

Google.Professional-Machine-Learning-Engineer.v2023-12-28.q120

Exam Code:	Professional-Machine-Learning-Engineer
Exam Name:	Google Professional Machine Learning Engineer
Certification Provider:	Google
Free Question Number:	120
Version:	v2023-12-28
# of views:	2434
# of Questions views:	1200
https://www.exam-tests.com/Professional-Machine-Learning-Engineer-exam/Google.Professional-Machine-Learning-Engineer.v2023-12-28.q120.html	

NEW QUESTION: 1

You work for an online travel agency that also sells advertising placements on its website to other companies. You have been asked to predict the most relevant web banner that a user should see next. Security is important to your company. The model latency requirements are 300ms@p99, the inventory is thousands of web banners, and your exploratory analysis has shown that navigation context is a good predictor. You want to Implement the simplest solution. How should you configure the prediction pipeline?

A. Embed the client on the website, and then deploy the model on AI Platform Prediction.

B. Embed the client on the website, deploy the gateway on App Engine, and then deploy the model on AI Platform Prediction.

C. Embed the client on the website, deploy the gateway on App Engine, deploy the database on Cloud Bigtable for writing and for reading the user's navigation context, and then deploy the model on AI Platform Prediction.

D. Embed the client on the website, deploy the gateway on App Engine, deploy the database on Memorystore for writing and for reading the user's navigation context, and then deploy the model on Google Kubernetes Engine.

Answer: C ([LEAVE A REPLY](#))

<https://medium.com/google-cloud/secure-cloud-run-cloud-functions-and-app-engine-with-api-key-73c57bededd1>

NEW QUESTION: 2

Your task is classify if a company logo is present on an image. You found out that 96% of a data does not include a logo. You are dealing with data imbalance problem. Which metric do you use to evaluate to model?

A. F1 Score

B. F Score with higher recall weighted than precision

C. F Score with higher precision weighting than recall

D. RMSE

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 3

An interactive online dictionary wants to add a widget that displays words used in similar contexts. A Machine Learning Specialist is asked to provide word features for the downstream nearest neighbor model powering the widget.

What should the Specialist do to meet these requirements?

- A. Create one-hot word encoding vectors.
- B. Produce a set of synonyms for every word using Amazon Mechanical Turk.
- C. Create word embedding vectors that store edit distance with every other word.
- D. Download word embeddings pre-trained on a large corpus.

Answer: A ([LEAVE A REPLY](#))

Explanation/Reference: <https://aws.amazon.com/blogs/machine-learning/amazon-sagemaker-object2vec-adds-new-features-that-support-automatic-negative-sampling-and-speed-up-training/>

NEW QUESTION: 4

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. Use one-hot encoding on all categorical features.
- B. Write your data in TFRecords.
- C. Z-normalize all the numeric features.
- D. Oversample the fraudulent transaction 10 times.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 5

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 of each version of your models using seasons and years as events in Vertex ML Metadata. Compare the results across the slices.
- B. 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.
- C. Create versions of your models for each season per year in Vertex AI. Compare the performance statistics across the models in the Evaluate tab of the Vertex AI UI.
- D. Store the performance statistics in Cloud SQL. Query that database to compare the performance statistics across the model versions.

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 6

While performing exploratory data analysis on a dataset, you find that an important categorical feature has 5% null values. You want to minimize the bias that could result from the missing values. How should you handle the missing values?

- A. Replace the missing values with a placeholder category indicating a missing value.
- B. Replace the missing values with the feature's mean.
- C. Move the rows with missing values to your validation dataset.

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

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 7

You are an ML engineer at a large grocery retailer with stores in multiple regions. You have been asked to create an inventory prediction model. Your models features include region, location, historical demand, and seasonal popularity. You want the algorithm to learn from new inventory data on a daily basis. Which algorithms should you use to build the model?

A. Classification

B. Reinforcement Learning

C. Recurrent Neural Networks (RNN)

D. Convolutional Neural Networks (CNN)

Answer: C ([LEAVE A REPLY](#))

"algorithm to learn from new inventory data on a daily basis" = time series model , best option to deal with time series is forsure RNN
<https://builtin.com/data-science/recurrent-neural-networks-and-lstm>

NEW QUESTION: 8

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. 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.

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

C. Remove the data transformation step from your pipeline.

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

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 9

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

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

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

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

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

Answer: A ([LEAVE A REPLY](#))

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

Explanation:

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

NEW QUESTION: 10

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 removing features highly correlated with the target value.
- C. Address data leakage by applying nested cross-validation during model training.
- D. Address the model overfitting by tuning the hyperparameters to reduce the AUC ROC value.

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 11

A credit card company wants to build a credit scoring model to help predict whether a new credit card applicant will default on a credit card payment. The company has collected data from a large number of sources with thousands of raw attributes. Early experiments to train a classification model revealed that many attributes are highly correlated, the large number of features slows down the training speed significantly, and that there are some overfitting issues.

The Data Scientist on this project would like to speed up the model training time without losing a lot of information from the original dataset.

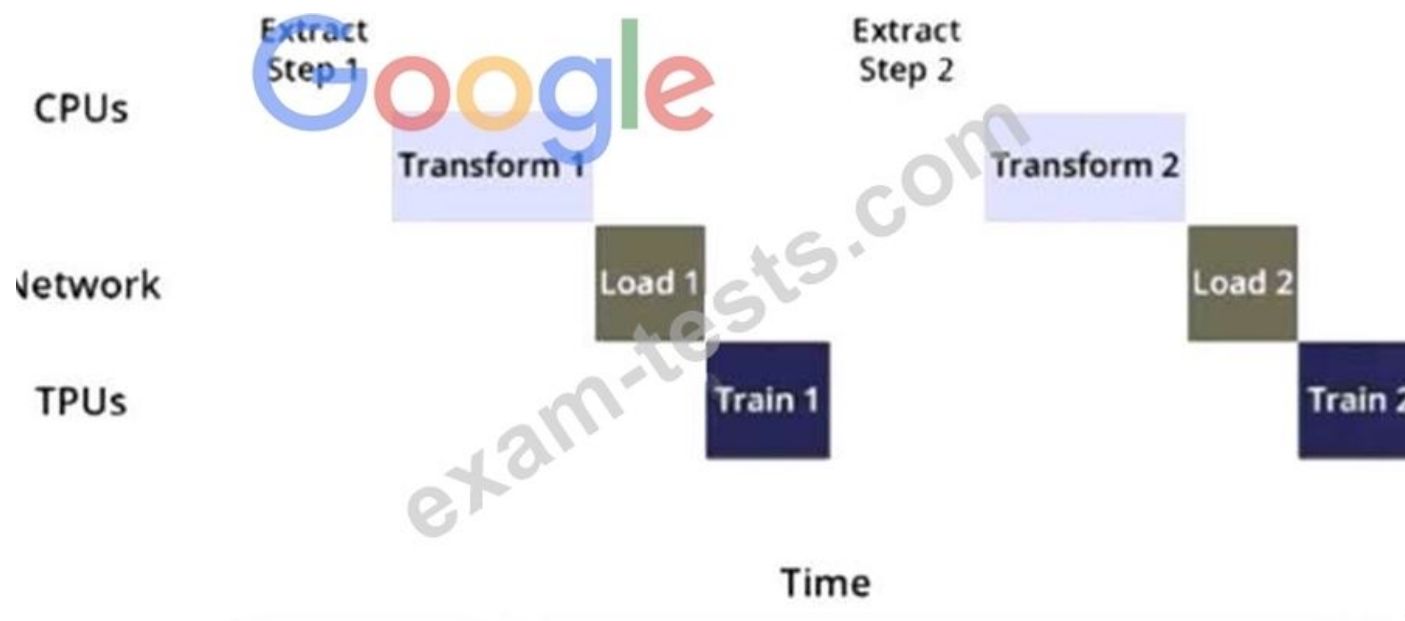
Which feature engineering technique should the Data Scientist use to meet the objectives?

- A. Normalize all numerical values to be between 0 and 1
- B. Run self-correlation on all features and remove highly correlated features
- C. Cluster raw data using k-means and use sample data from each cluster to build a new dataset
- D. Use an autoencoder or principal component analysis (PCA) to replace original features with new features

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 12

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 8 NVIDIA V100 GPUs and increase batch size.
- B. Rewrite your input function to resize and reshape the input images.
- C. Rewrite your input function using parallel reads, parallel processing, and prefetch.
- D. Move from Cloud TPU v2 to Cloud TPU v3 and increase batch size.

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 13

You are developing an image recognition model using PyTorch based on ResNet50 architecture. Your code is working fine on your local laptop on a small subsample. Your full dataset has 200k labeled images. You want to quickly scale your training workload while minimizing cost. You plan to use 4 V100 GPUs. What should you do? (Choose Correct Answer and Give Reference and Explanation)

- A. Configure a Compute Engine VM with all the dependencies that launches the training. Train your model with Vertex AI using a custom tier that contains the required GPUs.
- B. Package your code with Setuptools, and use a pre-built container. Train your model with Vertex AI using a custom tier that contains the required GPUs.
- C. Create a Vertex AI Workbench user-managed notebooks instance with 4 V100 GPUs, and use it to train your model.
- D. Create a Google Kubernetes Engine cluster with a node pool that has 4 V100 GPUs. Prepare and submit a TFJob operator to this node pool.

