Google.Professional-Machine-Learning-Engineer.v2023-02-17.q116

Exam Code:	Professional-Machine-Learning-Engineer
Exam Name:	Google Professional Machine Learning Engineer
Certification Provider:	Google
Free Question Number:	116
Version:	v2023-02-17
# of views:	1117
# of Questions views:	1160
https://www.dumpsdb.com/dumps/Google/Professional-Machine-Learning-Engineer/Google.Professional-Machine-Learning-Engineer.v2023-02-17.q116	

NEW QUESTION: 1
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. Downsample the data with upweighting to create a sample with 10% positive examples
B. Use the class distribution to generate 10% positive examples
C. Use a convolutional neural network with max pooling and softmax activation
D. Remove negative examples until the numbers of positive and negative examples are equal
Answer: C (LEAVE A REPLY)

NEW QUESTION: 2
You recently designed and built a custom neural network that uses critical dependencies specific to your organization's framework. You need to train the model using a managed training service on Google Cloud. However, the ML framework and related dependencies are not supported by AI Platform Training. Also, both your model and your data are too large to fit in memory on a single machine. Your ML framework of choice uses the scheduler, workers, and servers distribution structure. What should you do?
A. Reconfigure your code to a ML framework with dependencies that are supported by AI Platform Training
B. Build your custom container to run jobs on AI Platform Training
C. Use a built-in model available on AI Platform Training
D. Build your custom containers to run distributed training jobs on AI Platform Training
Answer: D (LEAVE A REPLY)

NEW QUESTION: 3

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: 4
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 a good result because predicting those who cancel their subscription is more difficult, since there is less data for this group.
B. This is a good result because the accuracy across both groups is greater than 80%.
C. This is not a good result because the model is performing worse than predicting that people will always renew their subscription.
D. 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.
Answer: (SHOW ANSWER)

NEW QUESTION: 5
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 Neo to take into account different types of medical devices. Programmatically control which model is invoked for inference based on the medical device type.
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. Create a single endpoint that accesses multiple models. Use Amazon SageMaker batch transform to control invoking the different models through the single endpoint.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 6
A web-based company wants to improve its conversion rate on its landing page. Using a large historical dataset of customer visits, the company has repeatedly trained a multi-class deep learning network algorithm on Amazon SageMaker. However, there is an overfitting problem: training data shows 90% accuracy in predictions, while test data shows 70% accuracy only.
The company needs to boost the generalization of its model before deploying it into production to maximize conversions of visits to purchases.
Which action is recommended to provide the HIGHEST accuracy model for the company's test and validation data?
A. Allocate a higher proportion of the overall data to the training dataset
B. Reduce the number of layers and units (or neurons) from the deep learning network
C. Apply L1 or L2 regularization and dropouts to the training
D. Increase the randomization of training data in the mini-batches used in training
Answer: B (LEAVE A REPLY)

NEW QUESTION: 7
A Machine Learning Specialist is working with a large company to leverage machine learning within its products. The company wants to group its customers into categories based on which customers will and will not churn within the next 6 months. The company has labeled the data available to the Specialist.
Which machine learning model type should the Specialist use to accomplish this task?
A. Linear regression
B. Classification
C. Clustering
D. Reinforcement learning
Answer: B (LEAVE A REPLY)
The goal of classification is to determine to which class or category a data point (customer in our case) belongs to. For classification problems, data scientists would use historical data with predefined target variables AKA labels (churner/non-churner) - answers that need to be predicted - to train an algorithm. With classification, businesses can answer the following questions:
* Will this customer churn or not?
* Will a customer renew their subscription?
* Will a user downgrade a pricing plan?
* Are there any signs of unusual customer behavior?
Reference: https://www.kdnuggets.com/2019/05/churn-prediction-machine-learning.html

NEW QUESTION: 8
You were asked to investigate failures of a production line component based on sensor readings. After receiving the dataset, you discover that less than 1% of the readings are positive examples representing failure incidents. You have tried to train several classification models, but none of them converge. How should you resolve the class imbalance problem?
A. Use the class distribution to generate 10% positive examples
B. Use a convolutional neural network with max pooling and softmax activation
C. Downsample the data with upweighting to create a sample with 10% positive examples
D. Remove negative examples until the numbers of positive and negative examples are equal

Answer: C (LEAVE A REPLY)
https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data#downsampling-and-upweighting
https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data

NEW QUESTION: 9
You work for a biotech startup that is experimenting with deep learning ML models based on properties of biological organisms. Your team frequently works on early-stage experiments with new architectures of ML models, and writes custom TensorFlow ops in C++. You train your models on large datasets and large batch sizes. Your typical batch size has 1024 examples, and each example is about 1 MB in size. The average size of a network with all weights and embeddings is 20 GB. What hardware should you choose for your models?
A. A cluster with 4 n1-highcpu-96 machines, each with 96 vCPUs and 86 GB RAM
B. A cluster with 2 a2-megagpu-16g machines, each with 16 NVIDIA Tesla A100 GPUs (640 GB GPU memory in total), 96 vCPUs, and 1.4 TB RAM
C. A cluster with an n1-highcpu-64 machine with a v2-8 TPU and 64 GB RAM
D. A cluster with 2 n1-highcpu-64 machines, each with 8 NVIDIA Tesla V100 GPUs (128 GB GPU memory in total), and a n1-highcpu-64 machine with 64 vCPUs and 58 GB RAM
Answer: B (LEAVE A REPLY)

NEW QUESTION: 10
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:
A)
B)
C)
D)
A. Option B
B. Option D
C. Option A
D. Option C
Answer: B (LEAVE A REPLY)

NEW QUESTION: 11
A Data Scientist is developing a machine learning model to classify whether a financial transaction is fraudulent. The labeled data available for training consists of 100,000 non-fraudulent observations and 1,000 fraudulent observations.
The Data Scientist applies the XGBoost algorithm to the data, resulting in the following confusion matrix when the trained model is applied to a previously unseen validation dataset. The accuracy of the model is 99.1%, but the Data Scientist has been asked to reduce the number of false negatives.
Which combination of steps should the Data Scientist take to reduce the number of false positive predictions by the model? (Choose two.)

A. Change the XGBoost eval_metric parameter to optimize based on rmse instead of error.
B. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.
C. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.
D. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.
E. Change the XGBoost eval_metric parameter to optimize based on AUC instead of error.
Answer: (SHOW ANSWER)

NEW QUESTION: 12
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 and 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. Train an anomaly detection model instead of an MLP
B. Add class weights to the MLP's loss function and then retrain
C. Gather more data using Amazon Mechanical Turk and then retrain
D. Train an XGBoost model instead of an MLP
Answer: D (LEAVE A REPLY)

NEW QUESTION: 13
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: 14
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 an Amazon Textract synchronous operation instead of an asynchronous operation.

