Amazon.AWS-Certified-Machine-Learning-Specialty.v2022-11-29.q177

Exam Code:	AWS-Certified-Machine-Learning-Specialty
Exam Name:	AWS Certified Machine Learning - Specialty
Certification Provider:	Amazon
Free Question Number:	177
Version:	v2022-11-29
# of views:	2240
# of Questions views:	1770
https://www.freepdfdumps.com/Amazon.AWS-Certified-Machine-Learning-Specialty.v2022-11-29.q177.html	

NEW QUESTION: 1
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. Use gated recurrent units (GRUs) instead of LSTM and run the training process until the validation loss stops decreasing.
B. 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.
C. Initialize the words by word2vec embeddings pretrained on a large collection of news articles related to the energy sector.
D. Reduce the learning rate and run the training process until the training loss stops decreasing.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 2
An ecommerce company is automating the categorization of its products based on images. A data scientist has trained a computer vision model using the Amazon SageMaker image classification algorithm. The images for each product are classified according to specific product lines. The accuracy of the model is too low when categorizing new products. All of the product images have the same dimensions and are stored within an Amazon S3 bucket. The company wants to improve the model so it can be used for new products as soon as possible.
Which steps would improve the accuracy of the solution? (Choose three.)

A. Use the SageMaker semantic segmentation algorithm to train a new model to achieve improved accuracy.
B. Use the Amazon Rekognition DetectLabels API to classify the products in the dataset.
C. Augment the images in the dataset. Use open source libraries to crop, resize, flip, rotate, and adjust the brightness and contrast of the images.
D. Use a SageMaker notebook to implement the normalization of pixels and scaling of the images. Store the new dataset in Amazon S3.
E. Use Amazon Rekognition Custom Labels to train a new model.
F. Check whether there are class imbalances in the product categories, and apply oversampling or undersampling as required. Store the new dataset in Amazon S3.
Answer: B,C,E (LEAVE A REPLY)
Reference:
https://towardsdatascience.com/image-processing-techniques-for-computer-vision-11f92f511e21
https://docs.aws.amazon.com/rekognition/latest/customlabels-dg/training-model.html

NEW QUESTION: 3
A Machine Learning team uses Amazon SageMaker to train an Apache MXNet handwritten digit classifier model using a research dataset. The team wants to receive a notification when the model is overfitting.
Auditors want to view the Amazon SageMaker log activity report to ensure there are no unauthorized API calls.
What should the Machine Learning team do to address the requirements with the least amount of code and fewest steps?
A. Implement an AWS Lambda function to long Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
B. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
C. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Set up Amazon SNS to receive a notification when the model is overfitting.
D. Implement an AWS Lambda function to log Amazon SageMaker API calls to AWS CloudTrail. Add code to push a custom metric to Amazon CloudWatch. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 4
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: (SHOW ANSWER)

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

NEW QUESTION: 5

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: 6

A machine learning (ML) specialist is administering a production Amazon SageMaker endpoint with model monitoring configured. Amazon SageMaker Model Monitor detects violations on the SageMaker endpoint, so the ML specialist retrains the model with the latest dataset. This dataset is statistically representative of the current production traffic. The ML specialist notices that even after deploying the new SageMaker model and running the first monitoring job, the SageMaker endpoint still has violations.

What should the ML specialist do to resolve the violations?

A. Delete the endpoint and recreate it with the original configuration.

B. Manually trigger the monitoring job to re-evaluate the SageMaker endpoint traffic sample.

C. Run the Model Monitor baseline job again on the new training set. Configure Model Monitor to use the new baseline.

D. Retrain the model again by using a combination of the original training set and the new training set.

Answer: C (LEAVE A REPLY)

NEW QUESTION: 7
A Machine Learning Engineer is preparing a data frame for a supervised learning task with the Amazon SageMaker Linear Learner algorithm. The ML Engineer notices the target label classes are highly imbalanced and multiple feature columns contain missing values. The proportion of missing values across the entire data frame is less than 5%.
What should the ML Engineer do to minimize bias due to missing values?
A. Replace each missing value by the mean or median across non-missing values in same row.
B. Delete observations that contain missing values because these represent less than 5% of the data.
C. Replace each missing value by the mean or median across non-missing values in the same column.
D. For each feature, approximate the missing values using supervised learning based on other features.
Answer: (SHOW ANSWER)
Use supervised learning to predict missing values based on the values of other features. Different supervised learning approaches might have different performances, but any properly implemented supervised learning approach should provide the same or better approximation than mean or median approximation, as proposed in responses A and C.
Supervised learning applied to the imputation of missing values is an active field of research.

NEW QUESTION: 8
A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:
Total number of images available = 1,000
Test set images = 100 (constant test set)
The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.
Which techniques can be used by the ML Specialist to improve this specific test error?
A. Increase the training data by adding variation in rotation for training images.
B. Increase the number of epochs for model training
C. Increase the number of layers for the neural network.
D. Increase the dropout rate for the second-to-last layer.
Answer: A (LEAVE A REPLY)
The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.

NEW QUESTION: 9
A company that runs an online library is implementing a chatbot using Amazon Lex to provide book recommendations based on category. This intent is fulfilled by an AWS Lambda function that queries an Amazon DynamoDB table for a list of book titles, given a particular category. For

testing, there are only three categories implemented as the custom slot types: "comedy," "adventure," and "documentary." A machine learning (ML) specialist notices that sometimes the request cannot be fulfilled because Amazon Lex cannot understand the category spoken by users with utterances such as "funny," "fun," and "humor." The ML specialist needs to fix the problem without changing the Lambda code or data in DynamoDB.
How should the ML specialist fix the problem?
A. Add the unrecognized words as synonyms in the custom slot type.
B. Use the AMAZON.SearchQuery built-in slot types for custom searches in the database.
C. Add the unrecognized words in the enumeration values list as new values in the slot type.
D. Create a new custom slot type, add the unrecognized words to this slot type as enumeration values, and use this slot type for the slot.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 10
A Machine Learning Specialist is using Amazon SageMaker to host a model for a highly available customer-facing application .
The Specialist has trained a new version of the model, validated it with historical data, and now wants to deploy it to production To limit any risk of a negative customer experience, the Specialist wants to be able to monitor the model and roll it back, if needed What is the SIMPLEST approach with the LEAST risk to deploy the model and roll it back, if needed?
A. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 5% of the traffic to the new variant. Revert traffic to the last version by resetting the weights if the model does not perform as expected.
B. Update the existing SageMaker endpoint to use a new configuration that is weighted to send 100% of the traffic to the new variant Revert traffic to the last version by resetting the weights if the model does not perform as expected.
C. Create a SageMaker endpoint and configuration for the new model version. Redirect production traffic to the new endpoint by updating the client configuration. Revert traffic to the last version if the model does not perform as expected.
D. Create a SageMaker endpoint and configuration for the new model version. Redirect production traffic to the new endpoint by using a load balancer Revert traffic to the last version if the model does not perform as expected.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 11
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. Change the cost function so that false positives have a higher impact on the cost value than false negatives.
B. Include a copy of the samples in the test dataset in the training dataset.
C. Add more deep trees to the random forest to enable the model to learn more features.
D. Change the cost function so that false negatives have a higher impact on the cost value than false positives.
E. Generate more positive samples by duplicating the positive samples and adding a small amount of noise to the duplicated data.
Answer: B,D (LEAVE A REPLY)

NEW QUESTION: 12
A data scientist has been running an Amazon SageMaker notebook instance for a few weeks. During this time, a new version of Jupyter Notebook was released along with additional software updates. The security team mandates that all running SageMaker notebook instances use the latest security and software updates provided by SageMaker.
How can the data scientist meet this requirements?
A. Create a new SageMaker notebook instance and mount the Amazon Elastic Block Store (Amazon EBS) volume from the original instance
B. Stop and then restart the SageMaker notebook instance
C. Call the UpdateNotebookInstanceLifecycleConfig API operation
D. Call the CreateNotebookInstanceLifecycleConfig API operation
Answer: B (LEAVE A REPLY)