Answer: D ([LEAVE A REPLY](#))

Google Kubernetes Engine (GKE) is a powerful and easy-to-use platform for deploying and managing containerized applications. It allows you to create a cluster of virtual machines that are pre-configured with the necessary dependencies and resources to run your machine learning workloads. By creating a GKE cluster with a node pool that has 4 V100 GPUs, you can take advantage of the powerful processing capabilities of these GPUs to train your model quickly and efficiently.

You can then use the Kubernetes Framework such as TFJob operator to submit the job of training your model, which will automatically distribute the workload across the available GPUs.

Reference:

Google Kubernetes Engine

TFJob operator

Vertex AI

NEW QUESTION: 14

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

```
signature_name: serving_default_text:0
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,
```

What is the correct way to write the predict request?

- A. data = json.dumps({"signature_name": "serving_default", "instances": [fab', 'be1, 'cd']})
- B. data = json.dumps({"signature_name": "serving_default", "instances": [['a', 'b', 'c', 'd', 'e', 'f']])
- C. data = json.dumps({"signature_name": "serving_default", "instances": [['a', 'b', 'c', 'd', 'e', 'f']])
- D. data = json.dumps({"signature_name": "serving_default", "instances": [['a', 'b'], ['c', 'd'], ['e', 'f']])

Answer: D ([LEAVE A REPLY](#))

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

NEW QUESTION: 15

A data scientist has developed a machine learning translation model for English to Japanese by using Amazon SageMaker's built-in seq2seq algorithm with 500,000 aligned sentence pairs. While testing with sample sentences, the data scientist finds that the translation quality is reasonable for an example as short as five words. However, the quality becomes unacceptable if the sentence is 100 words long.

Which action will resolve the problem?

- A. Choose a different weight initialization type.
- B. Adjust hyperparameters related to the attention mechanism.
- C. Add more nodes to the recurrent neural network (RNN) than the largest sentence's word count.
- D. Change preprocessing to use n-grams.

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 16

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. Vary features independently to identify the threshold per feature that changes the classification.
- B. Use the correlation with target values in the data summary page.
- C. Use the feature importance percentages in the model evaluation page.
- D. Use local feature importance from the predictions.

Answer: C ([LEAVE A REPLY](#))

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPASS.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)

NEW QUESTION: 17

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

- A. Keep the original test dataset unchanged even if newer products are incorporated into retraining
- B. Update your test dataset with images of the newer products when your evaluation metrics drop below a pre-decided threshold.
- C. Extend your test dataset with images of the newer products when they are introduced to retraining
- D. Replace your test dataset with images of the newer products when they are introduced to retraining.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 18

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 97.5% of the most popular clickbait videos measured by number of clicks.
- B. The Pearson correlation coefficient between the log-transformed number of views after 7 days and 30 days after publication is equal to 0.
- C. The model predicts videos as popular if the user who uploads them has over 10,000 likes.
- D. The model predicts 95% of the most popular videos measured by watch time within 30 days of being uploaded.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 19

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

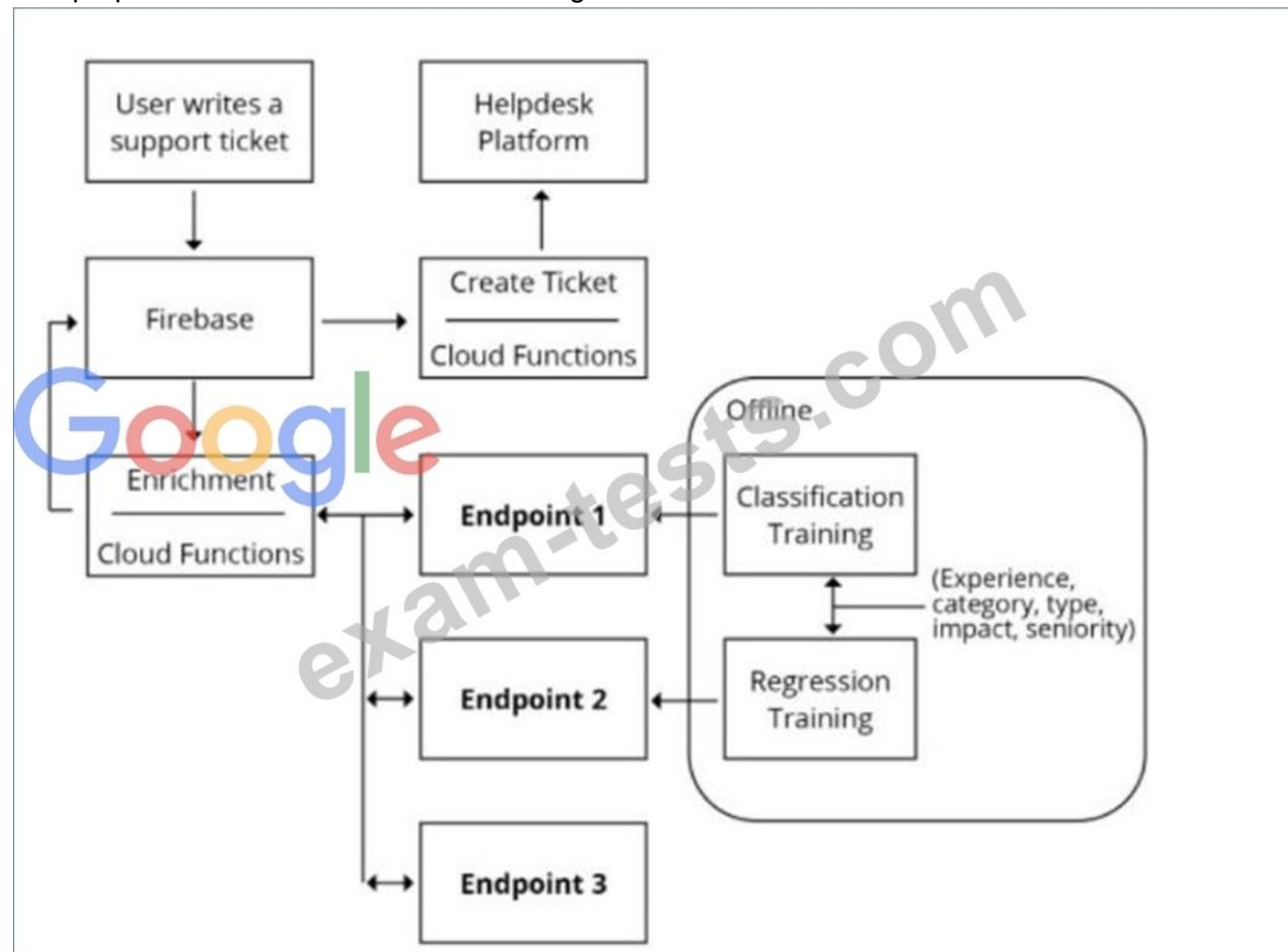
Answer: B (LEAVE A REPLY)

<https://cloud.google.com/bigquery-ml/docs/introduction>

NEW QUESTION: 20

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 (LEAVE A REPLY)

<https://cloud.google.com/architecture/architecture-of-a-serverless-ml-model#architecture> The architecture has the following flow:

A user writes a ticket to Firebase, which triggers a Cloud Function.

-The Cloud Function calls 3 different endpoints to enrich the ticket:

-An AI Platform endpoint, where the function can predict the priority.

-An AI Platform endpoint, where the function can predict the resolution time.

-The Natural Language API to do sentiment analysis and word salience.

-For each reply, the Cloud Function updates the Firebase real-time database.

-The Cloud Function then creates a ticket into the helpdesk platform using the RESTful API.

NEW QUESTION: 21

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 ([LEAVE A REPLY](#))

<https://www.oecd.org/finance/Impact-Big-Data-AI-in-the-Insurance-Sector.pdf>

<https://medium.com/artefact-engineering-and-data-science/including-ethics-best-practices-in-your-data-science-project-from-day-one-c15b26c2bf99>

NEW QUESTION: 22

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. Differential privacy federated learning, and explainability

C. Traceability, reproducibility, and explainability

D. Federated learning, reproducibility, and explainability

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 23

A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day, the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.

Which storage scheme is MOST adapted to this scenario?

A. Store datasets as tables in a multi-node Amazon Redshift cluster.

B. Store datasets as files in Amazon S3.

C. Store datasets as global tables in Amazon DynamoDB.

D. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 24

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

- A. Replace your test dataset with images of the newer products when they are introduced to retraining.
- B. Update your test dataset with images of the newer products when your evaluation metrics drop below a pre-decided threshold.
- C. Keep the original test dataset unchanged even if newer products are incorporated into retraining
- D. Extend your test dataset with images of the newer products when they are introduced to retraining

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 25

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. Dataprep
- B. Dataflow
- C. Cloud Data Fusion
- D. Apache Flink

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 26

You need to train a computer vision model that predicts the type of government ID present in a given image using a GPU-powered virtual machine on Compute Engine. You use the following parameters:

- * Optimizer: SGD
- * Image shape = 224x224
- * Batch size = 64
- * Epochs = 10
- * Verbose = 2

During training you encounter the following error: ResourceExhaustedError: out of Memory (oom) when allocating tensor. What should you do?

- A. Reduce the image shape
- B. Change the learning rate
- C. Change the optimizer
- D. Reduce the batch size

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 27

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. Recurrent Neural Networks (RNN)
- B. Reinforcement Learning
- C. Classification
- D. Convolutional Neural Networks (CNN)

Answer: [\(SHOW ANSWER\)](#)

NEW QUESTION: 28

An agency collects census information within a country to determine healthcare and social program needs by province and city. The census form collects responses for approximately 500 questions from each citizen.