B. 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.
C. 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.
D. 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.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 15
You are an ML engineer in the contact center of a large enterprise. You need to build a sentiment analysis tool that predicts customer sentiment from recorded phone conversations. You need to identify the best approach to building a model while ensuring that the gender, age, and cultural differences of the customers who called the contact center do not impact any stage of the model development pipeline and results. What should you do?
A. Convert the speech to text and extract sentiments based on the sentences
B. Extract sentiment directly from the voice recordings
C. Convert the speech to text and extract sentiment using syntactical analysis
D. Convert the speech to text and build a model based on the words
Answer: (SHOW ANSWER)

NEW QUESTION: 16
A Machine Learning Specialist is packaging a custom ResNet model into a Docker container so the company can leverage Amazon SageMaker for training. The Specialist is using Amazon EC2 P3 instances to train the model and needs to properly configure the Docker container to leverage the NVIDIA GPUs.
What does the Specialist need to do?
A. Bundle the NVIDIA drivers with the Docker image.
B. Build the Docker container to be NVIDIA-Docker compatible.
C. Organize the Docker container's file structure to execute on GPU instances.
D. Set the GPU flag in the Amazon SageMaker CreateTrainingJob request body.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 17

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

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

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

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

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

Answer: A (LEAVE A REPLY)

NEW QUESTION: 18

You are an ML engineer at an ecommerce company and have been tasked with building a model that predicts how much inventory the logistics team should order each month. Which approach should you take?

A. Use a regression model to predict how much additional inventory should be purchased each month. Give the results to the logistics team at the beginning of the month so they can increase inventory by the amount predicted by the model.

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

C. Use a time series forecasting model to predict each item's monthly sales. Give the results to the logistics team so they can base inventory on the amount predicted by the model.

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

Answer: (SHOW ANSWER)

NEW QUESTION: 19

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. Train local surrogate models to explain individual predictions.

C. Configure integrated gradients explanations on Vertex Explainable AI.

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

Answer: B (LEAVE A REPLY)

NEW QUESTION: 20
You work for a large hotel chain and have been asked to assist the marketing team in gathering predictions for a targeted marketing strategy. You need to make predictions about user lifetime value (LTV) over the next 30 days so that marketing can be adjusted accordingly. The customer dataset is in BigQuery, and you are preparing the tabular data for training with AutoML Tables. This data has a time signal that is spread across multiple columns. How should you ensure that AutoML fits the best model to your data?
A. Manually combine all columns that contain a time signal into an array Allow AutoML to interpret this array appropriately Choose an automatic data split across the training, validation, and testing sets
B. Submit the data for training without performing any manual transformations Allow AutoML to handle the appropriate transformations Choose an automatic data split across the training, validation, and testing sets
C. Submit the data for training without performing any manual transformations, and indicate an appropriate column as the Time column Allow AutoML to split your data based on the time signal provided, and reserve the more recent data for the validation and testing sets
D. Submit the data for training without performing any manual transformations Use the columns that have a time signal to manually split your data Ensure that the data in your validation set is from 30 days after the data in your training set and that the data in your testing set is from 30 days after your validation set
Answer: (SHOW ANSWER)
https://cloud.google.com/automl-tables/docs/data-best-practices#time

NEW QUESTION: 21
The displayed graph is from a forecasting model for testing a time series.
Considering the graph only, which conclusion should a Machine Learning Specialist make about the behavior of the model?
A. The model predicts both the trend and the seasonality well
B. The model predicts the trend well, but not the seasonality.
C. The model predicts the seasonality well, but not the trend.
D. The model does not predict the trend or the seasonality well.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 22
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: 23
A Machine Learning Specialist previously trained a logistic regression model using scikit-learn on a local machine, and the Specialist now wants to deploy it to production for inference only.
What steps should be taken to ensure Amazon SageMaker can host a model that was trained locally?
A. Build the Docker image with the inference code. Tag the Docker image with the registry hostname and upload it to Amazon ECR.
B. Build the Docker image with the inference code. Configure Docker Hub and upload the image to Amazon ECR.
C. Serialize the trained model so the format is compressed for deployment. Tag the Docker image with the registry hostname and upload it to Amazon S3.
D. Serialize the trained model so the format is compressed for deployment. Build the image and upload it to Docker Hub.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 24
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: B (LEAVE A REPLY)
Explanation/Reference: https://docs.aws.amazon.com/sagemaker/latest/dg/monitoring-cloudwatch.html

NEW QUESTION: 25
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 Vertex AI Pipelines with TensorFlow Extended (TFX) SDK.
B. Use Kubeflow Pipelines on Google Kubernetes Engine.
C. Use Cloud Composer for the orchestration.
D. Use Vertex AI Pipelines with Kubeflow Pipelines SDK.
Answer: (SHOW ANSWER)

NEW QUESTION: 26
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 AutoML Natural Language to build and test a classifier. Deploy the model as a REST API.
B. Use the Cloud Natural Language API to obtain metadata to classify the incoming cases.
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: A (LEAVE A REPLY)

NEW QUESTION: 27
You are going to train a DNN regression model with Keras APIs using this code:
How many trainable weights does your model have? (The arithmetic below is correct.)
A. 501*256+257*128+2 = 161154
B. 500*256+256*128+128*2 = 161024
C. 501*256+257*128+128*2=161408
D. 500*256*0 25+256*128*0 25+128*2 = 40448
Answer: C (LEAVE A REPLY)

NEW QUESTION: 28
You are training an LSTM-based model on AI Platform to summarize text using the following job submission script:
You want to ensure that training time is minimized without significantly compromising the accuracy of your model. What should you do?
A. Modify the batch size' parameter
B. Modify the 'learning rate' parameter
C. Modify the 'scale-tier' parameter
D. Modify the 'epochs' parameter
Answer: D (LEAVE A REPLY)

NEW QUESTION: 29
Your team is building an application for a global bank that will be used by millions of customers. You built a forecasting model that predicts customers1 account balances 3 days in the future. Your team will use the results in a new feature that will notify users when their account balance is likely to drop below $25. How should you serve your predictions?

