a system that stores and manages features such as parcels delivered and truck locations over time. The system must retrieve the features with low latency and feed those features into a model for online prediction. The data science team will retrieve historical data at a specific point in time for model training. You want to store the features with minimal effort. What should you do?

A.Store features in Bigtable as key/value data.

B.Store features in Vertex AI Feature Store.

C.Store features as a Vertex AI dataset, and use those features to train the models hosted in Vertex AI endpoints.

D.Store features in BigQuery timestamp partitioned tables, and use the BigQuery Storage Read API to serve the features.

Answer: B

Explanation:

Store features in Vertex AI Feature Store.

Question: 254 CertyIQ

You are working on a prototype of a text classification model in a managed Vertex AI Workbench notebook. You want to quickly experiment with tokenizing text by using a Natural Language Toolkit (NLTK) library. How should you add the library to your Jupyter kernel?

A.Install the NLTK library from a terminal by using the pip install nltk command.

B.Write a custom Dataflow job that uses NLTK to tokenize your text and saves the output to Cloud Storage.

C.Create a new Vertex AI Workbench notebook with a custom image that includes the NLTK library.

D.Install the NLTK library from a Jupyter cell by using the !pip install nltk --user command.

Answer: D

Explanation:

Direct Installation: It installs the library directly within the notebook environment, making it immediately available for use.

Simplicity: It requires a single command in a Jupyter cell, eliminating the need for external tools or configuration.

User-Specific Installation: The --user flag ensures the library is installed in your user space, avoiding conflicts with system-wide packages.

Question: 255 CertyIQ

You have recently used TensorFlow to train a classification model on tabular data. You have created a Dataflow pipeline that can transform several terabytes of data into training or prediction datasets consisting of TFRecords. You now need to productionize the model, and you want the predictions to be automatically uploaded to a BigQuery table on a weekly schedule. What should you do?

A.Import the model into Vertex AI and deploy it to a Vertex AI endpoint. On Vertex AI Pipelines, create a pipeline that uses the DataflowPythonJobOp and the ModelBacthPredictOp components.

B.Import the model into Vertex AI and deploy it to a Vertex AI endpoint. Create a Dataflow pipeline that reuses the data processing logic sends requests to the endpoint, and then uploads predictions to a BigQuery table.

C.Import the model into Vertex AI. On Vertex AI Pipelines, create a pipeline that uses the

DataflowPythonJobOp and the ModelBatchPredictOp components.

D.Import the model into BigQuery. Implement the data processing logic in a SQL query. On Vertex AI Pipelines create a pipeline that uses the BigquervQueryJobOp and the BigqueryPredictModelJobOp components.

Answer: B

Explanation:

Option A: Vertex AI Pipelines are excellent for orchestrating ML workflows but might not be as efficient as Dataflow for large-scale data processing, especially with existing Dataflow logic.

Option C: While Vertex AI Pipelines can handle model loading and prediction, Dataflow is better suited for large-scale data processing and BigQuery integration.

Option D: BigQuery ML is primarily for in-database model training and prediction, not ideal for external models or large-scale data processing.

Question: 256 CertyIQ

You work for an online grocery store. You recently developed a custom ML model that recommends a recipe when a user arrives at the website. You chose the machine type on the Vertex AI endpoint to optimize costs by using the queries per second (QPS) that the model can serve, and you deployed it on a single machine with 8 vCPUs and no accelerators.

A holiday season is approaching and you anticipate four times more traffic during this time than the typical daily traffic. You need to ensure that the model can scale efficiently to the increased demand. What should you do?

A.1. Maintain the same machine type on the endpoint.
2. Set up a monitoring job and an alert for CPU usage.
3. If you receive an alert, add a compute node to the endpoint.

B.1. Change the machine type on the endpoint to have 32 vCPUs.
2. Set up a monitoring job and an alert for CPU usage.
3. If you receive an alert, scale the vCPUs further as needed.

C.1. Maintain the same machine type on the endpoint Configure the endpoint to enable autoscaling based on vCPU usage.
2. Set up a monitoring job and an alert for CPU usage.
3. If you receive an alert, investigate the cause.

D.1. Change the machine type on the endpoint to have a GPU. Configure the endpoint to enable autoscaling based on the GPU usage.
2. Set up a monitoring job and an alert for GPU usage.
3. If you receive an alert, investigate the cause.

Answer: C

Explanation:

Cost Optimization: It starts with the current machine type, avoiding unnecessary upfront costs, and scales only when needed.Autoscaling: It automatically adjusts compute resources based on vCPU usage, ensuring the endpoint can handle traffic spikes without manual intervention.Monitoring and Alerting: It provides visibility into resource usage and triggers alerts for potential issues, enabling proactive actions.Investigation: It encourages investigation of alerts to identify any underlying problems beyond expected traffic growth, ensuring overall system health.

Question: 257 CertyIQ

You recently trained an XGBoost model on tabular data. You plan to expose the model for internal use as an HTTP microservice. After deployment, you expect a small number of incoming requests. You want to productionize the model with the least amount of effort and latency. What should you do?

A.Deploy the model to BigQuery ML by using CREATE MODEL with the BOOSTED_TREE_REGRESSOR statement, and invoke the BigQuery API from the microservice.

B.Build a Flask-based app. Package the app in a custom container on Vertex AI, and deploy it to Vertex AI Endpoints.