Which combination of algorithms would provide the appropriate insights? (Select TWO.)

- A. The factorization machines (FM) algorithm
- B. The Latent Dirichlet Allocation (LDA) algorithm
- C. The principal component analysis (PCA) algorithm
- D. The k-means algorithm
- E. The Random Cut Forest (RCF) algorithm

Answer: C,D [\(LEAVE A REPLY\)](#)

The PCA and K-means algorithms are useful in collection of data using census form.

NEW QUESTION: 29

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. Add a model monitoring job where 10% of incoming predictions are sampled every hour.
- B. Add a model monitoring job where 90% of incoming predictions are sampled 24 hours.
- C. Add a model monitoring job where 10% of incoming predictions are sampled 24 hours.
- D. Implement continuous retraining of the model daily using Vertex AI Pipelines.

Answer: B [\(LEAVE A REPLY\)](#)

NEW QUESTION: 30

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. Federated learning, reproducibility, and explainability
- C. Traceability, reproducibility, and explainability
- D. Differential privacy federated learning, and explainability

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 31

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: ['driverslicense', '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: D ([LEAVE A REPLY](#))

se `sparse_categorical_crossentropy`. Examples for above 3-class classification problem: [1] , [2], [3]

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPASS.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (**300 Q&As Dumps, 40%OFF Special Discount: Exam-Tests**)

NEW QUESTION: 32

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. Use TensorFlow I/O's BigQuery Reader to directly read the data.
- C. Export data to CSV files in Cloud Storage, and use `tf.data.TextLineDataset()` to read them.
- D. Convert the data into TFRecords, and use `tf.data.TFRecordDataset()` to read them.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 33

You have a functioning end-to-end ML pipeline that involves tuning the hyperparameters of your ML model using AI Platform, and then using the best-tuned parameters for training. Hypertuning is taking longer than expected and is delaying the downstream processes. You want to speed up the tuning job without significantly compromising its effectiveness. Which actions should you take? Choose 2 answers

- A. Set the early stopping parameter to TRUE
- B. Decrease the range of floating-point values

- C. Change the search algorithm from Bayesian search to random search.
- D. Decrease the maximum number of trials during subsequent training phases.
- E. Decrease the number of parallel trials

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 34

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. Automate multiple training runs using Cloud Composer
- B. Create multiple models using AutoML Tables
- C. Run multiple training jobs on AI Platform with similar job names
- D. Create an experiment in Kubeflow Pipelines to organize multiple runs

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 35

A retail company is using Amazon Personalize to provide personalized product recommendations for its customers during a marketing campaign. The company sees a significant increase in sales of recommended items to existing customers immediately after deploying a new solution version, but these sales decrease a short time after deployment. Only historical data from before the marketing campaign is available for training.

How should a data scientist adjust the solution?

- A. Implement a new solution using the built-in factorization machines (FM) algorithm in Amazon SageMaker.
- B. Add user metadata and use the HRNN-Metadata recipe in Amazon Personalize.
- C. Add event type and event value fields to the interactions dataset in Amazon Personalize.
- D. Use the event tracker in Amazon Personalize to include real-time user interactions.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 36

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

AuthorA: Political Party A

TextA1: [SentenceA11, SentenceA12, SentenceA13, ...]

TextA2: [SentenceA21, SentenceA22, SentenceA23, ...]

...

AuthorB: Political Party B

TextB1: [SentenceB11, SentenceB12, SentenceB13, ...]

TextB2: [SentenceB21, SentenceB22, SentenceB23, ...]

...

AuthorC: Political Party B

TextC1: [SentenceC11, SentenceC12, SentenceC13, ...]

TextC2: [SentenceC21, SentenceC22, SentenceC23, ...]

...

AuthorD: Political Party A

TextD1: [SentenceD11, SentenceD12, SentenceD13, ...]

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

...

...

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: [TextC1, 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, ...]

A. Option B

B. Option A

C. Option C

D. Option D

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 37

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. F1 score

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

D. RMSE

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 38

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: A ([LEAVE A REPLY](#))

<https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/aad9f93b86b7addfea4c419b9100c6cdd26cacea.pdf>

NEW QUESTION: 39

A gaming company has launched an online game where people can start playing for free, but they need to pay if they choose to use certain features. The company needs to build an automated system to predict whether or not a new user will become a paid user within 1 year. The company has gathered a labeled dataset from 1 million users.

The training dataset consists of 1,000 positive samples (from users who ended up paying within 1 year) and 999,000 negative samples (from users who did not use any paid features). Each data sample consists of 200 features including user age, device, location, and play patterns.

Using this dataset for training, the Data Science team trained a random forest model that converged with over 99% accuracy on the training set. However, the prediction results on a test dataset were not satisfactory. Which of the following approaches should the Data Science team take to mitigate this issue? (Choose two.)

A. Include a copy of the samples in the test dataset in the training dataset.

B. Add more deep trees to the random forest to enable the model to learn more features.

C. Generate more positive samples by duplicating the positive samples and adding a small amount of noise to the duplicated data.

D. Change the cost function so that false positives have a higher impact on the cost value than false negatives.

E. Change the cost function so that false negatives have a higher impact on the cost value than false positives.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 40

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 ([LEAVE A REPLY](#))

Reference:

<https://developers.google.com/machine-learning/recommendation/collaborative/basics>

https://cloud.google.com/architecture/recommendations-using-machine-learning-on-compute-engine#filtering_the_data

NEW QUESTION: 41

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 model
- C. AutoML Vision Edge mobile-versatile-1 model
- D. AutoML Vision Edge mobile-low-latency-1 model

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 42

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. Create an automated workflow in Cloud Composer that runs daily and looks for changes in code in Cloud Storage using a sensor.
- B. Use Cloud Functions to identify changes to your code in Cloud Storage and trigger a retraining job
- C. Use the gcloud command-line tool to submit training jobs on AI Platform when you update your code
- D. Use Cloud Build linked with Cloud Source Repositories to trigger retraining when new code is pushed to the repository

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 43

A company ingests machine learning (ML) data from web advertising clicks into an Amazon S3 data lake. Click data is added to an Amazon Kinesis data stream by using the Kinesis Producer Library (KPL). The data is loaded into the S3 data lake from the data stream by using an Amazon Kinesis Data Firehose delivery stream.

As the data volume increases, an ML specialist notices that the rate of data ingested into Amazon S3 is relatively constant. There also is an increasing backlog of data for Kinesis Data Streams and Kinesis Data Firehose to ingest.

Which next step is MOST likely to improve the data ingestion rate into Amazon S3?

- A. Increase the number of S3 prefixes for the delivery stream to write to.
- B. Decrease the retention period for the data stream.
- C. Increase the number of shards for the data stream.
- D. Add more consumers using the Kinesis Client Library (KCL).

Answer: C ([LEAVE A REPLY](#))

Explanation/Reference:

NEW QUESTION: 44

A real estate company wants to create a machine learning model for predicting housing prices based on a historical dataset. The dataset contains 32 features.

Which model will meet the business requirement?

- A. Linear regression
- B. K-means
- C. Principal component analysis (PCA)
- D. Logistic regression

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 45

A manufacturing company has a large set of labeled historical sales data. The manufacturer would like to predict how many units of a particular part should be produced each quarter.

Which machine learning approach should be used to solve this problem?

- A. Logistic regression
- B. Principal component analysis (PCA)
- C. Linear regression
- D. Random Cut Forest (RCF)

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 46

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: D ([LEAVE A REPLY](#))

<https://www.tensorflow.org/tfx/serving/performance>

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPass.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)

NEW QUESTION: 47

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_enqueued_batches TensorFlow Serving parameter
- B. Switch to the tensorflow-model-server-universal version of TensorFlow Serving
- C. Recompile TensorFlow Serving using the source to support CPU-specific optimizations Instruct GKE to choose an appropriate baseline minimum CPU platform for serving nodes
- D. Significantly increase the max_batch_size TensorFlow Serving parameter

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 48

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 k-fold cross-validation as a validation strategy to ensure that your model is ready for production.
- B. Use the TFX ModelValidator tools to specify performance metrics for production readiness
- 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: B ([LEAVE A REPLY](#))

NEW QUESTION: 49

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. Deploy the model to a Vertex AI endpoint, and invoke this endpoint in the Dataflow job.
- B. Load the model directly into the Dataflow job as a dependency, and use it for prediction.
- C. Deploy the model in a TFServing container on Google Kubernetes Engine, and invoke it in the Dataflow job.

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

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 50

A Data Engineer needs to build a model using a dataset containing customer credit card information. How can the Data Engineer ensure the data remains encrypted and the credit card information is secure?

A. Use a custom encryption algorithm to encrypt the data and store the data on an Amazon SageMaker instance in a VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.

B. Use an IAM policy to encrypt the data on the Amazon S3 bucket and Amazon Kinesis to automatically discard credit card numbers and insert fake credit card numbers.

C. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VPC. Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.

D. Use AWS KMS to encrypt the data on Amazon S3 and Amazon SageMaker, and redact the credit card numbers from the customer data with AWS Glue.

Answer: C ([LEAVE A REPLY](#))

Explanation/Reference: <https://docs.aws.amazon.com/sagemaker/latest/dg/pca.html>

NEW QUESTION: 51

A machine learning (ML) specialist wants to secure calls to the Amazon SageMaker Service API. The specialist has configured Amazon VPC with a VPC interface endpoint for the Amazon SageMaker Service API and is attempting to secure traffic from specific sets of instances and IAM users. The VPC is configured with a single public subnet.