NEW QUESTION: 13
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.
Predicted 0 1
Actual 0 99,966| 34 1 877|123

Which combination of steps should the Data Scientist take to reduce the number of false positive predictions by the model? (Select TWO.)
A. Change the XGBoost eval_metric parameter to optimize based on rmse instead of error.
B. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.
C. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.
D. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.
E. Change the XGBoost evaljnetric parameter to optimize based on AUC instead of error.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 14
A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket. A Machine Learning Specialist wants to use SQL to run queries on this data.
Which solution requires the LEAST effort to be able to query this data?
A. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.
B. Use AWS Batch to run ETL on the data and Amazon Aurora to run the queries.
C. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries.
D. Use AWS Glue to catalogue the data and Amazon Athena to run queries.
Answer: (SHOW ANSWER)

NEW QUESTION: 15
A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.
How should the Data Science team configure the notebook instance placement to meet these requirements?
A. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Use IAM policies to grant access to Amazon S3 and Amazon SageMaker.
B. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
C. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker.
D. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
Answer: (SHOW ANSWER)

NEW QUESTION: 16
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's hyperparameters should be periodically updated to prevent drift
B. The model needs to be completely re-engineered because it is unable to handle product inventory changes
C. The model should be periodically retrained from scratch using the original data while adding a regularization term to handle product inventory changes
D. The model should be periodically retrained using the original training data plus new data as product inventory changes
Answer: B (LEAVE A REPLY)

NEW QUESTION: 17
While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?

A. The model is predicting its target values perfectly.
B. There are too many variables in the model
C. The dataset cannot be accurately represented using the regression model
D. The model might have prediction errors over a range of target values.
Answer: (SHOW ANSWER)

NEW QUESTION: 18
A company needs to quickly make sense of a large amount of data and gain insight from it. The data is in different formats, the schemas change frequently, and new data sources are added regularly. The company wants to use AWS services to explore multiple data sources, suggest schemas, and enrich and transform the dat a. The solution should require the least possible coding effort for the data flows and the least possible infrastructure management.
Which combination of AWS services will meet these requirements?
A. Amazon EMR for data discovery, enrichment, and transformation
Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
Amazon QuickSight for reporting and getting insights
B. Amazon Kinesis Data Analytics for data ingestion
Amazon EMR for data discovery, enrichment, and transformation
Amazon Redshift for querying and analyzing the results in Amazon S3
C. AWS Glue for data discovery, enrichment, and transformation
Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL
Amazon QuickSight for reporting and getting insights
D. AWS Data Pipeline for data transfer

AWS Step Functions for orchestrating AWS Lambda jobs for data discovery, enrichment, and transformation Amazon Athena for querying and analyzing the results in Amazon S3 using standard SQL

Answer: A (LEAVE A REPLY)

Amazon QuickSight for reporting and getting insights

NEW QUESTION: 19

A real-estate company is launching a new product that predicts the prices of new houses. The historical data for the properties and prices is stored in .csv format in an Amazon S3 bucket. The data has a header, some categorical fields, and some missing values. The company's data scientists have used Python with a common open-source library to fill the missing values with zeros. The data scientists have dropped all of the categorical fields and have trained a model by using the open-source linear regression algorithm with the default parameters.

The accuracy of the predictions with the current model is below 50%. The company wants to improve the model performance and launch the new product as soon as possible.

Which solution will meet these requirements with the LEAST operational overhead?

A. Create an Amazon SageMaker notebook with a new IAM role that is associated with the notebook. Pull the dataset from the S3 bucket. Explore different combinations of feature engineering transformations, regression algorithms, and hyperparameters. Compare all the results in the notebook, and deploy the most accurate configuration in an endpoint for predictions.

B. Create an IAM role with access to Amazon S3, Amazon SageMaker, and AWS Lambda. Create a training job with the SageMaker built-in XGBoost model pointing to the bucket with the dataset. Specify the price as the target feature. Wait for the job to complete. Load the model artifact to a Lambda function for inference on prices of new houses.

C. Create an IAM role for Amazon SageMaker with access to the S3 bucket. Create a SageMaker AutoML job with SageMaker Autopilot pointing to the bucket with the dataset. Specify the price as the target attribute. Wait for the job to complete. Deploy the best model for predictions.

D. Create a service-linked role for Amazon Elastic Container Service (Amazon ECS) with access to the S3 bucket. Create an ECS cluster that is based on an AWS Deep Learning Containers image. Write the code to perform the feature engineering. Train a logistic regression model for predicting the price, pointing to the bucket with the dataset. Wait for the training job to complete. Perform the inferences.

Answer: D (LEAVE A REPLY)

NEW QUESTION: 20

A company wants to predict the sale prices of houses based on available historical sales dat a. 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: 21
A Machine Learning Specialist is developing a custom video recommendation model for an application The dataset used to train this model is very large with millions of data points and is hosted in an Amazon S3 bucket The Specialist wants to avoid loading all of this data onto an Amazon SageMaker notebook instance because it would take hours to move and will exceed the attached 5 GB Amazon EBS volume on the notebook instance.
Which approach allows the Specialist to use all the data to train the model?
A. Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
B. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to the instance. Train on a small amount of the data to verify the training code and hyperparameters. Go back to Amazon SageMaker and train using the full dataset
C. Use AWS Glue to train a model using a small subset of the data to confirm that the data will be compatible with Amazon SageMaker. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
D. Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 22
A Machine Learning Specialist is assigned a TensorFlow project using Amazon SageMaker for training, and needs to continue working for an extended period with no Wi-Fi access.
Which approach should the Specialist use to continue working?
A. Install Python 3 and boto3 on their laptop and continue the code development using that environment.
B. Download the SageMaker notebook to their local environment then install Jupyter Notebooks on their laptop and continue the development in a local notebook.

C. Download the TensorFlow Docker container used in Amazon SageMaker from GitHub to their local environment, and use the Amazon SageMaker Python SDK to test the code.
D. Download TensorFlow from tensorflow.org to emulate the TensorFlow kernel in the SageMaker environment.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 23
During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates.
What is the MOST likely cause of this issue?
A. The class distribution in the dataset is imbalanced.
B. Dataset shuffling is disabled.
C. The batch size is too big.
D. The learning rate is very high.
Answer: D (LEAVE A REPLY)
https://towardsdatascience.com/deep-learning-personal-notes-part-1-lesson-2-8946fe970b95

NEW QUESTION: 24
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 do1?
A. Build the Docker container to be NVIDIA-Docker compatible
B. Bundle the NVIDIA drivers with the Docker image
C. Set the GPU flag in the Amazon SageMaker Create TrainingJob request body
D. Organize the Docker container's file structure to execute on GPU instances.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 25
A 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 normalization to ensure each field will have a mean of 0 and a variance of 1 to remove any significant magnitude
B. Apply the orthogonal sparse Diagram (OSB) transformation to apply a fixed-size sliding window to generate new features of a similar magnitude.
C. Apply the Cartesian product transformation to create new combinations of fields that are independent of the magnitude

D. Apply quantile binning to group the data into categorical bins to keep any relationships in the data by replacing the magnitude with distribution
Answer: (SHOW ANSWER)

NEW QUESTION: 26
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? (Select 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 negatives have a higher impact on the cost value than false positives
C. Generate more positive samples by duplicating the positive samples and adding a small amount of noise to the duplicated data.
D. Change the cost function so that false positives have a higher impact on the cost value than false negatives
E. indicate a copy of the samples in the test database in the training dataset
Answer: B,E (LEAVE A REPLY)

NEW QUESTION: 27
A bank's Machine Learning team is developing an approach for credit card fraud detection The company has a large dataset of historical data labeled as fraudulent The goal is to build a model to take the information from new transactions and predict whether each transaction is fraudulent or not Which built-in Amazon SageMaker machine learning algorithm should be used for modeling this problem?
A. K-means
B. Seq2seq
C. Random Cut Forest (RCF)
D. XGBoost
Answer: A (LEAVE A REPLY)

NEW QUESTION: 28

A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket A Machine Learning Specialist wants to use SQL to run queries on this data. Which solution requires the LEAST effort to be able to query this data?

A. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.

B. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries

C. Use AWS Batch to run ETL on the data and Amazon Aurora to run the quenes

D. Use AWS Glue to catalogue the data and Amazon Athena to run queries

Answer: B (LEAVE A REPLY)

NEW QUESTION: 29

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. Write a direct connection to the SQL database within the notebook and pull data in

B. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.

C. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.

D. Move the data to Amazon DynamoDB and set up a connection to DynamoDB within the notebook to pull data in

Answer: B (LEAVE A REPLY)

NEW QUESTION: 30

A financial company is trying to detect credit card fraud. The company observed that, on average, 2% of credit card transactions were fraudulent. A data scientist trained a classifier on a year's worth of credit card transactions dat a. The model needs to identify the fraudulent transactions (positives) from the regular ones (negatives). The company's goal is to accurately capture as many positives as possible.

Which metrics should the data scientist use to optimize the model? (Choose two.)

A. Area under the precision-recall curve

B. False positive rate

C. Accuracy

D. True positive rate

E. Specificity

Answer: A,D (LEAVE A REPLY)

NEW QUESTION: 31

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 be fraudulent How should the Specialist frame this business problem'?

A. Binary classification

B. Regression classification

C. Multi-category classification

D. Streaming classification

Answer: D (LEAVE A REPLY)

NEW QUESTION: 32

A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.

The model produces the following confusion matrix after evaluating on a test dataset of 100 customers:

Based on the model evaluation results, why is this a viable model for production?

A. The precision of the model is 86%, which is greater than the accuracy of the model.

B. The precision of the model is 86%, which is less than the accuracy of the model.

C. The model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives.

D. The model is 86% accurate and the cost incurred by the company as a result of false negatives is less than the false positives.

Answer: D (LEAVE A REPLY)

NEW QUESTION: 33

A Machine Learning Specialist at a company sensitive to security is preparing a dataset for model training.

The dataset is stored in Amazon S3 and contains Personally Identifiable Information (PII). The dataset:
* Must be accessible from a VPC only.
* Must not traverse the public internet.
How can these requirements be satisfied?
A. Create a VPC endpoint and apply a bucket access policy that allows access from the given VPC endpoint and an Amazon EC2 instance.
B. Create a VPC endpoint and apply a bucket access policy that restricts access to the given VPC endpoint and the VPC.
C. Create a VPC endpoint and use Network Access Control Lists (NACLs) to allow traffic between only the given VPC endpoint and an Amazon EC2 instance.
D. Create a VPC endpoint and use security groups to restrict access to the given VPC endpoint and an Amazon EC2 instance.
Answer: (SHOW ANSWER)

NEW QUESTION: 34
A Machine Learning Specialist is designing a scalable data storage solution for Amazon SageMaker. There is an existing TensorFlow-based model implemented as a train.py script that relies on static training data that is currently stored as TFRecords.
Which method of providing training data to Amazon SageMaker would meet the business requirements with the LEAST development overhead?
A. Use Amazon SageMaker script mode and use train.py unchanged. Point the Amazon SageMaker training invocation to the local path of the data without reformatting the training data.
B. Use Amazon SageMaker script mode and use train.py unchanged. Put the TFRecord data into an Amazon S3 bucket. Point the Amazon SageMaker training invocation to the S3 bucket without reformatting the training data.
C. Rewrite the train.py script to add a section that converts TFRecords to protobuf and ingests the protobuf data instead of TFRecords.
D. Prepare the data in the format accepted by Amazon SageMaker. Use AWS Glue or AWS Lambda to reformat and store the data in an Amazon S3 bucket.
Answer: B (LEAVE A REPLY)
https://github.com/aws-samples/amazon-sagemaker-script-mode/blob/master/tf-horovod-inference-pipeline/train.py

NEW QUESTION: 35
A Marketing Manager at a pet insurance company plans to launch a targeted marketing campaign on social media to acquire new customers Currently, the company has the following data in Amazon Aurora
* Profiles for all past and existing customers
* Profiles for all past and existing insured pets
* Policy-level information

* Premiums received
* Claims paid
What steps should be taken to implement a machine learning model to identify potential new customers on social media?
A. Use regression on customer profile data to understand key characteristics of consumer segments Find similar profiles on social media.
B. Use a recommendation engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
C. Use a decision tree classifier engine on customer profile data to understand key characteristics of consumer segments. Find similar profiles on social media
D. Use clustering on customer profile data to understand key characteristics of consumer segments Find similar profiles on social media.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 36
A Machine Learning Specialist is creating a new natural language processing application that processes a dataset comprised of 1 million sentences The aim is to then run Word2Vec to generate embeddings of the sentences and enable different types of predictions - Here is an example from the dataset
"The quck BROWN FOX jumps over the lazy dog "
Which of the following are the operations the Specialist needs to perform to correctly sanitize and prepare the data in a repeatable manner? (Select THREE)
A. Normalize all words by making the sentence lowercase
B. Correct the typography on "quck" to "quick."
C. Tokenize the sentence into words.
D. Perform part-of-speech tagging and keep the action verb and the nouns only
E. One-hot encode all words in the sentence
F. Remove stop words using an English stopword dictionary.
Answer: B,D,E (LEAVE A REPLY)

NEW QUESTION: 37
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. Normal distribution
B. Binomial distribution
C. Uniform distribution
D. Poisson distribution
Answer: B (LEAVE A REPLY)

NEW QUESTION: 38
Example Corp has an annual sale event from October to December. The company has sequential sales data from the past 15 years and wants to use Amazon ML to predict the sales for this year's upcoming event. Which method should Example Corp use to split the data into a training dataset and evaluation dataset?
A. Have Amazon ML split the data randomly.
B. Pre-split the data before uploading to Amazon S3
C. Perform custom cross-validation on the data
D. Have Amazon ML split the data sequentially.
Answer: (SHOW ANSWER)

NEW QUESTION: 39
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. Compressed JSON
C. Parquet files
D. RecordIO
Answer: (SHOW ANSWER)

NEW QUESTION: 40
A company is launching a new product and needs to build a mechanism to monitor comments about the company and its new product on social medi a. The company needs to be able to evaluate the sentiment expressed in social media posts, and visualize trends and configure alarms based on various thresholds.
The company needs to implement this solution quickly, and wants to minimize the infrastructure and data science resources needed to evaluate the messages. The company already has a solution in place to collect posts and store them within an Amazon S3 bucket.
What services should the data science team use to deliver this solution?
A. Trigger an AWS Lambda function when social media posts are added to the S3 bucket. Call Amazon Comprehend for each post to capture the sentiment in the message and record the sentiment in an Amazon DynamoDB table. Schedule a second Lambda function to query recently added records and send an Amazon Simple Notification Service (Amazon SNS) notification to notify analysts of trends.
B. Train a model in Amazon SageMaker by using the BlazingText algorithm to detect sentiment in the corpus of social media posts. Expose an endpoint that can be called by AWS Lambda. Trigger a Lambda function when posts are added to the S3 bucket to invoke the endpoint and

record the sentiment in an Amazon DynamoDB table and in a custom Amazon CloudWatch metric. Use CloudWatch alarms to notify analysts of trends.