C.Build a Flask-based app. Package the app in a Docker image, and deploy it to Google Kubernetes Engine in Autopilot mode.

D.Use a prebuilt XGBoost Vertex container to create a model, and deploy it to Vertex AI Endpoints.

Answer: D

Explanation:

Prebuilt Container: It eliminates the need to build and manage a custom container, reducing development time and complexity.Vertex AI Endpoints: It provides a managed serving infrastructure with low latency and high availability, optimizing performance for predictions.Minimal Effort: It involves simple steps of creating a Vertex model and deploying it to an endpoint, streamlining the process.

Question: 258

CertyIQ

You work for an international manufacturing organization that ships scientific products all over the world. Instruction manuals for these products need to be translated to 15 different languages. Your organization's leadership team wants to start using machine learning to reduce the cost of manual human translations and increase translation speed. You need to implement a scalable solution that maximizes accuracy and minimizes operational overhead. You also want to include a process to evaluate and fix incorrect translations. What should you do?

A.Create a workflow using Cloud Function triggers. Configure a Cloud Function that is triggered when documents are uploaded to an input Cloud Storage bucket. Configure another Cloud Function that translates the documents using the Cloud Translation API, and saves the translations to an output Cloud Storage bucket. Use human reviewers to evaluate the incorrect translations.

B.Create a Vertex AI pipeline that processes the documents launches, an AutoML Translation training job, evaluates the translations and deploys the model to a Vertex AI endpoint with autoscaling and model monitoring. When there is a predetermined skew between training and live data, re-trigger the pipeline with the latest data.

C.Use AutoML Translation to train a model. Configure a Translation Hub project, and use the trained model to translate the documents. Use human reviewers to evaluate the incorrect translations.

D.Use Vertex AI custom training jobs to fine-tune a state-of-the-art open source pretrained model with your data. Deploy the model to a Vertex AI endpoint with autoscaling and model monitoring. When there is a predetermined skew between the training and live data, configure a trigger to run another training job with the latest data.

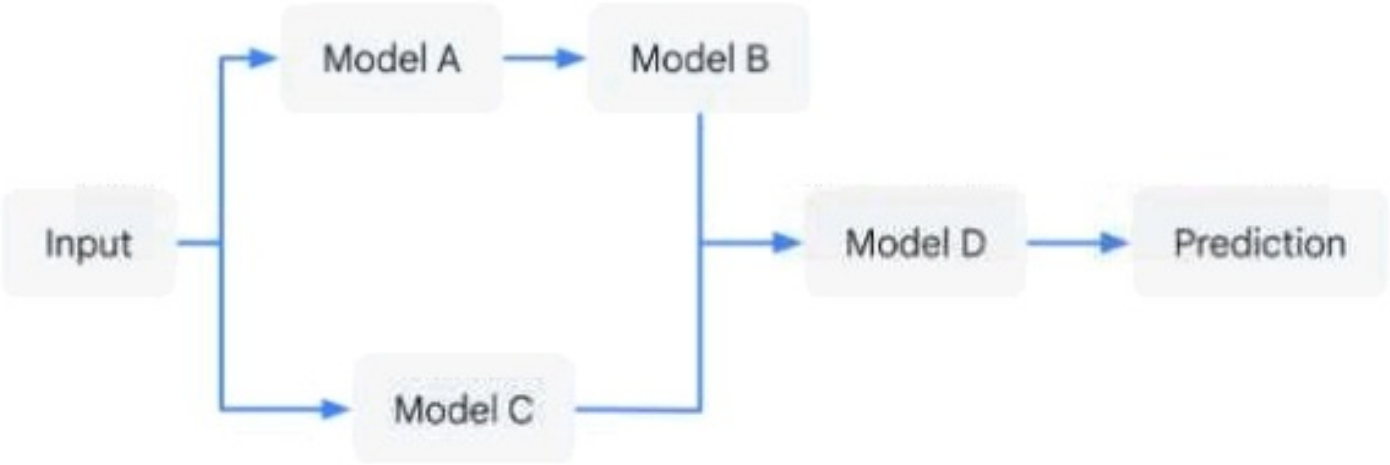
Answer: B

Explanation:

Option A: While Cloud Functions provide automation, the Cloud Translation API uses generic models that might not be as accurate for domain-specific content, potentially leading to more human corrections.Option C: Translation Hub offers collaboration features but lacks automated model training and pipeline orchestration, requiring more manual effort.Option D: Vertex AI custom training jobs provide flexibility but require more expertise and effort compared to AutoML Translation, and the pre-trained model might not be as well-suited for the specific domain.

Question: 259 CertyIQ

You have developed an application that uses a chain of multiple scikit-learn models to predict the optimal price for your company's products. The workflow logic is shown in the diagram. Members of your team use the individual models in other solution workflows. You want to deploy this workflow while ensuring version control for each individual model and the overall workflow. Your application needs to be able to scale down to zero. You want to minimize the compute resource utilization and the manual effort required to manage this solution. What should you do?



A.Expose each individual model as an endpoint in Vertex AI Endpoints. Create a custom container endpoint to orchestrate the workflow.

B.Create a custom container endpoint for the workflow that loads each model's individual files Track the versions of each individual model in BigQuery.