A. 1. Create a Pub/Sub topic for each user
2 Deploy a Cloud Function that sends a notification when your model predicts that a user's account balance will drop below the $25 threshold.
B. 1. Create a Pub/Sub topic for each user
2. Deploy an application on the App Engine standard environment that sends a notification when your model predicts that a user's account balance will drop below the $25 threshold
C. 1. Build a notification system on Firebase
2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when the average of all account balance predictions drops below the $25 threshold
D. 1 Build a notification system on Firebase
Answer: (SHOW ANSWER)
2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when your model predicts that a user's account balance will drop below the $25 threshold Explanation:
Firebase is designed for exactly this sort of scenario. Also, it would not be possible to create millions of pubsub topics due to GCP quotas https://cloud.google.com/pubsub/quotas#quotas
https://firebase.google.com/docs/cloud-messaging

NEW QUESTION: 30
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 AutoML Natural Language to build the support requests classifier
B. Use an established text classification model on AI Platform to perform transfer learning
C. Use an established text classification model on AI Platform as-is to classify support requests
D. Use the Natural Language API to classify support requests
Answer: (SHOW ANSWER)

NEW QUESTION: 31
A Machine Learning Specialist receives customer data for an online shopping website. The data includes demographics, past visits, and locality information. The Specialist must develop a machine learning approach to identify the customer shopping patterns, preferences, and trends to enhance the website for better service and smart recommendations.
Which solution should the Specialist recommend?
A. Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in the customer database.
B. A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database.
C. Collaborative filtering based on user interactions and correlations to identify patterns in the customer database.
D. Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database.
Answer: C (LEAVE A REPLY)
Explanation

NEW QUESTION: 32
You need to build classification workflows over several structured datasets currently stored in BigQuery. Because you will be performing the classification several times, you want to complete the following steps without writing code: exploratory data analysis, feature selection, model building, training, and hyperparameter tuning and serving. What should you do?
A. Configure AutoML Tables to perform the classification task
B. Run a BigQuery ML task to perform logistic regression for the classification
C. Use AI Platform Notebooks to run the classification model with pandas library
D. Use AI Platform to run the classification model job configured for hyperparameter tuning
Answer: A (LEAVE A REPLY)
https://cloud.google.com/automl-tables/docs/beginners-guide

NEW QUESTION: 33
You work for a bank and are building a random forest model for fraud detection. You have a dataset that includes transactions, of which 1% are identified as fraudulent.
Which data transformation strategy would likely improve the performance of your classifier?
A. Write your data in TFRecords.
B. Use one-hot encoding on all categorical features.
C. Z-normalize all the numeric features.
D. Oversample the fraudulent transaction 10 times.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 34
A retail company intends to use machine learning to categorize new products. A labeled dataset of current products was provided to the Data Science team. The dataset includes 1,200 products. The labeled dataset has 15 features for each product such as title dimensions, weight, and price. Each product is labeled as belonging to one of six categories such as books, games, electronics, and movies.
Which model should be used for categorizing new products using the provided dataset for training?
A. A DeepAR forecasting model based on a recurrent neural network (RNN)
B. AnXGBoost model where the objective parameter is set to multi:softmax
C. A regression forest where the number of trees is set equal to the number of product categories
D. A deep convolutional neural network (CNN) with a softmax activation function for the last layer
Answer: (SHOW ANSWER)

NEW QUESTION: 35
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: 36
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: 37
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 dat a. You want the model to be resilient to overfitting. Which strategy should you use when retraining the model?
A. Run a hyperparameter tuning job on Al Platform to optimize for the learning rate, and increase the number of neurons by a factor of 2.
B. Apply a 12 regularization parameter of 0.4, and decrease the learning rate by a factor of 10.
C. Run a hyperparameter tuning job on Al Platform to optimize for the L2 regularization and dropout parameters
D. Apply a dropout parameter of 0 2, and decrease the learning rate by a factor of 10
Answer: D (LEAVE A REPLY)

NEW QUESTION: 38
Machine Learning Specialist is building a model to predict future employment rates based on a wide range of economic factors. While exploring the data, the Specialist notices that the magnitude of the input features vary greatly. The Specialist does not want variables with a larger magnitude to dominate the model.

What should the Specialist do to prepare the data for model training?

A. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution.

B. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude.

C. Apply normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude.

D. Apply the orthogonal sparse bigram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.

Answer: C (LEAVE A REPLY)

Explanation/Reference: https://docs.aws.amazon.com/machine-learning/latest/dg/data-transformations-reference.html

NEW QUESTION: 39

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: 40

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 AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.

B. 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.

C. 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.

D. 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.

Answer: A (LEAVE A REPLY)

NEW QUESTION: 41
You work for an advertising company and want to understand the effectiveness of your company's latest advertising campaign. You have streamed 500 MB of campaign data into BigQuery. You want to query the table, and then manipulate the results of that query with a pandas dataframe in an AI Platform notebook. What should you do?
A. Use AI Platform Notebooks' BigQuery cell magic to query the data, and ingest the results as a pandas dataframe
B. Export your table as a CSV file from BigQuery to Google Drive, and use the Google Drive API to ingest the file into your notebook instance
C. Download your table from BigQuery as a local CSV file, and upload it to your AI Platform notebook instance Use pandas. read_csv to ingest the file as a pandas dataframe
D. From a bash cell in your AI Platform notebook, use the bq extract command to export the table as a CSV file to Cloud Storage, and then use gsutii cp to copy the data into the notebook Use pandas. read_csv to ingest the file as a pandas dataframe
Answer: A (LEAVE A REPLY)
Refer to this link for details: https://cloud.google.com/bigquery/docs/bigquery-storage-python-pandas First 2 points talks about querying the data.
Download query results to a pandas DataFrame by using the BigQuery Storage API from the IPython magics for BigQuery in a Jupyter notebook.
Download query results to a pandas DataFrame by using the BigQuery client library for Python.
Download BigQuery table data to a pandas DataFrame by using the BigQuery client library for Python.
Download BigQuery table data to a pandas DataFrame by using the BigQuery Storage API client library for Python.
https://googleapis.dev/python/bigquery/latest/magics.html#ipython-magics-for-bigquery
https://cloud.google.com/bigquery/docs/bigquery-storage-python-pandas

NEW QUESTION: 42
You are designing an ML recommendation model for shoppers on your company's ecommerce website. You will use Recommendations AI to build, test, and deploy your system. How should you develop recommendations that increase revenue while following best practices?
A. Use the "Other Products You May Like" recommendation type to increase the click-through rate
B. Use the "Frequently Bought Together' recommendation type to increase the shopping cart size for each order.
C. Import your user events and then your product catalog to make sure you have the highest quality event stream
D. Because it will take time to collect and record product data, use placeholder values for the product catalog to test the viability of the model.
Answer: B (LEAVE A REPLY)
Frequently bought together' recommendations aim to up-sell and cross-sell customers by providing product.

NEW QUESTION: 43
Your team is building a convolutional neural network (CNN)-based architecture from scratch. The preliminary experiments running on your on-premises CPU-only infrastructure were encouraging, but have slow convergence. You have been asked to speed up model training to reduce time-to-market. You want to experiment with virtual machines (VMs) on Google Cloud to leverage more powerful hardware. Your code does not include any manual device placement and has not been wrapped in Estimator model-level abstraction. Which environment should you train your model on?
A. AVM on Compute Engine and 1 TPU with all dependencies installed manually.