Which combination of steps should the ML specialist take to secure the traffic? (Choose two.)

A. Add a VPC endpoint policy to allow access to the IAM users.

B. Modify the users' IAM policy to allow access to Amazon SageMaker Service API calls only.

C. Modify the security group on the endpoint network interface to restrict access to the instances.

D. Modify the ACL on the endpoint network interface to restrict access to the instances.

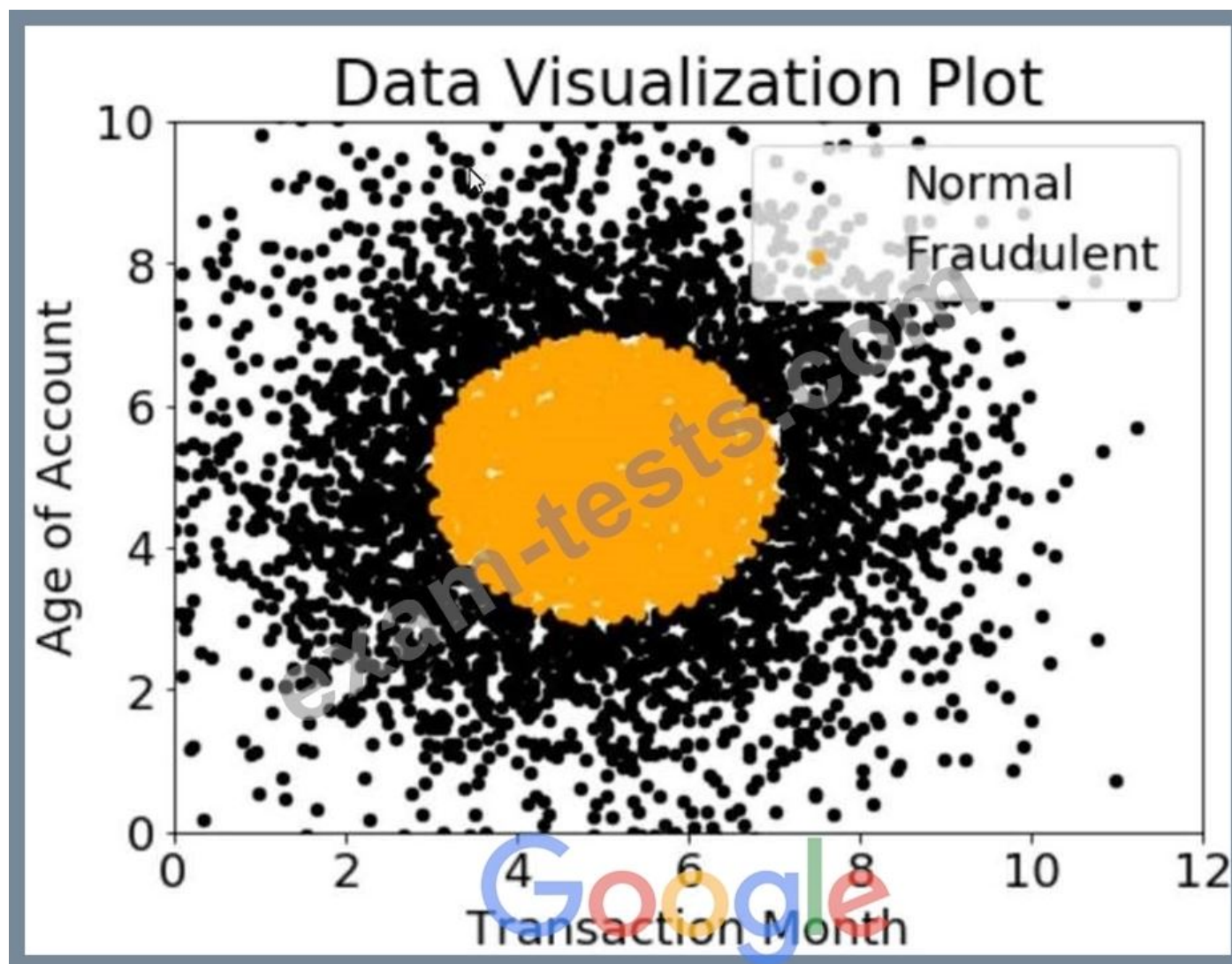
E. Add a SageMaker Runtime VPC endpoint interface to the VPC.

Answer: (SHOW ANSWER)

Explanation/Reference: <https://aws.amazon.com/blogs/machine-learning/private-package-installation-in-amazon-sagemaker-running-in-internet-free-mode/>

NEW QUESTION: 52

A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information, which model would have the HIGHEST accuracy?

- A. Logistic regression
- B. Long short-term memory (LSTM) model with scaled exponential linear unit (SELU)
- C. Single perceptron with tanh activation function
- D. Support vector machine (SVM) with non-linear kernel

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 53

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

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

- A. Use Amazon Rekognition's feature to detect text in an image to extract the data from scanned images. Use this information to process the loan applications.
- B. Configure Amazon Textract to route low-confidence predictions to Amazon Augmented AI (Amazon A2I). Perform a manual review on those words before performing a business validation.
- C. Use an Amazon Textract synchronous operation instead of an asynchronous operation.
- D. Configure Amazon Textract to route low-confidence predictions to Amazon SageMaker Ground Truth.

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

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 54

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. Use Dataflow with the SavedModel to read the data from BigQuery
- B. Submit a batch prediction job on AI Platform that points to the model location in Cloud Storage.
- C. Export the model to BigQuery ML.
- D. Deploy and version the model on AI Platform.

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 55

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 ([LEAVE A REPLY](#))

the model cannot work as-is as the classes to predict will likely not be the same; we need to use transfer learning to retrain the last layer and adapt it to the classes we need

NEW QUESTION: 56

You have been asked to develop an input pipeline for an ML training model that processes images from disparate sources at a low latency. You discover that your input data does not fit in memory. How should you create a dataset following Google-recommended best practices?

- A. Create a `tf.data.Dataset.prefetch` transformation
- B. Convert the images to `tf.Tensor` Objects, and then run `Dataset.from_tensor_slices()`.
- C. Convert the images to `tf.Tensor` Objects, and then run `tf.data.Dataset.from_tensors()`.
- D. Convert the images into TFRecords, store the images in Cloud Storage, and then use the `tf.data` API to read the images for training

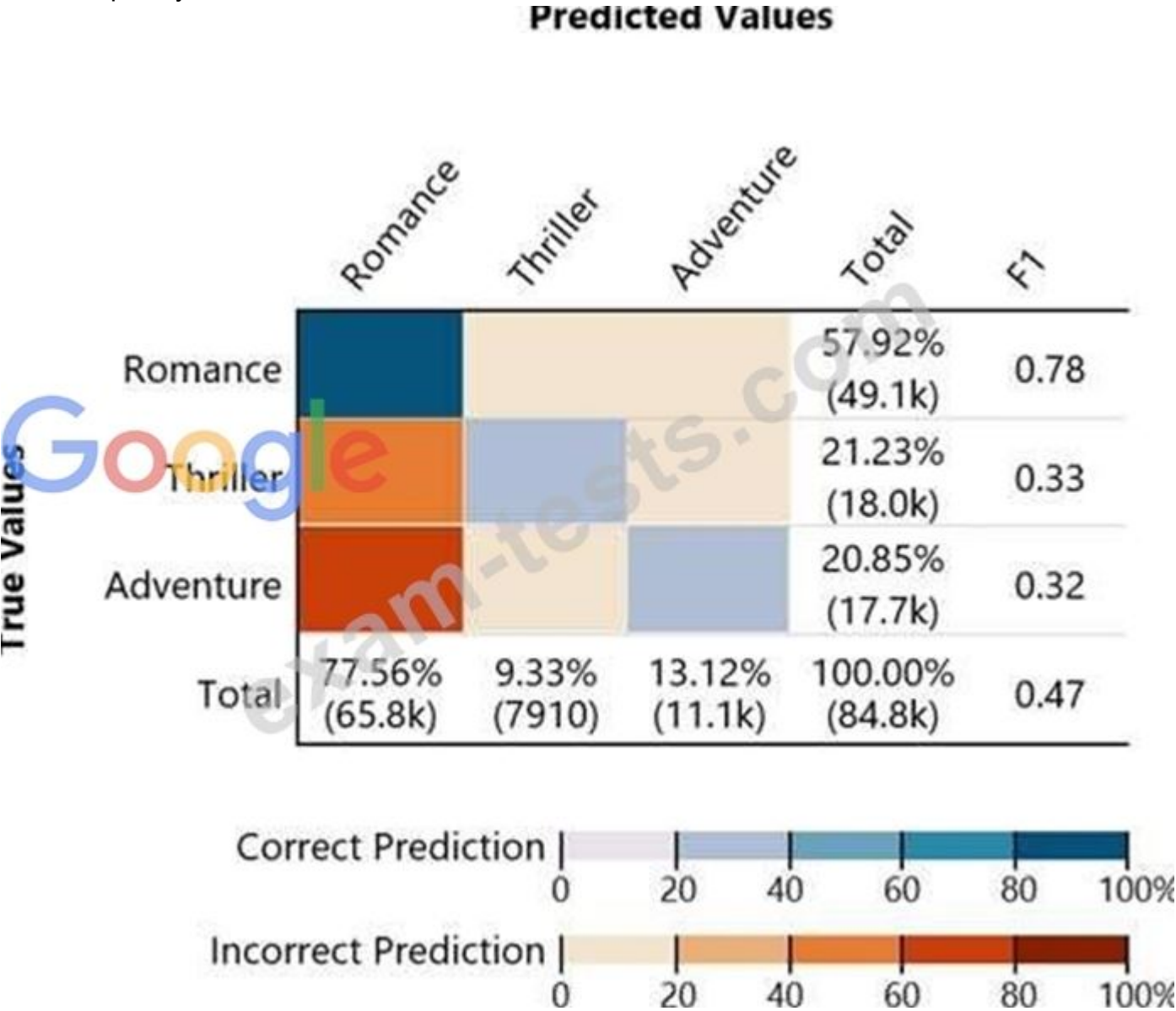
Answer: D ([LEAVE A REPLY](#))