C.Expose each individual model as an endpoint in Vertex AI Endpoints. Use Cloud Run to orchestrate the workflow.

D.Load each model's individual files into Cloud Run. Use Cloud Run to orchestrate the workflow. Track the versions of each individual model in BigQuery.

Answer: C

Explanation:

Option A: A custom container endpoint for orchestration adds complexity and management overhead.Option B: Loading model files directly into a custom container endpoint can lead to versioning challenges and potential conflicts if models are shared across workflows.Option D: Using BigQuery for model versioning is not its primary function and might introduce complexities in model loading and management.

Question: 260 CertyIQ

You are developing a model to predict whether a failure will occur in a critical machine part. You have a dataset consisting of a multivariate time series and labels indicating whether the machine part failed. You recently started experimenting with a few different preprocessing and modeling approaches in a Vertex AI Workbench notebook. You want to log data and track artifacts from each run. How should you set up your experiments?

A.1. Use the Vertex AI SDK to create an experiment and set up Vertex ML Metadata.
2. Use the log_time_series_metrics function to track the preprocessed data, and use the log_merrics function to log loss values.

B.1. Use the Vertex AI SDK to create an experiment and set up Vertex ML Metadata.
2. Use the log_time_series_metrics function to track the preprocessed data, and use the log_metrics function to log loss values.

C.1. Create a Vertex AI TensorBoard instance and use the Vertex AI SDK to create an experiment and associate the TensorBoard instance.

2. Use the assign_input_artifact method to track the preprocessed data and use the log_time_series_metrics function to log loss values.

D.1. Create a Vertex AI TensorBoard instance, and use the Vertex AI SDK to create an experiment and associate the TensorBoard instance.
2. Use the log_time_series_metrics function to track the preprocessed data, and use the log_metrics function to log loss values.

Answer: C

Explanation:

C.Tensorboard for experimentation and comparison of different model runs.assign_input_artifacts to track preprocessed data, since it links artifacts as inputs to the execution.
https://cloud.google.com/python/docs/reference/aiplatform/latest/google.cloud.aiplatform.Execution#

google_cloud_aiplatform_Execution_assign_input_artifactsUsing log_time_series_metrics would make sense if what we were doing is logging a metric, which we aren't when we track the preprocessed data not yet ran by the model.

Question: 261 CertyIQ

You are developing a recommendation engine for an online clothing store. The historical customer transaction data is stored in BigQuery and Cloud Storage. You need to perform exploratory data analysis (EDA), preprocessing and model training. You plan to rerun these EDA, preprocessing, and training steps as you experiment with different types of algorithms. You want to minimize the cost and development effort of running these steps as you experiment. How should you configure the environment?

A.Create a Vertex AI Workbench user-managed notebook using the default VM instance, and use the %%bigquery magic commands in Jupyter to query the tables.

B.Create a Vertex AI Workbench managed notebook to browse and query the tables directly from the JupyterLab interface.

C.Create a Vertex AI Workbench user-managed notebook on a Dataproc Hub, and use the %%bigquery magic commands in Jupyter to query the tables.

D.Create a Vertex AI Workbench managed notebook on a Dataproc cluster, and use the spark-bigquery-connector to access the tables.

Answer: B

Explanation:

Create a Vertex AI Workbench managed notebook to browse and query the tables directly from the JupyterLab interface.

Reference:

https://cloud.google.com/bigquery/docs/visualize-jupyter

Question: 262 CertyIQ

You recently deployed a model to a Vertex AI endpoint and set up online serving in Vertex AI Feature Store. You have configured a daily batch ingestion job to update your featurestore. During the batch ingestion jobs, you discover that CPU utilization is high in your featurestore's online serving nodes and that feature retrieval latency is high. You need to improve online serving performance during the daily batch ingestion. What should you do?

A.Schedule an increase in the number of online serving nodes in your featurestore prior to the batch ingestion

jobs

B.Enable autoscaling of the online serving nodes in your featurestore

C.Enable autoscaling for the prediction nodes of your DeployedModel in the Vertex AI endpoint

D.Increase the worker_count in the ImportFeatureValues request of your batch ingestion job

Answer: B

Explanation:

Option A: Manually scheduling node increases requires prior knowledge of batch ingestion times and might not be as responsive to unexpected workload spikes.Option C: Autoscaling prediction nodes in the Vertex AI endpoint might help with model prediction latency but doesn't directly address feature retrieval latency from the featurestore.Option D: Increasing worker_count in the batch ingestion job could speed up ingestion but might further strain online serving nodes, potentially worsening latency.

Question: 263 CertyIQ

You are developing a custom TensorFlow classification model based on tabular data. Your raw data is stored in BigQuery. contains hundreds of millions of rows, and includes both categorical and numerical features. You need to use a MaxMin scaler on some numerical features, and apply a one-hot encoding to some categorical features such as SKU names. Your model will be trained over multiple epochs. You want to minimize the effort and cost of your solution. What should you do?

A.1. Write a SQL query to create a separate lookup table to scale the numerical features.
2. Deploy a TensorFlow-based model from Hugging Face to BigQuery to encode the text features.
3. Feed the resulting BigQuery view into Vertex AI Training.

B.1. Use BigQuery to scale the numerical features.
2. Feed the features into Vertex AI Training.
3. Allow TensorFlow to perform the one-hot text encoding.