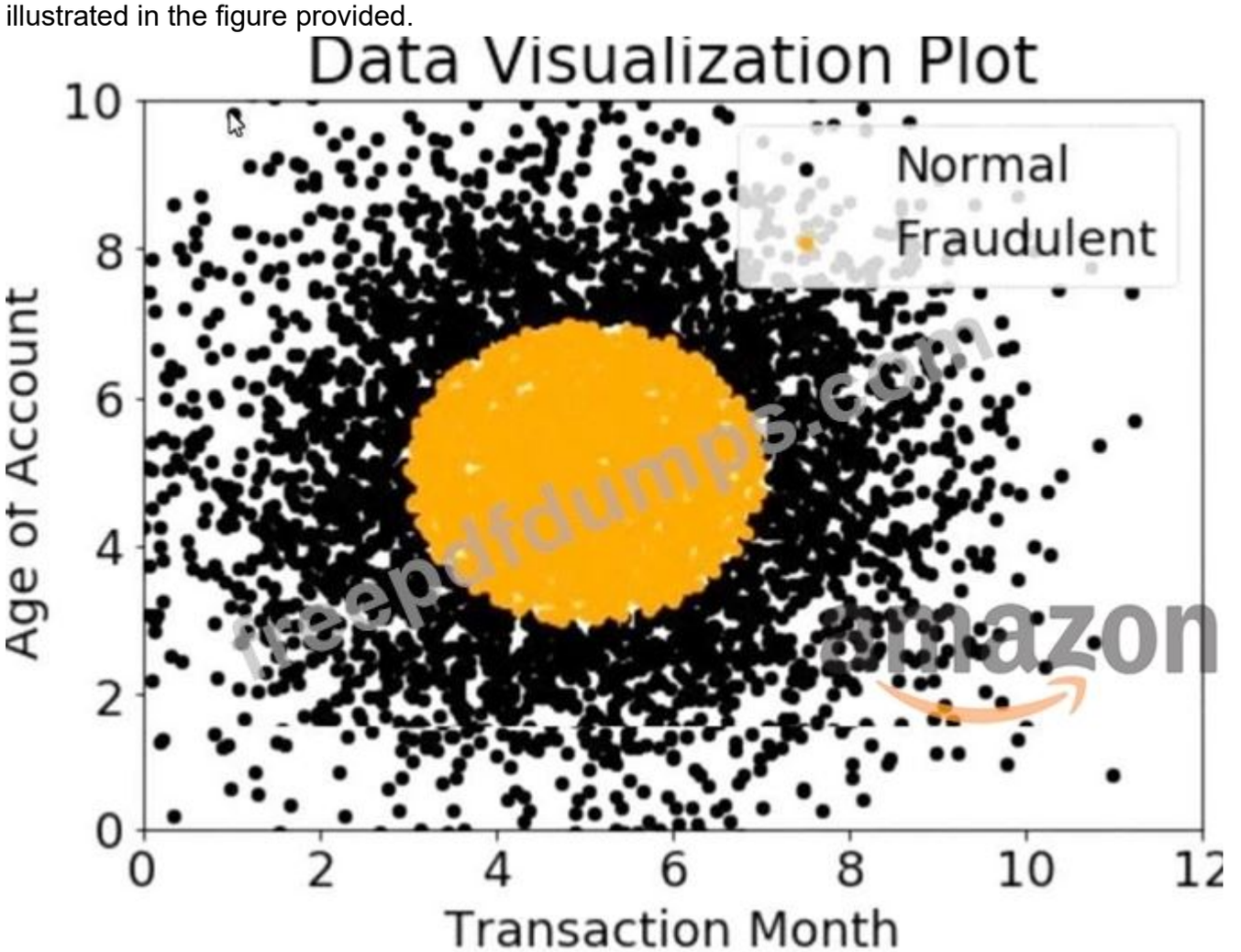
C. Trigger an AWS Lambda function when social media posts are added to the S3 bucket. Call Amazon Comprehend for each post to capture the sentiment in the message and record the sentiment in a custom Amazon CloudWatch metric and in S3. Use CloudWatch alarms to notify analysts of trends.

D. Train a model in Amazon SageMaker by using the semantic segmentation algorithm to model the semantic content in the corpus of social media posts. Expose an endpoint that can be called by AWS Lambda. Trigger a Lambda function when objects are added to the S3 bucket to invoke the endpoint and record the sentiment in an Amazon DynamoDB table. Schedule a second Lambda function to query recently added records and send an Amazon Simple Notification Service (Amazon SNS) notification to notify analysts of trends.

Answer: B (LEAVE A REPLY)

NEW QUESTION: 41

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



Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

A. Naive Bayesian classifier

B. Single Perceptron with sigmoidal activation function

C. Linear support vector machine (SVM)

D. Decision tree

Answer: B (LEAVE A REPLY)

NEW QUESTION: 42

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

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

A. Use AWS Glue to compress the data into the Apache Parquet format.

B. Transform the dataset into the RecordIO protobufformat.

C. Use the SageMaker batch transform feature to transform the training data into a DataFrame.

D. Use the SageMaker hyperparameter optimization feature to automatically optimize the data.

Answer: (SHOW ANSWER)

NEW QUESTION: 43

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? (Select TWO.)

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

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

C. indicate a copy of the samples in the test database in the training dataset

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

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

Answer: B,D (LEAVE A REPLY)

NEW QUESTION: 44
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. RecordIO
B. Parquet files
C. CSV files
D. Compressed JSON
Answer: (SHOW ANSWER)

NEW QUESTION: 45
A machine learning specialist needs to analyze comments on a news website with users across the globe. The specialist must find the most discussed topics in the comments that are in either English or Spanish.
What steps could be used to accomplish this task? (Choose two.)
A. Use Amazon Translate to translate from Spanish to English, if necessary. Use Amazon Comprehend topic modeling to find the topics.
B. Use an Amazon SageMaker BlazingText algorithm to find the topics independently from language. Proceed with the analysis.
C. Use Amazon Translate to translate from Spanish to English, if necessary. Use Amazon SageMaker Neural Topic Model (NTM) to find the topics.
D. Use an Amazon SageMaker seq2seq algorithm to translate from Spanish to English, if necessary. Use a SageMaker Latent Dirichlet Allocation (LDA) algorithm to find the topics.
E. Use Amazon Translate to translate from Spanish to English, if necessary. Use Amazon Lex to extract topics form the content.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 46
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 pre-trained weights in all layers including the last fully connected layer.
B. Initialize the model with random weights in all layers including 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 and replace the last fully connected layer.
Answer: (SHOW ANSWER)

NEW QUESTION: 47
A Machine Learning Specialist is designing a system for improving sales for a company. The objective is to use the large amount of information the company has on users' behavior and product preferences to predict which products users would like based on the users' similarity to other users.
What should the Specialist do to meet this objective?
A. Build a content-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
B. Build a collaborative filtering recommendation engine with Apache Spark ML on Amazon EMR.
C. Build a model-based filtering recommendation engine with Apache Spark ML on Amazon EMR.
D. Build a combinative filtering recommendation engine with Apache Spark ML on Amazon EMR.
Answer: B (LEAVE A REPLY)
Explanation
Many developers want to implement the famous Amazon model that was used to power the "People who bought this also bought these items" feature on Amazon.com. This model is based on a method called Collaborative Filtering. It takes items such as movies, books, and products that were rated highly by a set of users and recommending them to other users who also gave them high ratings. This method works well in domains where explicit ratings or implicit user actions can be gathered and analyzed.

NEW QUESTION: 48
While reviewing the histogram for residuals on regression evaluation data a Machine Learning Specialist notices that the residuals do not form a zero-centered bell shape as shown What does this mean?