Cite from Google Page: to construct a Dataset from data in memory, use `tf.data.Dataset.from_tensors()` or `tf.data.Dataset.from_tensor_slices()`. When input data is stored in a file (not in memory), the recommended TFRecord format, you can use `tf.data.TFRecordDataset()`. `tf.data.Dataset` is for data in memory. `tf.data.TFRecordDataset` is for data in non-memory storage. <https://cloud.google.com/architecture/ml-on-gcp-best-practices#store-image-video-audio-and-unstructured-data-on-cloud-storage>
" Store image, video, audio and unstructured data on Cloud Storage Store these data in large container formats on Cloud Storage. This applies to sharded TFRecord files if you're using TensorFlow, or Avro files if you're using any other framework. Combine many

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

NEW QUESTION: 57

Given the following confusion matrix for a movie classification model, what is the true class frequency for Romance and the predicted class frequency for Adventure?



- A. The true class frequency for Romance is 0.78 and the predicted class frequency for Adventure is (0.47-0.32)
- B. The true class frequency for Romance is 77.56% and the predicted class frequency for Adventure is 20.85%
- C. The true class frequency for Romance is 57.92% and the predicted class frequency for Adventure is 13.12%
- D. The true class frequency for Romance is 77.56% * 0.78 and the predicted class frequency for Adventure is 20.85%*0.32

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 58

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. Use one-hot encoding on all categorical features.
- B. Oversample the fraudulent transaction 10 times.
- C. Z-normalize all the numeric features.
- D. Write your data in TFRecords.

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 59

You work at a subscription-based company. You have trained an ensemble of trees and neural networks to predict customer churn, which is the likelihood that customers will not renew their yearly subscription. The average prediction is a 15% churn rate, but for a particular customer the model predicts that they are 70% likely to churn. The customer has a product usage history of 30%, is located in New York City, and became a customer in 1997. You need to explain the difference between the actual prediction, a 70% churn rate, and the average prediction. You want to use Vertex Explainable AI. What should you do?

- A. Configure sampled Shapley explanations on Vertex Explainable AI.
- B. Measure the effect of each feature as the weight of the feature multiplied by the feature value.
- C. Train local surrogate models to explain individual predictions.
- D. Configure integrated gradients explanations on Vertex Explainable AI.

Answer: ([SHOW ANSWER](#)**)**

NEW QUESTION: 60

A Machine Learning Specialist trained a regression model, but the first iteration needs optimizing. The Specialist needs to understand whether the model is more frequently overestimating or underestimating the target.

What option can the Specialist use to determine whether it is overestimating or underestimating the target value?

- A. Residual plots
- B. Confusion matrix
- C. Root Mean Square Error (RMSE)
- D. Area under the curve

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 61

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 BigQuery ML to build and test a logistic regression model to classify incoming requests. Use BigQuery ML to perform inference.
- B. Create a TensorFlow model using Google's BERT pre-trained model. Build and test a classifier, and deploy the model using Vertex AI.

- C. Use AutoML Natural Language to build and test a classifier. Deploy the model as a REST API.
- D. Use the Cloud Natural Language API to obtain metadata to classify the incoming cases.

Answer: C ([LEAVE A REPLY](#))

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPASS.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)

NEW QUESTION: 62

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: ([SHOW ANSWER](#)**)**

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

- NN models needs features with close ranges
- SGD converges well using features in [0, 1] scale
- The question specifically mention "different ranges"

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

NEW QUESTION: 63

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: B ([LEAVE A REPLY](#))

In this case, the model has a high accuracy of 99% for identifying customers who cancel their subscriptions, but a lower accuracy of 82% for identifying customers who renew their subscriptions. However, this does not necessarily mean that the model is performing

well, because 90% of the customers renew their subscription, so if the model always predicts that customers will renew, it will be correct 90% of the time. Therefore, the model's performance is worse than the baseline of always predicting that customers will renew their subscription.

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

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

NEW QUESTION: 64

A Machine Learning Specialist working for an online fashion company wants to build a data ingestion solution for the company's Amazon S3-based data lake.

The Specialist wants to create a set of ingestion mechanisms that will enable future capabilities comprised of:

- * Real-time analytics
- * Interactive analytics of historical data
- * Clickstream analytics
- * Product recommendations

Which services should the Specialist use?

- A.** AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for real-time data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
- B.** Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for near-real-time data insights; Amazon Kinesis Data Firehose for clickstream analytics; AWS Glue to generate personalized product recommendations
- C.** AWS Glue as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon Kinesis Data Firehose for delivery to Amazon ES for clickstream analytics; Amazon EMR to generate personalized product recommendations
- D.** Amazon Athena as the data catalog; Amazon Kinesis Data Streams and Amazon Kinesis Data Analytics for historical data insights; Amazon DynamoDB streams for clickstream analytics; AWS Glue to generate personalized product recommendations

Answer: ([SHOW ANSWER](#))

Explanation

NEW QUESTION: 65

A Marketing Manager at a pet insurance company plans to launch a targeted marketing campaign on social media to acquire new customers. Currently, the company has the following data in Amazon Aurora:

- * Profiles for all past and existing customers
- * Profiles for all past and existing insured pets
- * Policy-level information
- * Premiums received
- * Claims paid

What steps should be taken to implement a machine learning model to identify potential new customers on social media?

- A.** Use regression on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
- B.** Use clustering on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media

C. Use a decision tree classifier engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.

D. Use a recommendation engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 66

A Data Scientist wants to gain real-time insights into a data stream of GZIP files.

Which solution would allow the use of SQL to query the stream with the LEAST latency?

A. Amazon Kinesis Data Analytics with an AWS Lambda function to transform the data.

B. AWS Glue with a custom ETL script to transform the data.

C. An Amazon Kinesis Client Library to transform the data and save it to an Amazon ES cluster.

D. Amazon Kinesis Data Firehose to transform the data and put it into an Amazon S3 bucket.

Answer: **A** ([LEAVE A REPLY](#))

Explanation/Reference: <https://aws.amazon.com/big-data/real-time-analytics-featured-partners/>

NEW QUESTION: 67

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

A. Reduce the dimensions of the images used un the model

B. Reduce the number of layers in the model architecture

C. Configure your model to use bfloat16 instead float32

D. Reduce the global batch size from 1024 to 256

Answer: **D** ([LEAVE A REPLY](#))

NEW QUESTION: 68

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. Kubeflow Pipelines and App Engine

B. Cloud Composer, BigQuery ML , and AI Platform Prediction

C. Cloud Composer, AI Platform Training with custom containers , and App Engine

D. Kubeflow Pipelines and AI Platform Prediction

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 69

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: ([SHOW ANSWER](#))

Pub/Sub is a messaging service that can be used to stream data into BigQuery in real-time. Configuring Pub/Sub to stream the user activity data into BigQuery would ensure real-time ingestion of the data. Source: Google Cloud

NEW QUESTION: 70

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

Which techniques should be used to meet these requirements?

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

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 71

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

- A. Increase the size of the training batch
- B. Decrease the size of the training batch
- C. Increase the learning rate hyperparameter
- D. Decrease the learning rate hyperparameter