C.1. Use TFX components with Dataflow to encode the text features and scale the numerical features.
2. Export results to Cloud Storage as TFRecords.
3. Feed the data into Vertex AI Training.

D.1. Write a SQL query to create a separate lookup table to scale the numerical features.
2. Perform the one-hot text encoding in BigQuery.
3. Feed the resulting BigQuery view into Vertex AI Training.

Answer: B

Explanation:

Option A: Involves creating a separate lookup table and deploying a Hugging Face model in BigQuery, increasing complexity and cost.Option C: While TFX offers robust preprocessing capabilities, it adds overhead for this use case and requires knowledge of Dataflow.Option D: Performing one-hot encoding in BigQuery can be less efficient than TensorFlow's optimized implementation.

Question: 264 CertyIQ

You work for a retail company. You have been tasked with building a model to determine the probability of churn for each customer. You need the predictions to be interpretable so the results can be used to develop marketing campaigns that target at-risk customers. What should you do?

A.Build a random forest regression model in a Vertex AI Workbench notebook instance. Configure the model to generate feature importances after the model is trained.

B.Build an AutoML tabular regression model. Configure the model to generate explanations when it makes

predictions.

C.Build a custom TensorFlow neural network by using Vertex AI custom training. Configure the model to generate explanations when it makes predictions.

D.Build a random forest classification model in a Vertex AI Workbench notebook instance. Configure the model to generate feature importances after the model is trained.

Answer: D

Explanation:

Option A: Regression, not classification, is used for random forest model, which is not appropriate for predicting probabilities.

Option B: While AutoML tabular can generate model explanations, random forests inherently provide more granular insights into feature importance.

Option C: Neural networks can be less interpretable than tree-based models, and generating explanations for them often requires additional techniques and libraries.

Reference:

https://cloud.google.com/bigquery/docs/xai-overview

Question: 265 CertyIQ

You work for a company that is developing an application to help users with meal planning. You want to use machine learning to scan a corpus of recipes and extract each ingredient (e.g., carrot, rice, pasta) and each kitchen cookware (e.g., bowl, pot, spoon) mentioned. Each recipe is saved in an unstructured text file. What should you do?

A.Create a text dataset on Vertex AI for entity extraction Create two entities called “ingredient” and “cookware”, and label at least 200 examples of each entity. Train an AutoML entity extraction model to extract occurrences of these entity types. Evaluate performance on a holdout dataset.

B.Create a multi-label text classification dataset on Vertex AI. Create a test dataset, and label each recipe that corresponds to its ingredients and cookware. Train a multi-class classification model. Evaluate the model’s performance on a holdout dataset.

C.Use the Entity Analysis method of the Natural Language API to extract the ingredients and cookware from each recipe. Evaluate the model's performance on a prelabeled dataset.

D.Create a text dataset on Vertex AI for entity extraction. Create as many entities as there are different ingredients and cookware. Train an AutoML entity extraction model to extract those entities. Evaluate the model’s performance on a holdout dataset.

Answer: A

Explanation:

Correct Answer: A

Option B: Multi-label text classification is less suitable for identifying specific entities within text and would require labeling entire recipes with multiple classes, increasing complexity and reducing model specificity.

Option C: Natural Language API's Entity Analysis might not be as accurate for this specialized domain as a model trained on custom recipe data.

Option D: Creating separate entities for each ingredient and cookware type would significantly increase labeling effort and potentially hinder model generalization.

Question: 266 CertyIQ

You work for an organization that operates a streaming music service. You have a custom production model that is serving a “next song” recommendation based on a user's recent listening history. Your model is deployed on a Vertex AI endpoint. You recently retrained the same model by using fresh data. The model received positive test results offline. You now want to test the new model in production while minimizing complexity. What should you do?

A.Create a new Vertex AI endpoint for the new model and deploy the new model to that new endpoint. Build a service to randomly send 5% of production traffic to the new endpoint. Monitor end-user metrics such as listening time. If end-user metrics improve between models over time, gradually increase the percentage of production traffic sent to the new endpoint.

B.Capture incoming prediction requests in BigQuery. Create an experiment in Vertex AI Experiments. Run batch predictions for both models using the captured data. Use the user’s selected song to compare the models performance side by side. If the new model’s performance metrics are better than the previous model, deploy the new model to production.

C.Deploy the new model to the existing Vertex AI endpoint. Use traffic splitting to send 5% of production traffic to the new model. Monitor end-user metrics, such as listening time. If end-user metrics improve between models over time, gradually increase the percentage of production traffic sent to the new model.

D.Configure a model monitoring job for the existing Vertex AI endpoint. Configure the monitoring job to detect prediction drift and set a threshold for alerts. Update the model on the endpoint from the previous model to the new model. If you receive an alert of prediction drift, revert to the previous model.

Answer: C

Explanation:

Option A: Building a separate service adds unnecessary complexity and requires managing two endpoints.Option B: Batch predictions in Vertex AI Experiments might not reflect real-time user behavior and don't directly affect the production environment.Option D: Model monitoring alerts for prediction drift might be triggered by natural variations in user behavior instead of genuine performance issues and could lead to unnecessary model rollbacks.

Question: 267 CertyIQ

You created a model that uses BigQuery ML to perform linear regression. You need to retrain the model on the cumulative data collected every week. You want to minimize the development effort and the scheduling cost. What should you do?