A. The dataset cannot be accurately represented using the regression model
B. The model is predicting its target values perfectly.
C. The model might have prediction errors over a range of target values.
D. There are too many variables in the model
Answer: B (LEAVE A REPLY)

NEW QUESTION: 49
A machine learning (ML) specialist is using Amazon SageMaker hyperparameter optimization (HPO) to improve a model's accuracy. The learning rate parameter is specified in the following HPO configuration:

```
{
    "Name": "learning_rate",
    "MaxValue": "0.0001",
    "MinValue": "0.1"
}
```

During the results analysis, the ML specialist determines that most of the training jobs had a learning rate between 0.01 and 0.1. The best result had a learning rate of less than 0.01. Training jobs need to run regularly over a changing dataset. The ML specialist needs to find a tuning mechanism that uses different learning rates more evenly from the provided range between MinValue and MaxValue.

Which solution provides the MOST accurate result?

A. Modify the HPO configuration as follows:

```
{
    "Name": "learning_rate",
    "MaxValue" : "0.0001",
    "MinValue": "0.1"
    "ScalingType": "Logarithmic"
}
```

Select the most accurate hyperparameter configuration form this training job.

B. Modify the HPO configuration as follows:

```
{
    "Name": "learning_rate",
    "MaxValue" : "0.0001",
    "MinValue": "0.1"
    "ScalingType": "ReverseLogarithmic"
}
```

Select the most accurate hyperparameter configuration form this HPO job.

C. Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue while using the same number of training jobs for each HPO job:

[0.01, 0.1]

[0.001, 0.01]

[0.0001, 0.001]

Select the most accurate hyperparameter configuration form these three HPO jobs.

D. Run three different HPO jobs that use different learning rates form the following intervals for MinValue and MaxValue. Divide the number of training jobs for each HPO job by three:

[0.01, 0.1]

[0.001, 0.01]

[0.0001, 0.001]

Select the most accurate hyperparameter configuration form these three HPO jobs.

Answer: (SHOW ANSWER)

NEW QUESTION: 50

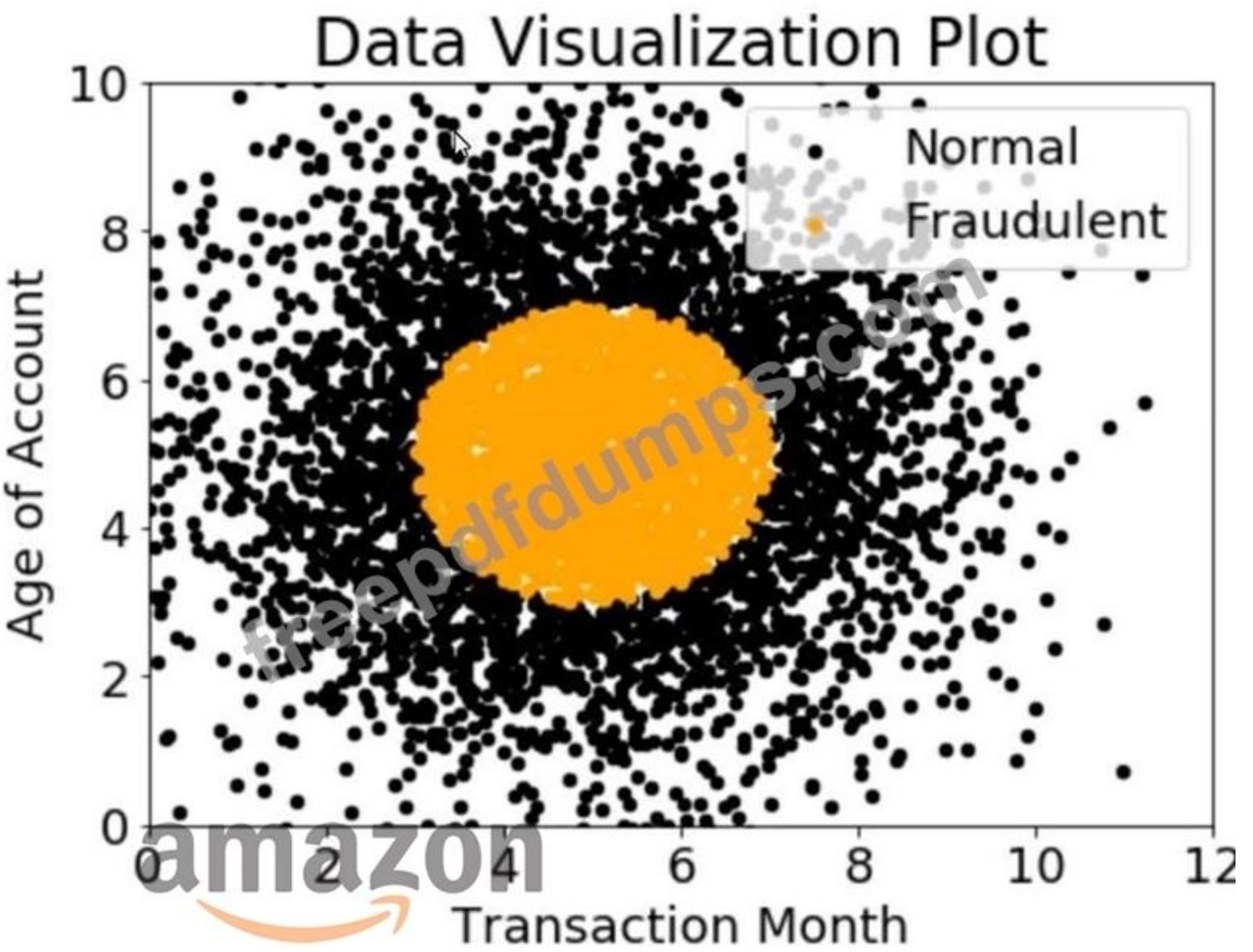
A Machine Learning team uses Amazon SageMaker to train an Apache MXNet handwritten digit classifier model using a research dataset. The team wants to receive a notification when the model is overfitting.

Auditors want to view the Amazon SageMaker log activity report to ensure there are no unauthorized API calls.

What should the Machine Learning team do to address the requirements with the least amount of code and fewest steps?

A. Implement an AWS Lambda function to long Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
B. Implement an AWS Lambda function to log Amazon SageMaker API calls to AWS CloudTrail. Add code to push a custom metric to Amazon CloudWatch. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
C. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Set up Amazon SNS to receive a notification when the model is overfitting.
D. Use AWS CloudTrail to log Amazon SageMaker API calls to Amazon S3. Add code to push a custom metric to Amazon CloudWatch. Create an alarm in CloudWatch with Amazon SNS to receive a notification when the model is overfitting.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 51
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

A. Decision tree

B. Linear support vector machine (SVM)

C. Naive Bayesian classifier

D. Single Perceptron with sigmoidal activation function

Answer: A (LEAVE A REPLY)

To identity fraudulent or not, you need a model with high accuracy and high recall (low false negative). The data in the graph is non-linear. Generally Decision tree gives better result for non-linear data than Naive Bayes classifier.

https://datascience.stackexchange.com/questions/6787/are-decision-tree-algorithms-linear-or-nonlinear

https://sebastianraschka.com/Articles/2014_naive_bayes_1.html

NEW QUESTION: 52

A company wants to create a data repository in the AWS Cloud for machine learning (ML) projects. The company wants to use AWS to perform complete ML lifecycles and wants to use Amazon S3 for the data storage. All of the company's data currently resides on premises and is 40 TB in size.

The company wants a solution that can transfer and automatically update data between the on-premises object storage and Amazon S3. The solution must support encryption, scheduling, monitoring, and data integrity validation.

Which solution meets these requirements?

A. Use the S3 sync command to compare the source S3 bucket and the destination S3 bucket. Determine which source files do not exist in the destination S3 bucket and which source files were modified.

B. Use S3 Batch Operations to pull data periodically from the on-premises storage. Enable S3 Versioning on the S3 bucket to protect against accidental overwrites.

C. Use AWS Transfer for FTPS to transfer the files from the on-premises storage to Amazon S3.

D. Use AWS DataSync to make an initial copy of the entire dataset. Schedule subsequent incremental transfers of changing data until the final cutover from on premises to AWS.