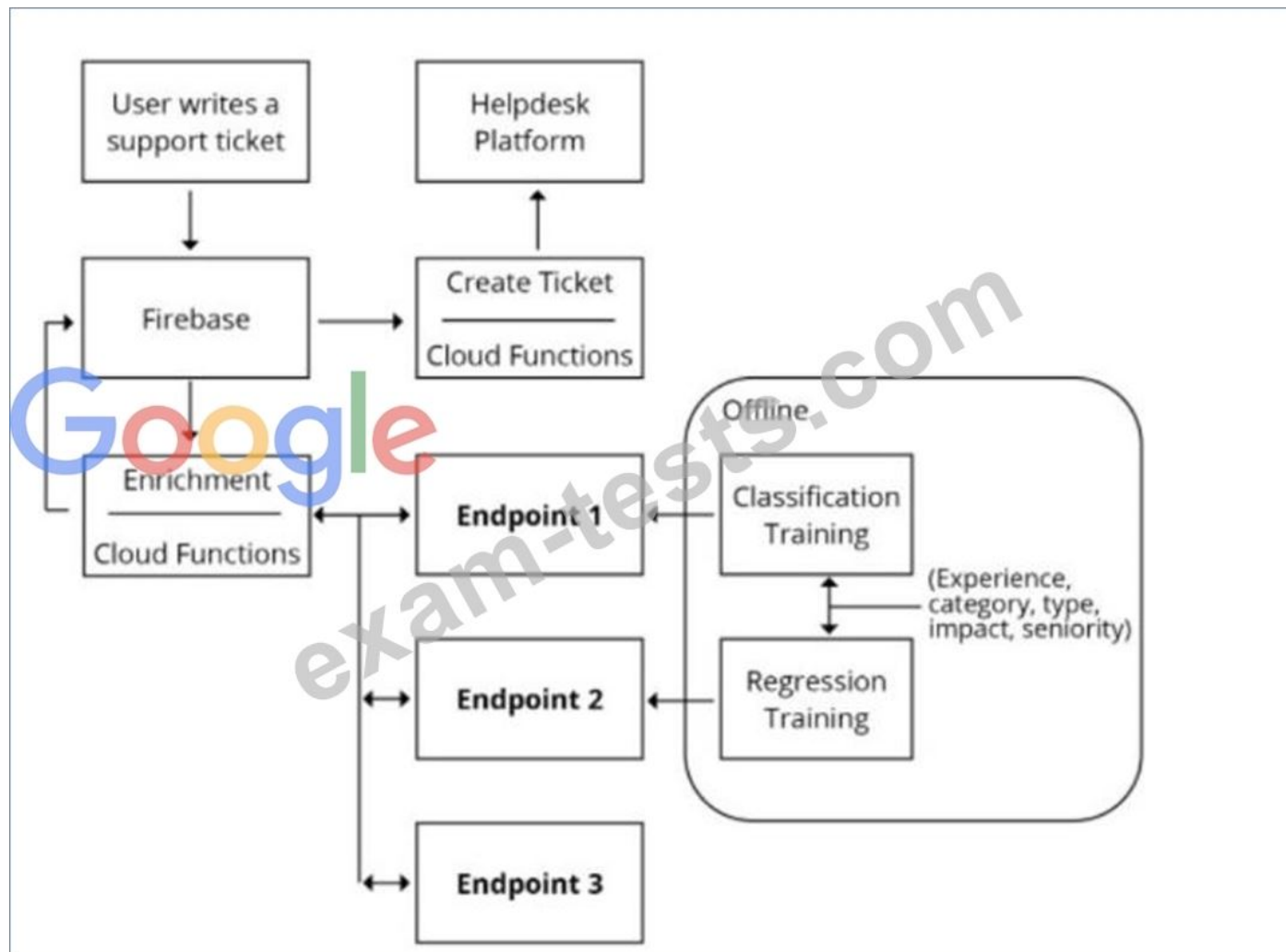
Answer: ([SHOW ANSWER](#))

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

NEW QUESTION: 72

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

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 73

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. 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.
- B. 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.
- C. 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.

D. 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.

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 74

A Data Scientist needs to analyze employment data. The dataset contains approximately 10 million observations on people across 10 different features. During the preliminary analysis, the Data Scientist notices that income and age distributions are not normal. While income levels shows a right skew as expected, with fewer individuals having a higher income, the age distribution also show a right skew, with fewer older individuals participating in the workforce.

Which feature transformations can the Data Scientist apply to fix the incorrectly skewed data? (Choose two.)

- A. One hot encoding
- B. High-degree polynomial transformation
- C. Cross-validation
- D. Logarithmic transformation
- E. Numerical value binning

Answer: C,E ([LEAVE A REPLY](#))

NEW QUESTION: 75

You manage a team of data scientists who use a cloud-based backend system to submit training jobs. This system has become very difficult to administer, and you want to use a managed service instead. The data scientists you work with use many different frameworks, including Keras, PyTorch, theano. Scikit-team, and custom libraries. What should you do?

- A. Use the 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 TFJob
- C. Create a library of VM images on Compute Engine; and publish these images on a centralized repository
- D. Set up Slurm workload manager to receive jobs that can be scheduled to run on your cloud infrastructure.

Answer: A ([LEAVE A REPLY](#))

because AI platform supported all the frameworks mentioned. And Kubeflow is not managed service in GCP.

<https://cloud.google.com/ai-platform/training/docs/getting-started-pytorch>

https://cloud.google.com/ai-platform/training/docs/containers-overview#advantages_of_custom_containers Use the ML framework of your choice. If you can't find an AI Platform Training runtime version that supports the ML framework you want to use, then you can build a custom container that installs your chosen framework and use it to run jobs on AI Platform Training.

NEW QUESTION: 76

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. 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
- 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. 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.
- 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: D ([LEAVE A REPLY](#))

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPASS.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)

NEW QUESTION: 77

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.** User engagement as measured by the number of battles played daily per user
- B.** Average time players wait before being assigned to a team
- C.** Precision and recall of assigning players to teams based on their predicted versus actual ability
- D.** Rate of return as measured by additional revenue generated minus the cost of developing a new model

Answer: ([SHOW ANSWER](#)**)**

NEW QUESTION: 78

A Machine Learning Specialist built an image classification deep learning model. However, the Specialist ran into an overfitting problem in which the training and testing accuracies were 99% and 75%, respectively.

How should the Specialist address this issue and what is the reason behind it?

- A.** The learning rate should be increased because the optimization process was trapped at a local minimum.
- B.** The dropout rate at the flatten layer should be increased because the model is not generalized enough.
- C.** The dimensionality of dense layer next to the flatten layer should be increased because the model is not complex enough.

D. The epoch number should be increased because the optimization process was terminated before it reached the global minimum.

Answer: D ([LEAVE A REPLY](#))

Explanation/Reference: https://www.tensorflow.org/tutorials/keras/overfit_and_underfit

NEW QUESTION: 79

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.** Add parallel interleave to the pipeline
- D.** Increase the network bandwidth

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 80

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 ([LEAVE A REPLY](#))

(<https://www.kubeflow.org/docs/components/pipelines/overview/pipelines-overview/>

<https://medium.com/google-cloud/how-to-build-an-end-to-end-propensity-to-purchase-solution-using-bigquery-ml-and-kubeflow-pipelines-cd4161f734d9#75c7>

NEW QUESTION: 81

You work for a credit card company and have been asked to create a custom fraud detection model based on historical data using AutoML Tables. You need to prioritize detection of fraudulent transactions while minimizing false positives. Which optimization objective should you use when training the model?

- A.** An optimization objective that minimizes Log loss
- B.** An optimization objective that maximizes the Precision at a Recall value of 0.50
- C.** An optimization objective that maximizes the area under the precision-recall curve (AUC PR) value
- D.** An optimization objective that maximizes the area under the receiver operating characteristic curve (AUC ROC) value

Answer: C ([LEAVE A REPLY](#))

<https://stats.stackexchange.com/questions/262616/roc-vs-precision-recall-curves-on-imbalanced-dataset>

<https://neptune.ai/blog/f1-score-accuracy-roc-auc-pr-auc>

https://icaiit.org/proceedings/6th_ICAIIT/1_3Fayzrakhmanov.pdf The problem of fraudulent transactions detection, which is an imbalanced classification problem (most transactions are not fraudulent), you want to maximize both precision and recall; so the area

under the PR curve. As a matter of fact, the question asks you to focus on detecting fraudulent transactions (maximize true positive rate, a.k.a. Recall) while minimizing false positives (a.k.a. maximizing Precision). Another way to see it is this: for imbalanced problems like this one you'll get a lot of true negatives even from a bad model (it's easy to guess a transaction as "non-fraudulent" because most of them are!), and with high TN the ROC curve goes high fast, which would be misleading. So you wanna avoid dealing with true negatives in your evaluation, which is precisely what the PR curve allows you to do.

NEW QUESTION: 82

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

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



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: [TextC1, 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, SentenceC13, SentenceD31, ...]

D)

Distribute paragraphs of texts (or chunks of consecutive sentences) across the train-test-eval subsets:

Train set: [SentenceA11, SentenceA12, SentenceB11, SentenceD12, ...]

Test set: [SentenceA13, SentenceB13, SentenceB21, SentenceD23, SentenceC12, SentenceD13, ...]

Eval set: [SentenceA11, SentenceA22, SentenceB13, SentenceD22, SentenceC23, SentenceD11, ...]

A. Option A

B. Option B

C. Option C

D. Option D

Answer: B ([LEAVE A REPLY](#))

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

<https://cloud.google.com/automl-tables/docs/prepare#ml-use>

NEW QUESTION: 83

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. 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.

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

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

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

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 84

A Machine Learning Specialist is developing a daily ETL workflow containing multiple ETL jobs. The workflow consists of the following processes:

* Start the workflow as soon as data is uploaded to Amazon S3.

* When all the datasets are available in Amazon S3, start an ETL job to join the uploaded datasets with multiple terabyte-sized datasets already stored in Amazon S3.

* Store the results of joining datasets in Amazon S3.

* If one of the jobs fails, send a notification to the Administrator.

Which configuration will meet these requirements?

A. Use AWS Lambda to trigger an AWS Step Functions workflow to wait for dataset uploads to complete in Amazon S3. Use AWS Glue to join the datasets. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.

B. Develop the ETL workflow using AWS Lambda to start an Amazon SageMaker notebook instance. Use a lifecycle configuration script to join the datasets and persist the results in Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.

C. Develop the ETL workflow using AWS Batch to trigger the start of ETL jobs when data is uploaded to Amazon S3. Use AWS Glue to join the datasets in Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.

D. Use AWS Lambda to chain other Lambda functions to read and join the datasets in Amazon S3 as soon as the data is uploaded to Amazon S3. Use an Amazon CloudWatch alarm to send an SNS notification to the Administrator in the case of a failure.

Answer: A ([LEAVE A REPLY](#))

Explanation/Reference: <https://aws.amazon.com/step-functions/use-cases/>

NEW QUESTION: 85

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 using Vertex AI Training with GPUs.