A.Use BigQuery’s scheduling service to run the model retraining query periodically.

B.Create a pipeline in Vertex AI Pipelines that executes the retraining query, and use the Cloud Scheduler API to run the query weekly.

C.Use Cloud Scheduler to trigger a Cloud Function every week that runs the query for retraining the model.

D.Use the BigQuery API Connector and Cloud Scheduler to trigger Workflows every week that retrains the model.

Answer: A

Explanation:

Option B: Vertex AI Pipelines offer flexibility for complex workflows, but it involves more development effort and potential costs for pipeline execution.Option C: Cloud Functions provide a serverless way to execute code, but they incur execution costs and require additional configuration for triggering and permissions.Option D: Workflows can manage complex orchestration, but configuring the BigQuery API Connector and Cloud Scheduler adds complexity and potential costs.

Question: 268 CertyIQ

You want to migrate a scikit-learn classifier model to TensorFlow. You plan to train the TensorFlow classifier model using the same training set that was used to train the scikit-learn model, and then compare the performances using a common test set. You want to use the Vertex AI Python SDK to manually log the evaluation metrics of each model and compare them based on their F1 scores and confusion matrices. How should you log the metrics?

A.Use the aiplatform.log_classification_metrics function to log the F1 score, and use the aiplatform.log_metrics function to log the confusion matrix.

B.Use the aiplatform.log_classification_metrics function to log the F1 score and the confusion matrix.

C.Use the aiplatform.log_metrics function to log the F1 score and the confusion matrix.

D.Use the aiplatform.log_metrics function to log the F1 score: and use the aiplatform.log_classification_metrics function to log the confusion matrix.

Answer: B

Explanation:

Option A: It's incorrect because aiplatform.log_metrics is a more general function that doesn't provide the same specialized structure for classification metrics.

Option C: While technically possible to log both metrics using aiplatform.log_metrics, it's less optimal as it requires manual formatting and might not be as easily interpreted by Vertex AI's visualization tools.

Option D: This is incorrect as it suggests using aiplatform.log_classification_metrics for the confusion matrix, but that function doesn't support logging confusion matrices directly.

Question: 269 CertyIQ

You are developing a model to help your company create more targeted online advertising campaigns. You need to create a dataset that you will use to train the model. You want to avoid creating or reinforcing unfair bias in the model. What should you do? (Choose two.)

A.Include a comprehensive set of demographic features

B.Include only the demographic groups that most frequently interact with advertisements

C.Collect a random sample of production traffic to build the training dataset

D.Collect a stratified sample of production traffic to build the training dataset

E.Conduct fairness tests across sensitive categories and demographics on the trained model

Answer: DE

Explanation:

D.Collect a stratified sample of production traffic to build the training dataset.

E.Conduct fairness tests across sensitive categories and demographics on the trained model.

Question: 270 CertyIQ

You are developing an ML model in a Vertex AI Workbench notebook. You want to track artifacts and compare models during experimentation using different approaches. You need to rapidly and easily transition successful experiments to production as you iterate on your model implementation. What should you do?

A.1. Initialize the Vertex SDK with the name of your experiment. Log parameters and metrics for each experiment, and attach dataset and model artifacts as inputs and outputs to each execution.
2. After a successful experiment create a Vertex AI pipeline.

B.1. Initialize the Vertex SDK with the name of your experiment. Log parameters and metrics for each experiment, save your dataset to a Cloud Storage bucket, and upload the models to Vertex AI Model Registry.
2. After a successful experiment, create a Vertex AI pipeline.

C.1. Create a Vertex AI pipeline with parameters you want to track as arguments to your PipelineJob. Use the Metrics, Model, and Dataset artifact types from the Kubeflow Pipelines DSL as the inputs and outputs of the components in your pipeline.
2. Associate the pipeline with your experiment when you submit the job.

D.1. Create a Vertex AI pipeline. Use the Dataset and Model artifact types from the Kubeflow Pipelines DSL as the inputs and outputs of the components in your pipeline.
2. In your training component, use the Vertex AI SDK to create an experiment run. Configure the log_params and log_metrics functions to track parameters and metrics of your experiment.

Answer: A

Explanation:

Option B: Manually saving datasets and models to Cloud Storage and Model Registry introduces extra steps and potential for inconsistencies.

Options C and D: Prioritizing pipeline creation limits flexibility and visibility during the experimentation phase, making it harder to track artifacts and compare models effectively.

Question: 271 CertyIQ

You recently created a new Google Cloud project. After testing that you can submit a Vertex AI Pipeline job from the Cloud Shell, you want to use a Vertex AI Workbench user-managed notebook instance to run your code from that instance. You created the instance and ran the code but this time the job fails with an insufficient permissions error. What should you do?

A.Ensure that the Workbench instance that you created is in the same region of the Vertex AI Pipelines resources you will use.

B.Ensure that the Vertex AI Workbench instance is on the same subnetwork of the Vertex AI Pipeline resources that you will use.

C.Ensure that the Vertex AI Workbench instance is assigned the Identity and Access Management (IAM) Vertex AI User role.

D.Ensure that the Vertex AI Workbench instance is assigned the Identity and Access Management (IAM) Notebooks Runner role.