Answer: D (LEAVE A REPLY)

NEW QUESTION: 53

A machine learning specialist stores IoT soil sensor data in Amazon DynamoDB table and stores weather event data as JSON files in Amazon S3. The dataset in DynamoDB is 10 GB in size and the dataset in Amazon S3 is 5 GB in size. The specialist wants to train a model on this data to help predict soil moisture levels as a function of weather events using Amazon SageMaker.

Which solution will accomplish the necessary transformation to train the Amazon SageMaker model with the LEAST amount of administrative overhead?

A. Crawl the data using AWS Glue crawlers. Write an AWS Glue ETL job that merges the two tables and writes the output in CSV format to Amazon S3.
B. Enable Amazon DynamoDB Streams on the sensor table. Write an AWS Lambda function that consumes the stream and appends the results to the existing weather files in Amazon S3.
C. Launch an Amazon EMR cluster. Create an Apache Hive external table for the DynamoDB table and S3 data. Join the Hive tables and write the results out to Amazon S3.
D. Crawl the data using AWS Glue crawlers. Write an AWS Glue ETL job that merges the two tables and writes the output to an Amazon Redshift cluster.
Answer: (SHOW ANSWER)

NEW QUESTION: 54
A large consumer goods manufacturer has the following products on sale:
* 34 different toothpaste variants
* 48 different toothbrush variants
* 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3. Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products. The company wants to predict the demand for a new product that will soon be launched.
Which solution should a Machine Learning Specialist apply?
A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product.
C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
D. Train a custom XGBoost model to forecast demand for the new product.
Answer: (SHOW ANSWER)
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.
Reference: https://docs.aws.amazon.com/sagemaker/latest/dg/deepar.html

NEW QUESTION: 55
An office security agency conducted a successful pilot using 100 cameras installed at key locations within the main office. Images from the cameras were uploaded to Amazon S3 and tagged using Amazon Rekognition, and the results were stored in Amazon ES. The agency is now looking to expand the pilot into a full production system using thousands of video cameras in its office locations globally. The goal is to identify activities performed by non-employees in real time.
Which solution should the agency consider?

A. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection on each stream, and alert when nonemployees are detected.
B. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection of known employees, and alert when non-employees are detected.
C. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream. On each stream, use Amazon Rekognition Image to detect faces from a collection of known employees and alert when non-employees are detected.
D. Install AWS DeepLens cameras and use the DeepLens_Kinesis_Video module to stream video to Amazon Kinesis Video Streams for each camera. On each stream, run an AWS Lambda function to capture image fragments and then call Amazon Rekognition Image to detect faces from a collection of known employees, and alert when non-employees are detected.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 56
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)
https://docs.aws.amazon.com/sagemaker/latest/dg/monitoring-cloudwatch.html

NEW QUESTION: 57
An e commerce company wants to launch a new cloud-based product recommendation feature for its web application. Due to data localization regulations, any sensitive data must not leave its on-premises data center, and the product recommendation model must be trained and tested using nonsensitive data only. Data transfer to the cloud must use IPsec. The web application is

hosted on premises with a PostgreSQL database that contains all the dat a. The company wants the data to be uploaded securely to Amazon S3 each day for model retraining.
How should a machine learning specialist meet these requirements?
A. Create an AWS Glue job to connect to the PostgreSQL DB instance. Ingest all data through an AWS Site- to-Site VPN connection into Amazon S3 while removing sensitive data using a PySpark job.
B. Create an AWS Glue job to connect to the PostgreSQL DB instance. Ingest tables without sensitive data through an AWS Site-to-Site VPN connection directly into Amazon S3.
C. Use PostgreSQL logical replication to replicate all data to PostgreSQL in Amazon EC2 through AWS Direct Connect with a VPN connection. Use AWS Glue to move data from Amazon EC2 to Amazon S3.
D. Use AWS Database Migration Service (AWS DMS) with table mapping to select PostgreSQL tables with no sensitive data through an SSL connection. Replicate data directly into Amazon S3.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 58
An office security agency conducted a successful pilot using 100 cameras installed at key locations within the main office. Images from the cameras were uploaded to Amazon S3 and tagged using Amazon Rekognition, and the results were stored in Amazon ES. The agency is now looking to expand the pilot into a full production system using thousands of video cameras in its office locations globally. The goal is to identify activities performed by non-employees in real time.
Which solution should the agency consider?
A. Install AWS DeepLens cameras and use the DeepLens_Kinesis_video module to stream video to Amazon Kinesis Video Streams for each camera. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection on each stream, and alert when non-employees are detected.
B. Install AWS DeepLens cameras and use the DeepLens_Kinesis_video module to stream video to Amazon Kinesis Video Streams for each camera. On each stream, run an AWS Lambda function to capture image fragments and then call Amazon Rekognition Image to detect faces from a collection of known employees, and alert when non-employees are detected.
C. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video strearw On each stream, use Amazon Rekognition Image to detect faces from a collection of known employees and alert when non-employees are detected.
D. Use a proxy server at each local office and for each camera, and stream the RTSP feed to a unique Amazon Kinesis Video Streams video stream. On each stream, use Amazon Rekognition Video and create a stream processor to detect faces from a collection of known employees, and alert when non-employees are detected
Answer: A (LEAVE A REPLY)

NEW QUESTION: 59
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. Root Mean Square Error (RMSE)
B. Residual plots
C. Area under the curve
D. Confusion matrix
Answer: B (LEAVE A REPLY)
https://docs.aws.amazon.com/machine-learning/latest/dg/regression-model-insights.html

NEW QUESTION: 60
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: (SHOW ANSWER)
As it is a interactive online dictionary, we need pre-trained word embedding thus the answer is D. In addition, there is no mention that the online dictonary is unique and does not have a pre-trained word embedding.

NEW QUESTION: 61
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)

NEW QUESTION: 62
A media company with a very large archive of unlabeled images, text, audio, and video footage wishes to index its assets to allow rapid identification of relevant content by the Research team. The company wants to use machine learning to accelerate the efforts of its in-house researchers who have limited machine learning expertise.
Which is the FASTEST route to index the assets?
A. Use Amazon Rekognition, Amazon Comprehend, and Amazon Transcribe to tag data into distinct categories/classes.
B. Create a set of Amazon Mechanical Turk Human Intelligence Tasks to label all footage.
C. Use Amazon Transcribe to convert speech to text. Use the Amazon SageMaker Neural Topic Model (NTM) and Object Detection algorithms to tag data into distinct categories/classes.
D. Use the AWS Deep Learning AMI and Amazon EC2 GPU instances to create custom models for audio transcription and topic modeling, and use object detection to tag data into distinct categories/classes.
Answer: (SHOW ANSWER)

NEW QUESTION: 63
An aircraft engine manufacturing company is measuring 200 performance metrics in a time-series. Engineers want to detect critical manufacturing defects in near-real time during testing. All of the data needs to be stored for offline analysis.
What approach would be the MOST effective to perform near-real time defect detection?
A. Use AWS IoT Analytics for ingestion, storage, and further analysis. Use Jupyter notebooks from within AWS IoT Analytics to carry out analysis for anomalies.
B. Use Amazon S3 for ingestion, storage, and further analysis. Use an Amazon EMR cluster to carry out Apache Spark ML k-means clustering to determine anomalies.
C. Use Amazon S3 for ingestion, storage, and further analysis. Use the Amazon SageMaker Random Cut Forest (RCF) algorithm to determine anomalies.
D. Use Amazon Kinesis Data Firehose for ingestion and Amazon Kinesis Data Analytics Random Cut Forest (RCF) to perform anomaly detection. Use Kinesis Data Firehose to store data in Amazon S3 for further
Answer: B (LEAVE A REPLY)

analysis.