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

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

Answer: ([SHOW ANSWER](#)**)**

NEW QUESTION: 86

A company that promotes healthy sleep patterns by providing cloud-connected devices currently hosts a sleep tracking application on AWS. The application collects device usage information from device users. The company's Data Science team is building a machine learning model to predict if and when a user will stop utilizing the company's devices. Predictions from this model are used by a downstream application that determines the best approach for contacting users.

The Data Science team is building multiple versions of the machine learning model to evaluate each version against the company's business goals. To measure long-term effectiveness, the team wants to run multiple versions of the model in parallel for long periods of time, with the ability to control the portion of inferences served by the models.

Which solution satisfies these requirements with MINIMAL effort?

A. Build and host multiple models in Amazon SageMaker. Create a single endpoint that accesses multiple models. Use Amazon SageMaker batch transform to control invoking the different models through the single endpoint.

B. Build and host multiple models in Amazon SageMaker. Create an Amazon SageMaker endpoint configuration with multiple production variants. Programmatically control the portion of the inferences served by the multiple models by updating the endpoint configuration.

C. Build and host multiple models in Amazon SageMaker. Create multiple Amazon SageMaker endpoints, one for each model. Programmatically control invoking different models for inference at the application layer.

D. Build and host multiple models in Amazon SageMaker Neo to take into account different types of medical devices. Programmatically control which model is invoked for inference based on the medical device type.

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 87

A data scientist needs to identify fraudulent user accounts for a company's ecommerce platform. The company wants the ability to determine if a newly created account is associated with a previously known fraudulent user.

The data scientist is using AWS Glue to cleanse the company's application logs during ingestion.

Which strategy will allow the data scientist to identify fraudulent accounts?

A. Execute the built-in FindDuplicates Amazon Athena query.

B. Create a FindMatches machine learning transform in AWS Glue.

C. Create an AWS Glue crawler to infer duplicate accounts in the source data.

D. Search for duplicate accounts in the AWS Glue Data Catalog.

Answer: B ([LEAVE A REPLY](#))

Explanation/Reference: <https://docs.aws.amazon.com/glue/latest/dg/machine-learning.html>

NEW QUESTION: 88

A large company has developed a BI application that generates reports and dashboards using data collected from various operational metrics. The company wants to provide executives with an enhanced experience so they can use natural language to get data from the reports. The company wants the executives to be able ask questions using written and spoken interfaces.

Which combination of services can be used to build this conversational interface? (Choose three.)

A. Amazon Comprehend

B. Amazon Lex

C. Amazon Polly

D. Amazon Connect

E. Amazon Transcribe

F. Alexa for Business

Answer: (SHOW ANSWER)

NEW QUESTION: 89

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 a convolutional neural network with max pooling and softmax activation

B. Remove negative examples until the numbers of positive and negative examples are equal

C. Downsample the data with upweighting to create a sample with 10% positive examples

D. Use the class distribution to generate 10% positive examples

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 90

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. Create a Managed Instance Group with autoscaling
- B. Create a cluster on Dataproc for training
- C. Use Kubeflow Pipelines to train on a Google Kubernetes Engine cluster.
- D. Use AI Platform for distributed training

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 91

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 ([LEAVE A REPLY](#))

<https://machinelearningmastery.com/introduction-to-regularization-to-reduce-overfitting-and-improve-generalization-error/>

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPass.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)

NEW QUESTION: 92

You need to train a computer vision model that predicts the type of government ID present in a given image using a GPU-powered virtual machine on Compute Engine. You use the following parameters:

- * Optimizer: SGD
- * Image shape = 224x224
- * Batch size = 64

* Epochs = 10

* Verbose = 2

During training you encounter the following error: ResourceExhaustedError: out of Memory (oom) when allocating tensor. What should you do?

- A. Change the learning rate
- B. Reduce the image shape
- C. Change the optimizer
- D. Reduce the batch size

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 93

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: ([SHOW ANSWER](#))

BigQuery ML supports supervised learning with the logistic regression model type.

NEW QUESTION: 94

A Machine Learning Specialist is building a model that will perform time series forecasting using Amazon SageMaker. The Specialist has finished training the model and is now planning to perform load testing on the endpoint so they can configure Auto Scaling for the model variant.

Which approach will allow the Specialist to review the latency, memory utilization, and CPU utilization during the load test?

- A. Review SageMaker logs that have been written to Amazon S3 by leveraging Amazon Athena and Amazon QuickSight to visualize logs as they are being produced.
- B. Generate an Amazon CloudWatch dashboard to create a single view for the latency, memory utilization, and CPU utilization metrics that are outputted by Amazon SageMaker.
- C. Build custom Amazon CloudWatch Logs and then leverage Amazon ES and Kibana to query and visualize the log data as it is generated by Amazon SageMaker.
- D. Send Amazon CloudWatch Logs that were generated by Amazon SageMaker to Amazon ES and use Kibana to query and visualize the log data.

Answer: ([SHOW ANSWER](#))

Explanation/Reference: <https://docs.aws.amazon.com/sagemaker/latest/dg/monitoring-cloudwatch.html>

NEW QUESTION: 95

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 ([LEAVE A REPLY](#))

AutoML Natural Language is a service that allows you to quickly build, test and deploy natural language processing (NLP) models without needing to have expertise in NLP or machine learning. You can use it to train a classifier on your corpus of written support cases, and then use the AutoML API to perform classification on new requests. Once the model is trained, it can be deployed as a REST API. This allows the classifier to be integrated into your pipeline and be easily consumed by other systems.

NEW QUESTION: 96

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: ([SHOW ANSWER](#)**)**

https://cloud.google.com/data-fusion/docs/concepts/overview#using_the_code-free_web_ui

NEW QUESTION: 97

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 into TFRecords, store the images in Cloud Storage, and then use the `tf.data` API to read the images for training
- C. Convert the images to `tf.Tensor` Objects, and then run `tf.data.Dataset.from_tensors()`.
- D. Convert the images to `tf.Tensor` Objects, and then run `Dataset.from_tensor_slices()`.

Answer: ([SHOW ANSWER](#)**)**

NEW QUESTION: 98

A Machine Learning Specialist is preparing data for training on Amazon SageMaker. The Specialist is using one of the SageMaker built-in algorithms for the training. The dataset is stored in .CSV format and is transformed into a `numpy.array`, which appears to be negatively affecting the speed of the training.

What should the Specialist do to optimize the data for training on SageMaker?

- A. Use the SageMaker batch transform feature to transform the training data into a DataFrame.
- B. Use the SageMaker hyperparameter optimization feature to automatically optimize the data.
- C. Use AWS Glue to compress the data into the Apache Parquet format.
- D. Transform the dataset into the RecordIO protobuf format.

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 99

A Machine Learning Specialist is implementing a full Bayesian network on a dataset that describes public transit in New York City. One of the random variables is discrete, and represents the number of minutes New Yorkers wait for a bus given that the buses cycle every 10 minutes, with a mean of 3 minutes.

Which prior probability distribution should the ML Specialist use for this variable?

- A. Uniform distribution
- B. Poisson distribution
- C. Normal distribution
- D. Binomial distribution

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 100

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 8 GPUs with all dependencies installed manually.
- B. A Deep Learning VM with more powerful CPU e2-highcpu-16 machines with all libraries pre-installed.
- C. A Deep Learning VM with an n1-standard-2 machine and 1 GPU with all libraries pre-installed.
- D. AVM on Compute Engine and 1 TPU with all dependencies installed manually.

Answer: D ([LEAVE A REPLY](#))

NEW QUESTION: 101

A financial services company is building a robust serverless data lake on Amazon S3. The data lake should be flexible and meet the following requirements:

- * Support querying old and new data on Amazon S3 through Amazon Athena and Amazon Redshift Spectrum.
- * Support event-driven ETL pipelines
- * Provide a quick and easy way to understand metadata

Which approach meets these requirements?

- A. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Glue ETL job, and an external Apache Hive metastore to search and discover metadata.
- B. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Batch job, and an external Apache Hive metastore to search and discover metadata.

- C. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.
- D. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 102

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

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



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: [TextC1, 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, SentenceC13, SentenceD31, ...]

D)

Distribute paragraphs of texts (or chunks of consecutive sentences) across the train-test-eval subsets:

Train set: [SentenceA11, SentenceA12, SentenceB11, SentenceD12, ...]

Test set: [SentenceA13, SentenceB13, SentenceB21, SentenceD23, SentenceC12, SentenceD13, ...]

Eval set: [SentenceA11, SentenceA22, SentenceB13, SentenceD22, SentenceC23, SentenceD11, ...]

A. Option A

B. Option D

C. Option C

D. Option B

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 103

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

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

C. 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

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

Answer: ([SHOW ANSWER](#)**)**

NEW QUESTION: 104

A Machine Learning Specialist at a company sensitive to security is preparing a dataset for model training. The dataset is stored in Amazon S3 and contains Personally Identifiable Information (PII).

The dataset:

* Must be accessible from a VPC only.

* Must not traverse the public internet.

How can these requirements be satisfied?

A. Create a VPC endpoint and apply a bucket access policy that restricts access to the given VPC endpoint and the VPC.

B. Create a VPC endpoint and apply a bucket access policy that allows access from the given VPC endpoint and an Amazon EC2 instance.

C. Create a VPC endpoint and use Network Access Control Lists (NACLs) to allow traffic between only the given VPC endpoint and an Amazon EC2 instance.