Answer: C

Explanation:

A. Region Compatibility: While regional compatibility is important, it's not the primary cause of this permission error.B. Subnet Matching: Subnet alignment is usually not a requirement for Vertex AI pipeline job submission.D. Notebooks Runner Role: This role is primarily for executing notebook code, not managing Vertex AI resources.

Question: 272 CertyIQ

You work for a semiconductor manufacturing company. You need to create a real-time application that automates the quality control process. High-definition images of each semiconductor are taken at the end of the assembly

line in real time. The photos are uploaded to a Cloud Storage bucket along with tabular data that includes each semiconductor's batch number, serial number, dimensions, and weight. You need to configure model training and serving while maximizing model accuracy. What should you do?

A.Use Vertex AI Data Labeling Service to label the images, and tram an AutoML image classification model. Deploy the model, and configure Pub/Sub to publish a message when an image is categorized into the failing class.

B.Use Vertex AI Data Labeling Service to label the images, and train an AutoML image classification model. Schedule a daily batch prediction job that publishes a Pub/Sub message when the job completes.

C.Convert the images into an embedding representation. Import this data into BigQuery, and train a BigQuery ML K-means clustering model with two clusters. Deploy the model and configure Pub/Sub to publish a message when a semiconductor's data is categorized into the failing cluster.

D.Import the tabular data into BigQuery, use Vertex AI Data Labeling Service to label the data and train an AutoML tabular classification model. Deploy the model, and configure Pub/Sub to publish a message when a semiconductor's data is categorized into the failing class.

Answer: A

Explanation:

Use Vertex AI Data Labeling Service to label the images, and tram an AutoML image classification model. Deploy the model, and configure Pub/Sub to publish a message when an image is categorized into the failing class.

Question: 273 CertyIQ

You work for a rapidly growing social media company. Your team builds TensorFlow recommender models in an on-premises CPU cluster. The data contains billions of historical user events and 100,000 categorical features. You notice that as the data increases, the model training time increases. You plan to move the models to Google Cloud. You want to use the most scalable approach that also minimizes training time. What should you do?

A.Deploy the training jobs by using TPU VMs with TPUv3 Pod slices, and use the TPUEmbeading API

B.Deploy the training jobs in an autoscaling Google Kubernetes Engine cluster with CPUs

C.Deploy a matrix factorization model training job by using BigQuery ML

D.Deploy the training jobs by using Compute Engine instances with A100 GPUs, and use the tf.nn.embedding_lookup API

Answer: A

Explanation:

TPU (Tensor Processing Units) VMs are specialized hardware accelerators designed by Google specifically for machine learning tasks.TPUv3 Pod slices offer high scalability and are excellent for distributed training tasks.The TPUEmbedding API is optimized for handling large volumes of categorical features, which fits your scenario with 100,000 categorical features.This option is likely to offer the fastest training times due to specialized hardware and optimized APIs for large-scale machine learning tasks.

Question: 274 CertyIQ

You are training and deploying updated versions of a regression model with tabular data by using Vertex AI Pipelines, Vertex AI Training, Vertex AI Experiments, and Vertex AI Endpoints. The model is deployed in a Vertex AI endpoint, and your users call the model by using the Vertex AI endpoint. You want to receive an email when the feature data distribution changes significantly, so you can retrigger the training pipeline and deploy an updated version of your model. What should you do?

A.Use Vertex AI Model Monitoring. Enable prediction drift monitoring on the endpoint, and specify a notification email.

B.In Cloud Logging, create a logs-based alert using the logs in the Vertex AI endpoint. Configure Cloud Logging to send an email when the alert is triggered.

C.In Cloud Monitoring create a logs-based metric and a threshold alert for the metric. Configure Cloud Monitoring to send an email when the alert is triggered.

D.Export the container logs of the endpoint to BigQuery. Create a Cloud Function to run a SQL query over the exported logs and send an email. Use Cloud Scheduler to trigger the Cloud Function.

Answer: A

Explanation:

Options B and C: While Cloud Logging and Cloud Monitoring can be used for general monitoring, they don't have the same specialized focus on prediction drift, potentially requiring more complex setup and analysis.

Option D: Exporting logs to BigQuery and creating a Cloud Function for analysis can be time-consuming and less efficient compared to Vertex AI Model Monitoring's out-of-the-box capabilities.

Question: 275 CertyIQ

You have trained an XGBoost model that you plan to deploy on Vertex AI for online prediction. You are now uploading your model to Vertex AI Model Registry, and you need to configure the explanation method that will serve online prediction requests to be returned with minimal latency. You also want to be alerted when feature attributions of the model meaningfully change over time. What should you do?

A.1. Specify sampled Shapley as the explanation method with a path count of 5.
2. Deploy the model to Vertex AI Endpoints.
3. Create a Model Monitoring job that uses prediction drift as the monitoring objective.

B.1. Specify Integrated Gradients as the explanation method with a path count of 5.
2. Deploy the model to Vertex AI Endpoints.
3. Create a Model Monitoring job that uses prediction drift as the monitoring objective.

C.1. Specify sampled Shapley as the explanation method with a path count of 50.
2. Deploy the model to Vertex AI Endpoints.
3. Create a Model Monitoring job that uses training-serving skew as the monitoring objective.

D.1. Specify Integrated Gradients as the explanation method with a path count of 50.
2. Deploy the model to Vertex AI Endpoints.
3. Create a Model Monitoring job that uses training-serving skew as the monitoring objective.