NEW QUESTION: 64
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. Push the data from Microsoft SQL Server to Amazon S3 using an AWS Data Pipeline and provide the S3 location within the notebook.
D. Move the data to Amazon ElastiCache using AWS DMS and set up a connection within the notebook to pull data in for fast access.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 65
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)
1. Use Glue Crawler to build scheme from the structured csv file.
2. Configure and run a job to transform the data from CSV to Parquet.
https://aws.amazon.com/blogs/big-data/build-a-data-lake-foundation-with-aws-glue-and-amazon-s3/

NEW QUESTION: 66
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. Compressed JSON
B. CSV files
C. Parquet files
D. RecordIO
Answer: C (LEAVE A REPLY)

NEW QUESTION: 67
A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.
How should the Data Science team configure the notebook instance placement to meet these requirements?
A. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker.
B. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Use IAM policies to grant access to Amazon S3 and Amazon SageMaker.
C. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
D. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 68
Amazon Connect has recently been tolled out across a company as a contact call center The solution has been configured to store voice call recordings on Amazon S3 The content of the voice calls are being analyzed for the incidents being discussed by the call operators Amazon Transcribe is being used to convert the audio to text, and the output is stored on Amazon S3
Which approach will provide the information required for further analysis?
A. Use Amazon Translate with the transcribed files to train and build a model for the key topics
B. Use Amazon Comprehend with the transcribed files to build the key topics
C. Use the Amazon SageMaker k-Nearest-Neighbors (kNN) algorithm on the transcribed files to generate a word embeddings dictionary for the key topics
D. Use the AWS Deep Learning AMI with Gluon Semantic Segmentation on the transcribed files to train and build a model for the key topics
Answer: A (LEAVE A REPLY)

NEW QUESTION: 69

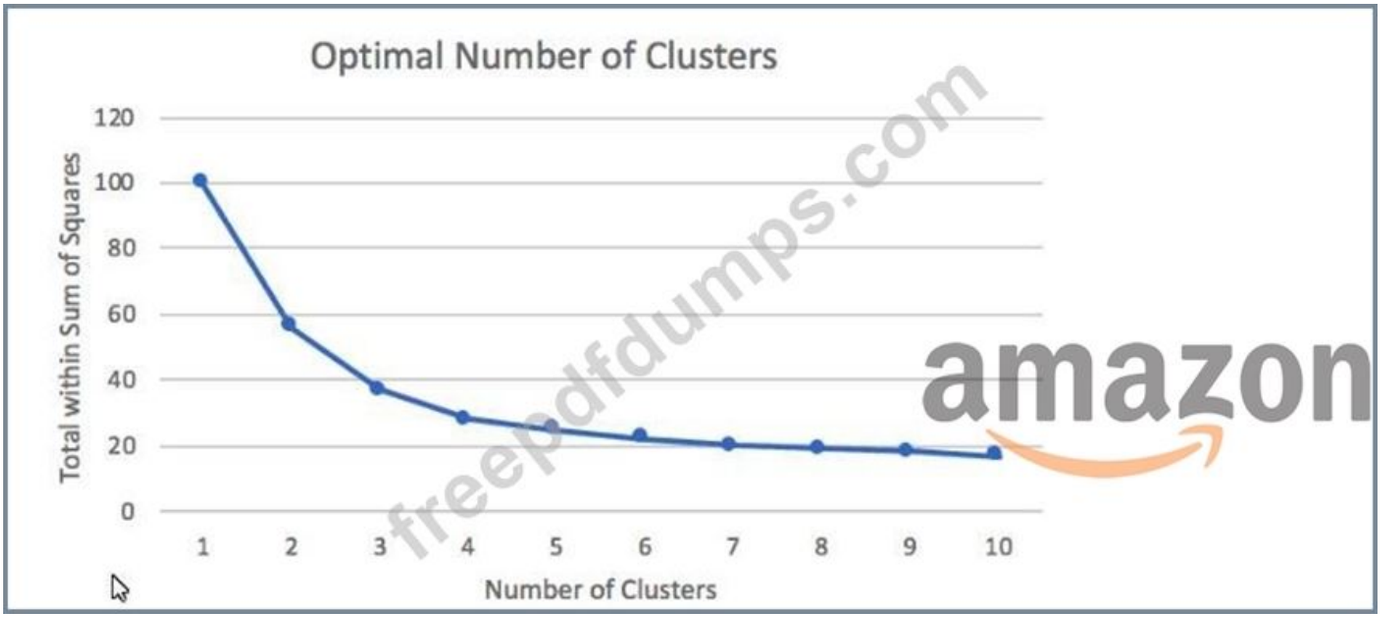
A Data Scientist is training a multilayer perception (MLP) on a dataset with multiple classes. The target class of interest is unique compared to the other classes within the dataset, but it does not achieve and acceptable recall metric. The Data Scientist has already tried varying the number and size 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. Gather more data using Amazon Mechanical Turk and then retrain
B. Train an XGBoost model instead of an MLP
C. Add class weights to the MLP's loss function and then retrain
D. Train an anomaly detection model instead of an MLP
Answer: B (LEAVE A REPLY)

NEW QUESTION: 70
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 historical data insights; Amazon DynamoDB streams 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 near-real-time data insights; Amazon Kinesis Data Firehose for clickstream analytics; AWS Glue to generate personalized product recommendations
Answer: A (LEAVE A REPLY)

NEW QUESTION: 71
A Machine Learning Specialist is working with a large cybersecurity company that manages security events in real time for companies around the world. The cybersecurity company wants to design a solution that will allow it to use machine learning to score malicious events as anomalies

on the data as it is being ingested. The company also wants be able to save the results in its data lake for later processing and analysis.
What is the MOST efficient way to accomplish these tasks?
A. Ingest the data using Amazon Kinesis Data Firehose, and use Amazon Kinesis Data Analytics Random Cut Forest (RCF) for anomaly detection. Then use Kinesis Data Firehose to stream the results to Amazon S3.
B. Ingest the data into Apache Spark Streaming using Amazon EMR, and use Spark MLlib with k-means to perform anomaly detection. Then store the results in an Apache Hadoop Distributed File System (HDFS) using Amazon EMR with a replication factor of three as the data lake.
C. Ingest the data and store it in Amazon S3. Use AWS Batch along with the AWS Deep Learning AMIs to train a k-means model using TensorFlow on the data in Amazon S3.
D. Ingest the data and store it in Amazon S3. Have an AWS Glue job that is triggered on demand transform the new data. Then use the built-in Random Cut Forest (RCF) model within Amazon SageMaker to detect anomalies in the data.
Answer: A (LEAVE A REPLY)
https://aws.amazon.com/tw/blogs/machine-learning/use-the-built-in-amazon-sagemaker-random-cut-forest-algorithm-for-anomaly-detection/

NEW QUESTION: 72
A Machine Learning Specialist prepared the following graph displaying the results of k-means for k = [1:10]



Considering the graph, what is a reasonable selection for the optimal choice of k?
A. 7
B. 1
C. 10
D. 4
Answer: A (LEAVE A REPLY)

NEW QUESTION: 73
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 served by the models.
Which solution satisfies these requirements with MINIMAL effort?
A. 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.
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 Neo to take into account different types of medical devices. Programmatically control which model is invoked for inference based on the medical device type.
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: B (LEAVE A REPLY)
A/B testing with Amazon SageMaker is required in the Exam. In A/B testing, you test different variants of your models and compare how each variant performs. Amazon SageMaker enables you to test multiple models or model versions behind the `same endpoint` using `production variants`. Each production variant identifies a machine learning (ML) model and the resources deployed for hosting the model. To test multiple models by `distributing traffic` between them, specify the `percentage of the traffic` that gets routed to each model by specifying the `weight` for each `production variant` in the endpoint configuration.
https://docs.aws.amazon.com/sagemaker/latest/dg/model-ab-testing.html#model-testing-target-variant

NEW QUESTION: 74
A Machine Learning Specialist uploads a dataset to an Amazon S3 bucket protected with server-side encryption using AWS KMS.
How should the ML Specialist define the Amazon SageMaker notebook instance so it can read the same dataset from Amazon S3?
A. Define security group(s) to allow all HTTP inbound/outbound traffic and assign those security group(s) to the Amazon SageMaker notebook instance.