B. AVM on Compute Engine and 8 GPUs with all dependencies installed manually.
C. A Deep Learning VM with an n1-standard-2 machine and 1 GPU with all libraries pre-installed.
D. A Deep Learning VM with more powerful CPU e2-highcpu-16 machines with all libraries pre-installed.
Answer: C (LEAVE A REPLY)
https://cloud.google.com/deep-learning-vm/docs/cli#creating_an_instance_with_one_or_more_gpus
https://cloud.google.com/deep-learning-vm/docs/introduction#pre-installed_packages
"speed up model training" will make us biased towards GPU,TPU options by options eliminations we may need to stay away of any manual installations , so using preconfigered deep learning will speed up time to market

NEW QUESTION: 44
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: (SHOW ANSWER)

NEW QUESTION: 45
You are training a TensorFlow model on a structured data set with 100 billion records stored in several CSV files. You need to improve the input/output execution performance. What should you do?
A. Load the data into BigQuery and read the data from BigQuery.
B. Convert the CSV files into shards of TFRecords, and store the data in the Hadoop Distributed File System (HDFS)
C. Load the data into Cloud Bigtable, and read the data from Bigtable
D. Convert the CSV files into shards of TFRecords, and store the data in Cloud Storage
Answer: (SHOW ANSWER)

NEW QUESTION: 46
You are an ML engineer at a travel company. You have been researching customers' travel behavior for many years, and you have deployed models that predict customers' vacation patterns. You have observed that customers' vacation destinations vary based on seasonality and holidays; however, these seasonal variations are similar across years. You want to quickly and easily store and compare the model versions and performance statistics across years. What should you do?
A. Store the performance statistics in Cloud SQL. Query that database to compare the performance statistics across the model versions.
B. 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. Store the performance statistics of each version of your models using seasons and years as events in Vertex ML Metadata. Compare the results across the slices.

D. 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.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 47
A retail chain has been ingesting purchasing records from its network of 20,000 stores to Amazon S3 using Amazon Kinesis Data Firehose. To support training an improved machine learning model, training records will require new but simple transformations, and some attributes will be combined. The model needs to be retrained daily.
Given the large number of stores and the legacy data ingestion, which change will require the LEAST amount of development effort?
A. Require that the stores to switch to capturing their data locally on AWS Storage Gateway for loading into Amazon S3, then use AWS Glue to do the transformation.
B. Insert an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehose stream that transforms raw record attributes into simple transformed values using SQL.
C. Deploy an Amazon EMR cluster running Apache Spark with the transformation logic, and have the cluster run each day on the accumulating records in Amazon S3, outputting new/transformed records to Amazon S3.
D. Spin up a fleet of Amazon EC2 instances with the transformation logic, have them transform the data records accumulating on Amazon S3, and output the transformed records to Amazon S3.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 48
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 clustering on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
B. Use a recommendation engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.

C. Use regression on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
D. Use a decision tree classifier engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 49
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. Create an experiment in Kubeflow Pipelines to organize multiple runs
D. Run multiple training jobs on AI Platform with similar job names
Answer: D (LEAVE A REPLY)

NEW QUESTION: 50
A Machine Learning Specialist is training a model to identify the make and model of vehicles in images. The Specialist wants to use transfer learning and an existing model trained on images of general objects. The Specialist collated a large custom dataset of pictures containing different vehicle makes and models.
What should the Specialist do to initialize the model to re-train it with the custom data?
A. Initialize the model with random weights in all layers and replace the last fully connected layer.
B. Initialize the model with pre-trained weights in all layers including the last fully connected layer.
C. Initialize the model with pre-trained weights in all layers and replace the last fully connected layer.
D. Initialize the model with random weights in all layers including the last fully connected layer.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 51
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. Change the search algorithm from Bayesian search to random search.
B. Set the early stopping parameter to TRUE
C. Decrease the number of parallel trials
D. Decrease the range of floating-point values
E. Decrease the maximum number of trials during subsequent training phases.
Answer: A,D (LEAVE A REPLY)

NEW QUESTION: 52

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. Decrease the range of floating-point values

C. Set the early stopping parameter to TRUE

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

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

Answer: (SHOW ANSWER)

Reference:
https://cloud.google.com/ai-platform/training/docs/using-hyperparameter-tuning#early-stopping

NEW QUESTION: 53

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. Differential privacy federated learning, and explainability

B. Traceability, reproducibility, and explainability

C. Redaction, reproducibility, and explainability

D. Federated learning, reproducibility, and explainability

Answer: B (LEAVE A REPLY)

NEW QUESTION: 54

Your team is building an application for a global bank that will be used by millions of customers. You built a forecasting model that predicts customers1 account balances 3 days in the future. Your team will use the results in a new feature that will notify users when their account balance is likely to drop below $25. How should you serve your predictions?

A. 1 Build a notification system on Firebase
2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when your model predicts that a user's account balance will drop below the $25 threshold

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

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

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

Answer: D (LEAVE A REPLY)

NEW QUESTION: 55

You are working on a classification problem with time series data and achieved an area under the receiver operating characteristic curve (AUC ROC) value of 99% for training data after just a few experiments. You haven't explored using any sophisticated algorithms or spent any time on hyperparameter tuning. What should your next step be to identify and fix the problem?
A. Address the model overfitting by using a less complex algorithm.
B. Address data leakage by applying nested cross-validation during model training.
C. Address data leakage by removing features highly correlated with the target value.
D. Address the model overfitting by tuning the hyperparameters to reduce the AUC ROC value.
Answer: (SHOW ANSWER)
https://towardsdatascience.com/time-series-nested-cross-validation-76adba623eb9

NEW QUESTION: 56
You work for a large hotel chain and have been asked to assist the marketing team in gathering predictions for a targeted marketing strategy. You need to make predictions about user lifetime value (LTV) over the next 30 days so that marketing can be adjusted accordingly. The customer dataset is in BigQuery, and you are preparing the tabular data for training with AutoML Tables. This data has a time signal that is spread across multiple columns. How should you ensure that AutoML fits the best model to your data?
A. Submit the data for training without performing any manual transformations Use the columns that have a time signal to manually split your data Ensure that the data in your validation set is from 30 days after the data in your training set and that the data in your testing set is from 30 days after your validation set
B. Submit the data for training without performing any manual transformations, and indicate an appropriate column as the Time column Allow AutoML to split your data based on the time signal provided, and reserve the more recent data for the validation and testing sets
C. Submit the data for training without performing any manual transformations Allow AutoML to handle the appropriate transformations Choose an automatic data split across the training, validation, and testing sets
D. Manually combine all columns that contain a time signal into an array Allow AutoML to interpret this array appropriately Choose an automatic data split across the training, validation, and testing sets
Answer: A (LEAVE A REPLY)

NEW QUESTION: 57
You are developing models to classify customer support emails. You created models with TensorFlow Estimators using small datasets on your on-premises system, but you now need to train the models using large datasets to ensure high performance. You will port your models to Google Cloud and want to minimize code refactoring and infrastructure overhead for easier migration from on-prem to cloud. What should you do?
A. Use AI Platform for distributed training
B. Create a cluster on Dataproc for training
C. Create a Managed Instance Group with autoscaling
D. Use Kubeflow Pipelines to train on a Google Kubernetes Engine cluster.
Answer: A (LEAVE A REPLY)
AI platform also contains kubeflow pipelines. you don't need to set up infrastructure to use it. For D you need to set up a kubernetes cluster engine. The question asks us to minimize infrastructure overheard.