Answer: A

Explanation:

Sampled Shapley is a fast and scalable approximation of the Shapley value, which is a game-theoretic concept that measures the contribution of each feature to the model prediction. Sampled Shapley is suitable for online prediction requests, as it can return feature attributions with minimal latency. The path count parameter controls the number of samples used to estimate the Shapley value, and a lower value means faster computation. Integrated Gradients is another explanation method that computes the average gradient along the path from a baseline input to the actual input. Integrated Gradients is more accurate than Sampled Shapley, but also more computationally intensive

not B as integrated gradients is only for Custom-trained TensorFlow models that use a TensorFlow prebuilt container to serve predictions and AutoML image models

Question: 276 CertyIQ

You work at a gaming startup that has several terabytes of structured data in Cloud Storage. This data includes gameplay time data, user metadata, and game metadata. You want to build a model that recommends new games to users that requires the least amount of coding. What should you do?

A.Load the data in BigQuery. Use BigQuery ML to train an Autoencoder model.

B.Load the data in BigQuery. Use BigQuery ML to train a matrix factorization model.

C.Read data to a Vertex AI Workbench notebook. Use TensorFlow to train a two-tower model.

D.Read data to a Vertex AI Workbench notebook. Use TensorFlow to train a matrix factorization model.

Answer: B

Explanation:

Load the data in BigQuery. Use BigQuery ML to train a matrix factorization model.

Reference:

https://developers.google.com/machine-learning/recommendation/collaborative/matrix

Question: 277 CertyIQ

You work for a large bank that serves customers through an application hosted in Google Cloud that is running in the US and Singapore. You have developed a PyTorch model to classify transactions as potentially fraudulent or not. The model is a three-layer perceptron that uses both numerical and categorical features as input, and hashing happens within the model.

You deployed the model to the us-central1 region on nl-highcpu-16 machines, and predictions are served in real time. The model's current median response latency is 40 ms. You want to reduce latency, especially in Singapore, where some customers are experiencing the longest delays. What should you do?

A.Attach an NVIDIA T4 GPU to the machines being used for online inference.

B.Change the machines being used for online inference to nl-highcpu-32.

C.Deploy the model to Vertex AI private endpoints in the us-central1 and asia-southeast1 regions, and allow the application to choose the appropriate endpoint.

D.Create another Vertex AI endpoint in the asia-southeast1 region, and allow the application to choose the appropriate endpoint.

Answer: C

Explanation:

Deploy the model to Vertex AI private endpoints in the us-central1 and asia-southeast1 regions, and allow the application to choose the appropriate endpoint.

Question: 278 CertyIQ

You need to train an XGBoost model on a small dataset. Your training code requires custom dependencies. You want to minimize the startup time of your training job. How should you set up your Vertex AI custom training job?

A.Store the data in a Cloud Storage bucket, and create a custom container with your training application. In your training application, read the data from Cloud Storage and train the model.

B.Use the XGBoost prebuilt custom container. Create a Python source distribution that includes the data and installs the dependencies at runtime. In your training application, load the data into a pandas DataFrame and train the model.

C.Create a custom container that includes the data. In your training application, load the data into a pandas DataFrame and train the model.

D.Store the data in a Cloud Storage bucket, and use the XGBoost prebuilt custom container to run your training application. Create a Python source distribution that installs the dependencies at runtime. In your training application, read the data from Cloud Storage and train the model.

Answer: A

Explanation:

Store the data in a Cloud Storage bucket, and create a custom container with your training application. In your training application, read the data from Cloud Storage and train the model.

Question: 279 CertyIQ

You are creating an ML pipeline for data processing, model training, and model deployment that uses different Google Cloud services. You have developed code for each individual task, and you expect a high frequency of new files. You now need to create an orchestration layer on top of these tasks. You only want this orchestration pipeline to run if new files are present in your dataset in a Cloud Storage bucket. You also want to minimize the compute node costs. What should you do?

A.Create a pipeline in Vertex AI Pipelines. Configure the first step to compare the contents of the bucket to the last time the pipeline was run. Use the scheduler API to run the pipeline periodically.

B.Create a Cloud Function that uses a Cloud Storage trigger and deploys a Cloud Composer directed acyclic graph (DAG).

C.Create a pipeline in Vertex AI Pipelines. Create a Cloud Function that uses a Cloud Storage trigger and deploys the pipeline.

D.Deploy a Cloud Composer directed acyclic graph (DAG) with a GCSObjectUpdateSensor class that detects when a new file is added to the Cloud Storage bucket.

Answer: C

Explanation:

Create a pipeline in Vertex AI Pipelines. Create a Cloud Function that uses a Cloud Storage trigger and deploys the pipeline.

Question: 280 CertyIQ

You are using Kubeflow Pipelines to develop an end-to-end PyTorch-based MLOps pipeline. The pipeline reads data from BigQuery, processes the data, conducts feature engineering, model training, model evaluation, and deploys the model as a binary file to Cloud Storage. You are writing code for several different versions of the feature engineering and model training steps, and running each new version in Vertex AI Pipelines. Each pipeline run is taking over an hour to complete. You want to speed up the pipeline execution to reduce your development time, and you want to avoid additional costs. What should you do?

A.Comment out the part of the pipeline that you are not currently updating.