D. Create a VPC endpoint and use security groups to restrict access to the given VPC endpoint and an Amazon EC2 instance

Answer: B ([LEAVE A REPLY](#))

Explanation/Reference: <https://docs.aws.amazon.com/AmazonS3/latest/dev/example-bucket-policies-vpc-endpoint.html>

NEW QUESTION: 105

A machine learning specialist is running an Amazon SageMaker endpoint using the built-in object detection algorithm on a P3 instance for real-time predictions in a company's production application. When evaluating the model's resource utilization, the specialist notices that the model is using only a fraction of the GPU.

Which architecture changes would ensure that provisioned resources are being utilized effectively?

- A. Deploy the model onto an Amazon Elastic Container Service (Amazon ECS) cluster using a P3 instance.
- B. Redeploy the model on an M5 instance. Attach Amazon Elastic Inference to the instance.
- C. Redeploy the model as a batch transform job on an M5 instance.
- D. Redeploy the model on a P3dn instance.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 106

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. 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
- C. Convert your PySpark into SparkSQL queries to transform the data and then run your pipeline on Dataproc to write the data into BigQuery.
- D. 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

Answer: C ([LEAVE A REPLY](#))

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPass.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)

NEW QUESTION: 107

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. Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs without checkpoints.
- C. Migrate to training with Kuberflow on Google Kubernetes Engine, and use preemptible VMs with checkpoints.
- D. Use AI Platform to run distributed training jobs without checkpoints.

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 108

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 Cloud Composer for the orchestration.
- B. Use Vertex AI Pipelines with TensorFlow Extended (TFX) SDK.
- C. Use Kubeflow Pipelines on Google Kubernetes Engine.
- D. Use Vertex AI Pipelines with Kubeflow Pipelines SDK.

Answer: C ([LEAVE A REPLY](#))

NEW QUESTION: 109

Your data science team is training a PyTorch model for image classification based on a pre-trained ResNet 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. Convert the model to a TensorFlow model, and run a hyperparameter tuning job on AI Platform.
- D. Create a Kuberflow Pipelines instance, and run a hyperparameter tuning job on Katib.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 110

A logistics company needs a forecast model to predict next month's inventory requirements for a single item in 10 warehouses. A machine learning specialist uses Amazon Forecast to develop a forecast model from 3 years of monthly data. There is no missing data. The specialist selects the DeepAR+ algorithm to train a predictor. The predictor means absolute percentage error (MAPE) is much larger than the MAPE produced by the current human forecasters. Which changes to the CreatePredictor API call could improve the MAPE? (Choose two.)

- A. Set PerformAutoML to true.
- B. Set ForecastHorizon to 4.
- C. Set ForecastFrequency to W for weekly.
- D. Set PerformHPO to true.
- E. Set FeaturizationMethodName to filling.

Answer: C,D ([LEAVE A REPLY](#))

Explanation/Reference: <https://docs.aws.amazon.com/forecast/latest/dg/forecast.dg.pdf>

NEW QUESTION: 111

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

Choose 2 answers

- A. Reduce the value of the repeat parameter
- B. Use the interleave option for reading data
- C. Decrease the batch size argument in your transformation
- D. Increase the buffer size for the shuffle option.
- E. Set the prefetch option equal to the training batch size

Answer: B,E ([LEAVE A REPLY](#))

NEW QUESTION: 112

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

Choose 2 answers

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

Answer: C,D ([LEAVE A REPLY](#))

NEW QUESTION: 113

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 ([LEAVE A REPLY](#))

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

<https://github.com/kubeflow/pipelines/blob/master/components/gcp/bigquery/query/sample.ipynb>

; <https://v0-5.kubeflow.org/docs/pipelines/reusable-components/>

NEW QUESTION: 114

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 ([LEAVE A REPLY](#))

Google has bought this software and support for this tool is not good. SQL can work in Cloud fusion pipelines too but I would prefer to use a single tool like Bigquery to both transform and store data.

NEW QUESTION: 115

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.** Apply feature crossing with another column that does not have missing values.
- B.** Predict the missing values using linear regression.
- C.** Replace the missing values with zeros.
- D.** Delete the rows that have missing values.

Answer: ([SHOW ANSWER](#)**)**

NEW QUESTION: 116

You are building a linear model with over 100 input features, all with values between -1 and 1. You suspect that many features are non-informative. You want to remove the non-informative features from your model while keeping the informative ones in their original form. Which technique should you use?

- A.** Use Principal Component Analysis to eliminate the least informative features.
- B.** Use L1 regularization to reduce the coefficients of uninformative features to 0.
- C.** After building your model, use Shapley values to determine which features are the most informative.
- D.** Use an iterative dropout technique to identify which features do not degrade the model when removed.

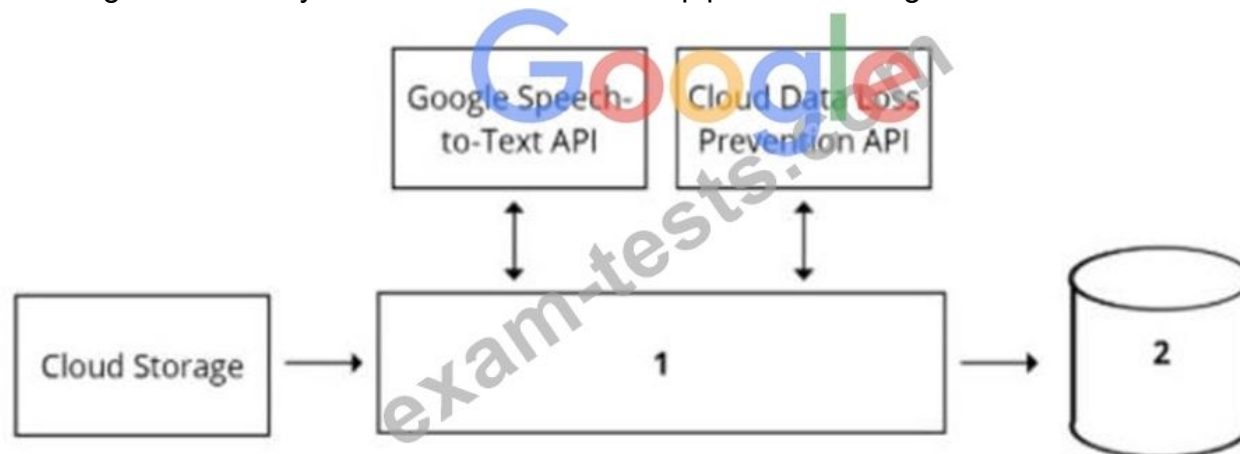
Answer: B ([LEAVE A REPLY](#))

<https://cloud.google.com/ai-platform/prediction/docs/ai-explanations/overview#sampled-shapley>

NEW QUESTION: 117

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 ([LEAVE A REPLY](#))

<https://github.com/GoogleCloudPlatform/dataflow-contact-center-speech-analysis>

NEW QUESTION: 118

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 BigQueryAPI.
- 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 ([LEAVE A REPLY](#))

<https://codelabs.developers.google.com/codelabs/cloud-kubeflow-pipelines-gis> Kubeflow 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.

<https://www.kubeflow.org/docs/components/pipelines/introduction/#what-is-kubeflow-pipelines>

"Kubeflow Pipelines supports the export of scalar metrics. You can write a list of metrics to a local file to describe the performance of the model. The pipeline agent uploads the local file as your run-time metrics. You can view the uploaded metrics as a visualization in the Runs page for a particular experiment in the Kubeflow Pipelines UI."

<https://www.kubeflow.org/docs/components/pipelines/sdk/pipelines-metrics/>

NEW QUESTION: 119

You need to build an ML model for a social media application to predict whether a user's submitted profile photo meets the requirements. The application will inform the user if the picture meets the requirements. How should you build a model to ensure that the application does not falsely accept a non-compliant picture?

- A. Use Vertex AI Workbench user-managed notebooks to build a custom model that has three times as many examples of pictures that do not meet the profile photo requirements.
- B. Use Vertex AI Workbench user-managed notebooks to build a custom model that has three times as many examples of pictures that meet the profile photo requirements.
- C. Use AutoML to optimize the model's recall in order to minimize false negatives.
- D. Use AutoML to optimize the model's F1 score in order to balance the accuracy of false positives and false negatives.

Answer: B ([LEAVE A REPLY](#))

NEW QUESTION: 120

A company is using Amazon Polly to translate plaintext documents to speech for automated company announcements. However, company acronyms are being mispronounced in the current documents.

How should a Machine Learning Specialist address this issue for future documents?

- A. Convert current documents to SSML with pronunciation tags.
- B. Create an appropriate pronunciation lexicon.
- C. Output speech marks to guide in pronunciation.
- D. Use Amazon Lex to preprocess the text files for pronunciation

Answer: A ([LEAVE A REPLY](#))

Explanation/Reference: <https://docs.aws.amazon.com/polly/latest/dg/ssml.html>

Valid Professional-Machine-Learning-Engineer Dumps shared by BraindumpsPass.com for Helping Passing Professional-Machine-Learning-Engineer Exam! BraindumpsPass.com now offer the **newest Professional-Machine-Learning-Engineer exam dumps**, the BraindumpsPass.com Professional-Machine-Learning-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** BraindumpsPass.com Professional-Machine-Learning-Engineer dumps with Test Engine here: <https://www.braindumpsPass.com/Google/Professional-Machine-Learning-Engineer-practice-exam-dumps.html> (300 Q&As Dumps, **40%OFF** Special Discount: **Exam-Tests**)