NEW QUESTION: 58

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: 59
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: (SHOW ANSWER)
(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: 60
Machine Learning Specialist is training a model to identify the make and model of vehicles in images. The Specialist wants to use transfer learning and an existing model trained on images of general objects. The Specialist collated a large custom dataset of pictures containing different vehicle makes and models.
What should the Specialist do to initialize the model to re-train it with the custom data?

A. Initialize the model with random weights in all layers including the last fully connected layer.
B. Initialize the model with pre-trained weights in all layers and replace the last fully connected layer.
C. Initialize the model with random weights in all layers and replace the last fully connected layer.
D. Initialize the model with pre-trained weights in all layers including the last fully connected layer.
Answer: B (LEAVE A REPLY)
Explanation/Reference:

NEW QUESTION: 61
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: 62
You work for an online publisher that delivers news articles to over 50 million readers. You have built an AI model that recommends content for the company's weekly newsletter. A recommendation is considered successful if the article is opened within two days of the newsletter's published date and the user remains on the page for at least one minute.
All the information needed to compute the success metric is available in BigQuery and is updated hourly. The model is trained on eight weeks of data, on average its performance degrades below the acceptable baseline after five weeks, and training time is 12 hours. You want to ensure that the model's performance is above the acceptable baseline while minimizing cost. How should you monitor the model to determine when retraining is necessary?
A. Schedule a cron job in Cloud Tasks to retrain the model every week before the newsletter is created.
B. Use Vertex AI Model Monitoring to detect skew of the input features with a sample rate of 100% and a monitoring frequency of two days.
C. Schedule a daily Dataflow job in Cloud Composer to compute the success metric.
D. Schedule a weekly query in BigQuery to compute the success metric.

Answer: C (LEAVE A REPLY)

NEW QUESTION: 63
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. Area under the curve
B. Root Mean Square Error (RMSE)
C. Residual plots
D. Confusion matrix
Answer: A (LEAVE A REPLY)

NEW QUESTION: 64
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 cluster on Dataproc for training
B. Use AI Platform for distributed training
C. Use Kubeflow Pipelines to train on a Google Kubernetes Engine cluster.
D. Create a Managed Instance Group with autoscaling
Answer: C (LEAVE A REPLY)

NEW QUESTION: 65
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. Add more deep trees to the random forest to enable the model to learn more features.
B. Change the cost function so that false positives have a higher impact on the cost value than false negatives.
C. Change the cost function so that false negatives have a higher impact on the cost value than false positives.
D. Include a copy of the samples in the test dataset in the training dataset.
E. Generate more positive samples by duplicating the positive samples and adding a small amount of noise to the duplicated data.
Answer: (SHOW ANSWER)

NEW QUESTION: 66

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: ['driversjicense', '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: (SHOW ANSWER)

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

NEW QUESTION: 67

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

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

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

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

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

Answer: D (LEAVE A REPLY)

NEW QUESTION: 68

An online reseller has a large, multi-column dataset with one column missing 30% of its data. A Machine Learning Specialist believes that certain columns in the dataset could be used to reconstruct the missing data.

Which reconstruction approach should the Specialist use to preserve the integrity of the dataset?

A. Listwise deletion

B. Last observation carried forward

C. Multiple imputation

D. Mean substitution

Answer: C (LEAVE A REPLY)

Explanation/Reference: https://worldwidescience.org/topicpages/i/imputing+missing+values.html

NEW QUESTION: 69

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: 70
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. Z-normalize all the numeric features.
B. Write your data in TFRecords.
C. Oversample the fraudulent transaction 10 times.
D. Use one-hot encoding on all categorical features.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 71
You work for a retailer that sells clothes to customers around the world. You have been tasked with ensuring that ML models are built in a secure manner. Specifically, you need to protect sensitive customer data that might be used in the models. You have identified four fields containing sensitive data that are being used by your data science team: AGE, IS_EXISTING_CUSTOMER, LATITUDE_LONGITUDE, and SHIRT_SIZE. What should you do with the data before it is made available to the data science team for training purposes?
A. Use principal component analysis (PCA) to reduce the four sensitive fields to one PCA vector.
B. Coarsen the data by putting AGE into quantiles and rounding LATITUDE_LONGTTUDE into single precision. The other two fields are already as coarse as possible.
C. Remove all sensitive data fields, and ask the data science team to build their models using non-sensitive data.
D. Tokenize all of the fields using hashed dummy values to replace the real values.
Answer: (SHOW ANSWER)

NEW QUESTION: 72
You have trained a model on a dataset that required computationally expensive preprocessing operations. You need to execute the same preprocessing at prediction time. You deployed the model on AI Platform for high-throughput online prediction. Which architecture should you use?
A. * Validate the accuracy of the model that you trained on preprocessed data
* Create a new model that uses the raw data and is available in real time
* Deploy the new model onto AI Platform for online prediction
B. * Send incoming prediction requests to a Pub/Sub topic
* Transform the incoming data using a Dataflow job
* Submit a prediction request to AI Platform using the transformed data
* Write the predictions to an outbound Pub/Sub queue
C. * Stream incoming prediction request data into Cloud Spanner
* Create a view to abstract your preprocessing logic.

* Query the view every second for new records
* Submit a prediction request to AI Platform using the transformed data
* Write the predictions to an outbound Pub/Sub queue.

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

Answer: B (LEAVE A REPLY)

https://cloud.google.com/architecture/data-preprocessing-for-ml-with-tf-transform-pt1#where_to_do_preprocessing

NEW QUESTION: 73

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

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

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

C. Use local feature importance from the predictions.

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

Answer: (SHOW ANSWER)

NEW QUESTION: 74

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. Redeploy the model as a batch transform job on an M5 instance.

B. Redeploy the model on a P3dn instance.

C. Redeploy the model on an M5 instance. Attach Amazon Elastic Inference to the instance.

D. Deploy the model onto an Amazon Elastic Container Service (Amazon ECS) cluster using a P3 instance.

Answer: D (LEAVE A REPLY)

NEW QUESTION: 75

A monitoring service generates 1 TB of scale metrics record data every minute. A Research team performs queries on this data using Amazon Athena. The queries run slowly due to the large volume of data, and the team requires better performance.

How should the records be stored in Amazon S3 to improve query performance?