B.Enable caching in all the steps of the Kubeflow pipeline.

C.Delegate feature engineering to BigQuery and remove it from the pipeline.

D.Add a GPU to the model training step.

Answer: B

Explanation:

Enable caching in all the steps of the Kubeflow pipeline.

Question: 281 CertyIQ

You work at a large organization that recently decided to move their ML and data workloads to Google Cloud. The data engineering team has exported the structured data to a Cloud Storage bucket in Avro format. You need to propose a workflow that performs analytics, creates features, and hosts the features that your ML models use for online prediction. How should you configure the pipeline?

A.Ingest the Avro files into Cloud Spanner to perform analytics. Use a Dataflow pipeline to create the features, and store them in Vertex AI Feature Store for online prediction.

B.Ingest the Avro files into BigQuery to perform analytics. Use a Dataflow pipeline to create the features, and store them in Vertex AI Feature Store for online prediction.

C.Ingest the Avro files into Cloud Spanner to perform analytics. Use a Dataflow pipeline to create the features, and store them in BigQuery for online prediction.

D.Ingest the Avro files into BigQuery to perform analytics. Use BigQuery SQL to create features and store them in a separate BigQuery table for online prediction.

Answer: B

Explanation:

Ingest the Avro files into BigQuery to perform analytics. Use a Dataflow pipeline to create the features, and store them in Vertex AI Feature Store for online prediction.

Question: 282 CertyIQ

You work at an organization that maintains a cloud-based communication platform that integrates conventional chat, voice, and video conferencing into one platform. The audio recordings are stored in Cloud Storage. All recordings have an 8 kHz sample rate and are more than one minute long. You need to implement a new feature in the platform that will automatically transcribe voice call recordings into a text for future applications, such as call summarization and sentiment analysis. How should you implement the voice call transcription feature following Google-recommended best practices?

A.Use the original audio sampling rate, and transcribe the audio by using the Speech-to-Text API with synchronous recognition.

B.Use the original audio sampling rate, and transcribe the audio by using the Speech-to-Text API with asynchronous recognition.

C.Upsample the audio recordings to 16 kHz, and transcribe the audio by using the Speech-to-Text API with synchronous recognition.

D.Upsample the audio recordings to 16 kHz, and transcribe the audio by using the Speech-to-Text API with asynchronous recognition.

Answer: B

Explanation:

Use the original audio sampling rate, and transcribe the audio by using the Speech-to-Text API with asynchronous recognition.

Question: 283 CertyIQ

You work for a multinational organization that has recently begun operations in Spain. Teams within your organization will need to work with various Spanish documents, such as business, legal, and financial documents. You want to use machine learning to help your organization get accurate translations quickly and with the least effort. Your organization does not require domain-specific terms or jargon. What should you do?

A.Create a Vertex AI Workbench notebook instance. In the notebook, extract sentences from the documents, and train a custom AutoML text model.

B.Use Google Translate to translate 1,000 phrases from Spanish to English. Using these translated pairs, train a custom AutoML Translation model.

C.Use the Document Translation feature of the Cloud Translation API to translate the documents.

D.Create a Vertex AI Workbench notebook instance. In the notebook, convert the Spanish documents into plain text, and create a custom TensorFlow seq2seq translation model.

Answer: C

Explanation:

Use the Document Translation feature of the Cloud Translation API to translate the documents.

Question: 284 CertyIQ

You have a custom job that runs on Vertex AI on a weekly basis. The job is implemented using a proprietary ML workflow that produces the datasets, models, and custom artifacts, and sends them to a Cloud Storage bucket. Many different versions of the datasets and models were created. Due to compliance requirements, your company needs to track which model was used for making a particular prediction, and needs access to the artifacts for each model. How should you configure your workflows to meet these requirements?

A.Use the Vertex AI Metadata API inside the custom job to create context, execution, and artifacts for each model, and use events to link them together.

B.Create a Vertex AI experiment, and enable autologging inside the custom job.

C.Configure a TensorFlow Extended (TFX) ML Metadata database, and use the ML Metadata API.

D.Register each model in Vertex AI Model Registry, and use model labels to store the related dataset and model information.

Answer: A

Explanation:

Use the Vertex AI Metadata API inside the custom job to create context, execution, and artifacts for each model, and use events to link them together.

Question: 285 CertyIQ

You have recently developed a custom model for image classification by using a neural network. You need to automatically identify the values for learning rate, number of layers, and kernel size. To do this, you plan to run multiple jobs in parallel to identify the parameters that optimize performance. You want to minimize custom code development and infrastructure management. What should you do?

A.Train an AutoML image classification model.

B.Create a custom training job that uses the Vertex AI Vizier SDK for parameter optimization.

C.Create a Vertex AI hyperparameter tuning job.

D.Create a Vertex AI pipeline that runs different model training jobs in parallel.

Answer: C

Explanation:

Create a Vertex AI hyperparameter tuning job.

Thank you

Thank you for being so interested in the premium exam material.
I'm glad to hear that you found it informative and helpful.

If you have any feedback or thoughts on the bumps, I would love to hear them.
Your insights can help me improve our writing and better understand our readers.

Best of Luck

You have worked hard to get to this point, and you are well-prepared for the exam
Keep your head up, stay positive, and go show that exam what you're made of!

Feedback

More Papers