A. CSV files

B. RecordIO

C. Parquet files

D. Compressed JSON

Answer: (SHOW ANSWER)

NEW QUESTION: 76
A company wants to predict the sale prices of houses based on available historical sales data. The target variable in the company's dataset is the sale price. The features include parameters such as the lot size, living area measurements, non-living area measurements, number of bedrooms, number of bathrooms, year built, and postal code. The company wants to use multi-variable linear regression to predict house sale prices.
Which step should a machine learning specialist take to remove features that are irrelevant for the analysis and reduce the model's complexity?
A. Plot a histogram of the features and compute their standard deviation. Remove features with high variance.
B. Plot a histogram of the features and compute their standard deviation. Remove features with low variance.
C. Build a heatmap showing the correlation of the dataset against itself. Remove features with low mutual correlation scores.
D. Run a correlation check of all features against the target variable. Remove features with low target variable correlation scores.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 77
A Machine Learning Specialist must build out a process to query a dataset on Amazon S3 using Amazon Athena. The dataset contains more than 800,000 records stored as plaintext CSV files. Each record contains
200 columns and is approximately 1.5 MB in size. Most queries will span 5 to 10 columns only.
How should the Machine Learning Specialist transform the dataset to minimize query runtime?
A. Convert the records to Apache Parquet format.
B. Convert the records to JSON format.
C. Convert the records to GZIP CSV format.
D. Convert the records to XML format.
Answer: A (LEAVE A REPLY)
Using compressions will reduce the amount of data scanned by Amazon Athena, and also reduce your S3 bucket storage. It's a Win-Win for your AWS bill. Supported formats: GZIP, LZO, SNAPPY (Parquet) and ZLIB.
Reference: https://www.cloudforecast.io/blog/using-parquet-on-athena-to-save-money-on-aws/

NEW QUESTION: 78

A Machine Learning Specialist kicks off a hyperparameter tuning job for a tree-based ensemble model using Amazon SageMaker with Area Under the ROC Curve (AUC) as the objective metric. This workflow will eventually be deployed in a pipeline that retrains and tunes hyperparameters each night to model click-through on data that goes stale every 24 hours.
With the goal of decreasing the amount of time it takes to train these models, and ultimately to decrease costs, the Specialist wants to reconfigure the input hyperparameter range(s).
Which visualization will accomplish this?
A. A scatter plot showing the performance of the objective metric over each training iteration.
B. A histogram showing whether the most important input feature is Gaussian.
C. A scatter plot with points colored by target variable that uses t-Distributed Stochastic Neighbor Embedding (t-SNE) to visualize the large number of input variables in an easier-to-read dimension.
D. A scatter plot showing the correlation between maximum tree depth and the objective metric.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 79
A company uses a long short-term memory (LSTM) model to evaluate the risk factors of a particular energy sector. The model reviews multi-page text documents to analyze each sentence of the text and categorize it as either a potential risk or no risk. The model is not performing well, even though the Data Scientist has experimented with many different network structures and tuned the corresponding hyperparameters.
Which approach will provide the MAXIMUM performance boost?
A. Reduce the learning rate and run the training process until the training loss stops decreasing.
B. Use gated recurrent units (GRUs) instead of LSTM and run the training process until the validation loss stops decreasing.
C. Initialize the words by term frequency-inverse document frequency (TF-IDF) vectors pretrained on a large collection of news articles related to the energy sector.
D. Initialize the words by word2vec embeddings pretrained on a large collection of news articles related to the energy sector.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 80
One of your models is trained using data provided by a third-party data broker. The data broker does not reliably notify you of formatting changes in the dat a. You want to make your model training pipeline more robust to issues like this. What should you do?
A. Use TensorFlow Data Validation to detect and flag schema anomalies.
B. Use custom TensorFlow functions at the start of your model training to detect and flag known formatting errors.
C. Use TensorFlow Transform to create a preprocessing component that will normalize data to the expected distribution, and replace values that don't match the schema with 0.
D. Use tf.math to analyze the data, compute summary statistics, and flag statistical anomalies.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 81
A Mobile Network Operator is building an analytics platform to analyze and optimize a company's operations using Amazon Athena and Amazon S3.
The source systems send data in .CSV format in real time. The Data Engineering team wants to transform the data to the Apache Parquet format before storing it on Amazon S3.

Which solution takes the LEAST effort to implement?
A. Ingest .CSV data using Apache Kafka Streams on Amazon EC2 instances and use Kafka Connect S3 to serialize data as Parquet
B. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Glue to convert data into Parquet.
C. Ingest .CSV data using Apache Spark Structured Streaming in an Amazon EMR cluster and use Apache Spark to convert data into Parquet.
D. Ingest .CSV data from Amazon Kinesis Data Streams and use Amazon Kinesis Data Firehose to convert data into Parquet.
Answer: B (LEAVE A REPLY)
Explanation/Reference:

NEW QUESTION: 82
A Machine Learning Specialist works for a credit card processing company and needs to predict which transactions may be fraudulent in near-real time. Specifically, the Specialist must train a model that returns the probability that a given transaction may fraudulent.
How should the Specialist frame this business problem?
A. Regression classification
B. Multi-category classification
C. Binary classification
D. Streaming classification
Answer: B (LEAVE A REPLY)

NEW QUESTION: 83
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. Write a Cloud Functions script that launches a training and deploying job on Ai Platform that is triggered by Cloud Scheduler
C. Use a model trained and deployed on BigQuery ML and trigger retraining with the scheduled query feature in BigQuery
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)

NEW QUESTION: 84
You recently developed a deep learning model using Keras, and now you are experimenting with different training strategies. First, you trained the model using a single GPU, but the training process was too slow. Next, you distributed the training across 4 GPUs using tf.distribute.MirroredStrategy (with no other changes), but you did not observe a decrease in training time. What should you do?
A. Create a custom training loop.
B. Distribute the dataset with tf.distribute.Strategy.experimental_distribute_dataset
C. Increase the batch size.
D. Use a TPU with tf.distribute.TPUStrategy.
Answer: (SHOW ANSWER)

NEW QUESTION: 85
You are training an LSTM-based model on AI Platform to summarize text using the following job submission script:
You want to ensure that training time is minimized without significantly compromising the accuracy of your model. What should you do?
A. Modify the 'epochs' parameter
B. Modify the 'scale-tier' parameter
C. Modify the batch size' parameter
D. Modify the 'learning rate' parameter
Answer: (SHOW ANSWER)
https://cloud.google.com/ai-platform/training/docs/machine-types#scale_tiers Google may optimize the configuration of the scale tiers for different jobs over time, based on customer feedback and the availability of cloud resources. Each scale tier is defined in terms of its suitability for certain types of jobs. Generally, the more advanced the tier, the more machines are allocated to the cluster, and the more powerful the specifications of each virtual machine. As you increase the complexity of the scale tier, the hourly cost of training jobs, measured in training units, also increases. See the pricing page to calculate the cost of your job.

NEW QUESTION: 86
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. Rewrite your input function to resize and reshape the input images.
B. Move from Cloud TPU v2 to Cloud TPU v3 and increase batch size.
C. Move from Cloud TPU v2 to 8 NVIDIA V100 GPUs and increase batch size.
D. Rewrite your input function using parallel reads, parallel processing, and prefetch.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 87
A Data Scientist needs to create a serverless ingestion and analytics solution for high-velocity, real-time streaming data.
The ingestion process must buffer and convert incoming records from JSON to a query-optimized, columnar format without data loss.
The output datastore must be highly available, and Analysts must be able to run SQL queries against the data and connect to existing business intelligence dashboards.
Which solution should the Data Scientist build to satisfy the requirements?
A. Create a schema in the AWS Glue Data Catalog of the incoming data format. Use an Amazon Kinesis Data Firehose delivery stream to stream the data and transform the data to Apache Parquet or ORC format using the AWS Glue Data Catalog before delivering to Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena, and connect to BI tools using the Athena Java Database Connectivity (JDBC) connector.
B. Write each JSON record to a staging location in Amazon S3. Use the S3 Put event to trigger an AWS Lambda function that transforms the data into Apache Parquet or ORC format and writes the data to a processed data location in Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena, and connect to BI tools using the Athena Java Database Connectivity (JDBC) connector.

C. Write each JSON record to a staging location in Amazon S3. Use the S3 Put event to trigger an AWS Lambda function that transforms the data into Apache Parquet or ORC format and inserts it into an Amazon RDS PostgreSQL database. Have the Analysts query and run dashboards from the RDS database.
D. Use Amazon Kinesis Data Analytics to ingest the streaming data and perform real-time SQL queries to convert the records to Apache Parquet before delivering to Amazon S3. Have the Analysts query the data directly from Amazon S3 using Amazon Athena and connect to BI tools using the Athena Java Database Connectivity (JDBC) connector.
Answer: A (LEAVE A REPLY)
Explanation/Reference:

NEW QUESTION: 88
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. Logarithmic transformation
B. Cross-validation
C. High-degree polynomial transformation
D. One hot encoding
E. Numerical value binning
Answer: (SHOW ANSWER)

NEW QUESTION: 89
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: 90
A trucking company is collecting live image data from its fleet of trucks across the globe. The data is growing rapidly and approximately 100 GB of new data is generated every day. The company wants to explore machine learning uses cases while ensuring the data is only accessible to specific IAM users.
Which storage option provides the most processing flexibility and will allow access control with IAM?
A. Use a database, such as Amazon DynamoDB, to store the images, and set the IAM policies to restrict access to only the desired IAM users.

B. Use an Amazon S3-backed data lake to store the raw images, and set up the permissions using bucket policies.
C. Setup up Amazon EMR with Hadoop Distributed File System (HDFS) to store the files, and restrict access to the EMR instances using IAM policies.
D. Configure Amazon EFS with IAM policies to make the data available to Amazon EC2 instances owned by the IAM users.
Answer: C (LEAVE A REPLY)
Explanation

NEW QUESTION: 91
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: ['driversjicense', 'passport', 'credit_card']. Which loss function should you use?
A. Sparse categorical cross-entropy
B. Binary cross-entropy
C. Categorical cross-entropy
D. Categorical hinge
Answer: B (LEAVE A REPLY)

NEW QUESTION: 92
You work on a growing team of more than 50 data scientists who all use AI Platform. You are designing a strategy to organize your jobs, models, and versions in a clean and scalable way. Which strategy should you choose?
A. Set up restrictive IAM permissions on the AI Platform notebooks so that only a single user or group can access a given instance.
B. Separate each data scientist's work into a different project to ensure that the jobs, models, and versions created by each data scientist are accessible only to that user.
C. Use labels to organize resources into descriptive categories. Apply a label to each created resource so that users can filter the results by label when viewing or monitoring the resources.
D. Set up a BigQuery sink for Cloud Logging logs that is appropriately filtered to capture information about AI Platform resource usage. In BigQuery, create a SQL view that maps users to the resources they are using
Answer: C (LEAVE A REPLY)
https://cloud.google.com/ai-platform/prediction/docs/resource-labels#overview_of_labels You can add labels to your AI Platform Prediction jobs, models, and model versions, then use those labels to organize resources into categories when viewing or monitoring the resources.

NEW QUESTION: 93
A data scientist wants to use Amazon Forecast to build a forecasting model for inventory demand for a retail company. The company has provided a dataset of historic inventory demand for its products as a .csv file stored in an Amazon S3 bucket. The table below shows a sample of the dataset.
How should the data scientist transform the data?
A. Use a Jupyter notebook in Amazon SageMaker to separate the dataset into a related time series dataset and an item metadata dataset. Upload both datasets as tables in Amazon Aurora.
B. Use a Jupyter notebook in Amazon SageMaker to transform the data into the optimized protobuf recordIO format. Upload the dataset in this format to Amazon S3.
C. Use AWS Batch jobs to separate the dataset into a target time series dataset, a related time series dataset, and an item metadata dataset. Upload them directly to Forecast from a local machine.
D. Use ETL jobs in AWS Glue to separate the dataset into a target time series dataset and an item metadata dataset. Upload both datasets as .csv files to Amazon S3.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 94
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. K-means
B. Linear regression
C. Principal component analysis (PCA)
D. Logistic regression
Answer: B (LEAVE A REPLY)

NEW QUESTION: 95
You have trained a model on a dataset that required computationally expensive preprocessing operations. You need to execute the same preprocessing at prediction time. You deployed the model on AI Platform for high-throughput online prediction. Which architecture should you use?
A. Send incoming prediction requests to a Pub/Sub topic
* Transform the incoming data using a Dataflow job
* Submit a prediction request to AI Platform using the transformed data
* Write the predictions to an outbound Pub/Sub queue
B. Validate the accuracy of the model that you trained on preprocessed data
* Create a new model that uses the raw data and is available in real time
* Deploy the new model onto AI Platform for online prediction
C. Stream incoming prediction request data into Cloud Spanner
* Create a view to abstract your preprocessing logic.
* Query the view every second for new records
* Submit a prediction request to AI Platform using the transformed data

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

NEW QUESTION: 96
Your team is building an application for a global bank that will be used by millions of customers. You built a forecasting model that predicts customers1 account balances 3 days in the future. Your team will use the results in a new feature that will notify users when their account balance is likely to drop below $25. How should you serve your predictions?
A. 1. Create a Pub/Sub topic for each user
2 Deploy a Cloud Function that sends a notification when your model predicts that a user's account balance will drop below the $25 threshold.
B. 1. Create a Pub/Sub topic for each user
2. Deploy an application on the App Engine standard environment that sends a notification when your model predicts that a user's account balance will drop below the $25 threshold
C. 1 Build a notification system on Firebase
2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when your model predicts that a user's account balance will drop below the $25 threshold
D. 1. Build a notification system on Firebase
2. Register each user with a user ID on the Firebase Cloud Messaging server, which sends a notification when the average of all account balance predictions drops below the $25 threshold
Answer: A (LEAVE A REPLY)

NEW QUESTION: 97
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: 98
You need to design an architecture that serves asynchronous predictions to determine whether a particular mission-critical machine part will fail. Your system collects data from multiple sensors from the machine. You want to build a model that will predict a failure in the next N minutes, given the average of each sensor's data from the past 12 hours. How should you design the architecture?

A. 1. Export 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.
B. 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.
C. 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.
D. 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.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 99
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: 100
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: (SHOW ANSWER)
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: 101
A Machine Learning Specialist has completed a proof of concept for a company using a small data sample, and now the Specialist is ready to implement an end-to-end solution in AWS using Amazon SageMaker. The historical training data is stored in Amazon RDS. Which approach should the Specialist use for training a model using that data?
A. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in.
B. Write a direct connection to the SQL database within the notebook and pull data in
C. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.
D. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 102
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. Change the optimizer
C. Reduce the image shape
D. Reduce the batch size
Answer: D (LEAVE A REPLY)

NEW QUESTION: 103
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: 104

You work for the AI team of an automobile company, and you are developing a visual defect detection model using TensorFlow and Keras. To improve your model performance, you want to incorporate some image augmentation functions such as translation, cropping, and contrast tweaking. You randomly apply these functions to each training batch. You want to optimize your data processing pipeline for run time and compute resources utilization. What should you do?
A. Embed the augmentation functions dynamically as part of Keras generators.
B. Use Dataflow to create all possible augmentations, and store them as TFRecords.
C. Use Dataflow to create the augmentations dynamically per training run, and stage them as TFRecords.
D. Embed the augmentation functions dynamically in the tf.Data pipeline.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 105
A city wants to monitor its air quality to address the consequences of air pollution. A Machine Learning Specialist needs to forecast the air quality in parts per million of contaminates for the next 2 days in the city. As this is a prototype, only daily data from the last year is available.
Which model is MOST likely to provide the best results in Amazon SageMaker?
A. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the single time series consisting of the full year of data with a predictor_typeof regressor.
B. Use Amazon SageMaker Random Cut Forest (RCF) on the single time series consisting of the full year of data.
C. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full year of data with a predictor_typeof regressor.
D. Use the Amazon SageMaker Linear Learner algorithm on the single time series consisting of the full year of data with a predictor_typeof classifier.
Answer: C (LEAVE A REPLY)
Explanation/Reference: https://aws.amazon.com/blogs/machine-learning/build-a-model-to-predict-the-impact-of-weather- on-urban-air-quality-using-amazon-sagemaker/?ref=Welcome.AI

NEW QUESTION: 106
You are designing an architecture with a serveress 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: (SHOW ANSWER)
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: 107
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 all hyperparameters are tuned
B. Ensure that training is reproducible
C. Ensure that feature expectations are captured in the schema
D. Ensure that model performance is monitored
Answer: B (LEAVE A REPLY)

NEW QUESTION: 108
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. Containerize the PySpark transformation step, and add it to your pipeline.
B. 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.
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: B (LEAVE A REPLY)

NEW QUESTION: 109
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 the feature's mean.
B. Replace the missing values with a placeholder category indicating a missing value.
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: (SHOW ANSWER)

NEW QUESTION: 110
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: A (LEAVE A REPLY)
Explanation

NEW QUESTION: 111
You work for a company that manages a ticketing platform for a large chain of cinemas. Customers use a mobile app to search for movies they're interested in and purchase tickets in the app. Ticket purchase requests are sent to Pub/Sub and are processed with a Dataflow streaming pipeline configured to conduct the following steps:
1. Check for availability of the movie tickets at the selected cinema.
2. Assign the ticket price and accept payment.
3. Reserve the tickets at the selected cinema.
4. Send successful purchases to your database.
Each step in this process has low latency requirements (less than 50 milliseconds). You have developed a logistic regression model with BigQuery ML that predicts whether offering a promo code for free popcorn increases the chance of a ticket purchase, and this prediction should be added to the ticket purchase process. You want to identify the simplest way to deploy this model to production while adding minimal latency. What should you do?
A. Export your model in TensorFlow format, deploy it on Vertex AI, and query the prediction endpoint from your streaming pipeline.

B. Run batch inference with BigQuery ML every five minutes on each new set of tickets issued.
C. Convert your model with TensorFlow Lite (TFLite), and add it to the mobile app so that the promo code and the incoming request arrive together in Pub/Sub.
D. Export your model in TensorFlow format, and add a tfx_bsl.public.beam.RunInference step to the Dataflow pipeline.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 112
You are building a model to predict daily temperatures. You split the data randomly and then transformed the training and test datasets. Temperature data for model training is uploaded hourly. During testing, your model performed with 97% accuracy; however, after deploying to production, the model's accuracy dropped to 66%. How can you make your production model more accurate?
A. Normalize the data for the training, and test datasets as two separate steps.
B. Split the training and test data based on time rather than a random split to avoid leakage
C. Add more data to your test set to ensure that you have a fair distribution and sample for testing
D. Apply data transformations before splitting, and cross-validate to make sure that the transformations are applied to both the training and test sets.
Answer: (SHOW ANSWER)
https://community.rapidminer.com/discussion/32592/normalising-data-before-data-split-or-after

NEW QUESTION: 113
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a machine learning specialist will build a binary classifier based on two features: age of account, denoted by x, and transaction month, denoted by y. The class distributions are illustrated in the provided figure. The positive class is portrayed in red, while the negative class is portrayed in black.
Which model would have the HIGHEST accuracy?
A. Single perceptron with a Tanh activation function
B. Decision tree
C. Support vector machine (SVM) with a radial basis function kernel
D. Linear support vector machine (SVM)
Answer: C (LEAVE A REPLY)

NEW QUESTION: 114
A Machine Learning Specialist deployed a model that provides product recommendations on a company's website. Initially, the model was performing very well and resulted in customers buying more products on average. However, within the past few months, the Specialist has noticed that the effect of product recommendations has diminished and customers are starting to return to their original habits of spending less.
The Specialist is unsure of what happened, as the model has not changed from its initial deployment over a year ago.
Which method should the Specialist try to improve model performance?
A. The model should be periodically retrained using the original training data plus new data as product inventory changes.
B. The model should be periodically retrained from scratch using the original data while adding a regularization term to handle product inventory changes
C. The model needs to be completely re-engineered because it is unable to handle product inventory changes.
D. The model's hyperparameters should be periodically updated to prevent drift.

Answer: A (LEAVE A REPLY)

NEW QUESTION: 115
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. Configure Kubeflow to run on Google Kubernetes Engine and receive training jobs through TFJob
B. Set up Slurm workload manager to receive jobs that can be scheduled to run on your cloud infrastructure.
C. Use the AI Platform custom containers feature to receive training jobs using any framework
D. Create a library of VM images on Compute Engine; and publish these images on a centralized repository
Answer: (SHOW ANSWER)

NEW QUESTION: 116
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. Set the prefetch option equal to the training batch size
B. Increase the buffer size for the shuffle option.
C. Reduce the value of the repeat parameter
D. Use the interleave option for reading data
E. Decrease the batch size argument in your transformation
Answer: D,E (LEAVE A REPLY)