B. onfigure the Amazon SageMaker notebook instance to have access to the VPC. Grant permission in the KMS key policy to the notebook's KMS role.
C. Assign an IAM role to the Amazon SageMaker notebook with S3 read access to the dataset. Grant permission in the KMS key policy to that role.
D. Assign the same KMS key used to encrypt data in Amazon S3 to the Amazon SageMaker notebook instance.
Answer: (SHOW ANSWER)
https://docs.aws.amazon.com/sagemaker/latest/dg/encryption-at-rest.html

NEW QUESTION: 75
An insurance company is developing a new device for vehicles that uses a camera to observe drivers' behavior and alert them when they appear distracted. The company created approximately 10,000 training images in a controlled environment that a Machine Learning Specialist will use to train and evaluate machine learning models.
During the model evaluation, the Specialist notices that the training error rate diminishes faster as the number of epochs increases and the model is not accurately inferring on the unseen test images.
Which of the following should be used to resolve this issue? (Choose two.)
A. Use gradient checking in the model.
B. Make the neural network architecture complex.
C. Add vanishing gradient to the model.
D. Perform data augmentation on the training data.
E. Add L2 regularization to the model.
Answer: (SHOW ANSWER)

NEW QUESTION: 76
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. 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.
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. Build the image and upload it to Docker Hub.
D. Build the Docker image with the inference code. Tag the Docker image with the registry hostname and upload it to Amazon ECR.
Answer: (SHOW ANSWER)

NEW QUESTION: 77
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. Decrease the retention period for the data stream.
B. Increase the number of S3 prefixes for the delivery stream to write to.
C. Add more consumers using the Kinesis Client Library (KCL).
D. Increase the number of shards for the data stream.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 78
A Machine Learning Specialist observes several performance problems with the training portion of a machine learning solution on Amazon SageMaker The solution uses a large training dataset 2 TB in size and is using the SageMaker k-means algorithm The observed issues include the unacceptable length of time it takes before the training job launches and poor I/O throughput while training the model What should the Specialist do to address the performance issues with the current solution?
A. Ensure that the input mode for the training job is set to Pipe.
B. Compress the training data into Apache Parquet format.
C. Copy the training dataset to an Amazon EFS volume mounted on the SageMaker instance.
D. Use the SageMaker batch transform feature
Answer: B (LEAVE A REPLY)

NEW QUESTION: 79
A financial services company is building a robust serverless data lake on Amazon S3. The data lake should be flexible and meet the following requirements:

* Support querying old and new data on Amazon S3 through Amazon Athena and Amazon Redshift Spectrum.
* Support event-driven ETL pipelines
* Provide a quick and easy way to understand metadata
Which approach meets these requirements?
A. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Glue ETL job, and an external Apache Hive metastore to search and discover metadata.
B. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Batch job, and an external Apache Hive metastore to search and discover metadata.
C. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.
D. 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.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 80
A Data Science team is designing a dataset repository where it will store a large amount of training data commonly used in its machine learning models. As Data Scientists may create an arbitrary number of new datasets every day the solution has to scale automatically and be cost-effective. Also, it must be possible to explore the data using SQL.
Which storage scheme is MOST adapted to this scenario?
A. Store datasets as files in an Amazon EBS volume attached to an Amazon EC2 instance.
B. Store datasets as tables in a multi-node Amazon Redshift cluster.
C. Store datasets as files in Amazon S3.
D. Store datasets as global tables in Amazon DynamoDB.
Answer: (SHOW ANSWER)

NEW QUESTION: 81
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 an Amazon S3-backed data lake to store the raw images, and set up the permissions using bucket policies.
B. Setup up Amazon EMR with Hadoop Distributed File System (HDFS) to store the files, and restrict access to the EMR instances using IAM policies.
C. 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.

D. Configure Amazon EFS with IAM policies to make the data available to Amazon EC2 instances owned by the IAM users.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 82
A Machine Learning Specialist is configuring Amazon SageMaker so multiple Data Scientists can access notebooks, train models, and deploy endpoints. To ensure the best operational performance, the Specialist needs to be able to track how often the Scientists are deploying models, GPU and CPU utilization on the deployed SageMaker endpoints, and all errors that are generated when an endpoint is invoked.
Which services are integrated with Amazon SageMaker to track this information? (Choose two.)
A. AWS CloudTrail
B. AWS Health
C. AWS Trusted Advisor
D. Amazon CloudWatch
E. AWS Config
Answer: A,D (LEAVE A REPLY)
https://aws.amazon.com/sagemaker/faqs/

NEW QUESTION: 83
A large consumer goods manufacturer has the following products on sale
* 34 different toothpaste variants
* 48 different toothbrush variants
* 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched Which solution should a Machine Learning Specialist apply?
A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product
C. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
D. Train a custom XGBoost model to forecast demand for the new product
Answer: B (LEAVE A REPLY)
Explanation
The Amazon SageMaker DeepAR forecasting algorithm is a supervised learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). Classical forecasting methods, such as autoregressive integrated moving average (ARIMA) or exponential smoothing (ETS), fit a single model to each individual time series. They then use that model to extrapolate the time series into the future.

NEW QUESTION: 84
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. Random Cut Forest (RCF) over random subsamples to identify patterns in the customer database
B. Latent Dirichlet Allocation (LDA) for the given collection of discrete data to identify patterns in the customer database.
C. Collaborative filtering based on user interactions and correlations to identify patterns in the customer database
D. A neural network with a minimum of three layers and random initial weights to identify patterns in the customer database
Answer: C (LEAVE A REPLY)

NEW QUESTION: 85
A technology startup is using complex deep neural networks and GPU compute to recommend the company's products to its existing customers based upon each customer's habits and interactions. The solution currently pulls each dataset from an Amazon S3 bucket before loading the data into a TensorFlow model pulled from the company's Git repository that runs locally. This job then runs for several hours while continually outputting its progress to the same S3 bucket. The job can be paused, restarted, and continued at any time in the event of a failure, and is run from a central queue.
Senior managers are concerned about the complexity of the solution's resource management and the costs involved in repeating the process regularly. They ask for the workload to be automated so it runs once a week, starting Monday and completing by the close of business Friday.
Which architecture should be used to scale the solution at the lowest cost?
A. Implement the solution using Amazon ECS running on Spot Instances and schedule the task using the ECS service scheduler
B. Implement the solution using a low-cost GPU-compatible Amazon EC2 instance and use the AWS Instance Scheduler to schedule the task
C. Implement the solution using AWS Deep Learning Containers, run the workload using AWS Fargate running on Spot Instances, and then schedule the task using the built-in task scheduler
D. Implement the solution using AWS Deep Learning Containers and run the container as a job using AWS Batch on a GPU-compatible Spot Instance
Answer: C (LEAVE A REPLY)

NEW QUESTION: 86
A Machine Learning Specialist wants to bring a custom algorithm to Amazon SageMaker. The Specialist implements the algorithm in a Docker container supported by Amazon SageMaker.

How should the Specialist package the Docker container so that Amazon SageMaker can launch the training correctly?

A. Use CMD configin the Dockerfile to add the training program as a CMD of the image

B. Copy the training program to directory /opt/ml/train

C. Configure the training program as an ENTRYPOINTnamed train

D. Modify the bash_profile file in the container and add a bashcommand to start the training program

Answer: (SHOW ANSWER)

NEW QUESTION: 87

A manufacturing company asks its Machine Learning Specialist to develop a model that classifies defective parts into one of eight defect types. The company has provided roughly 100000 images per defect type for training During the injial training of the image classification model the Specialist notices that the validation accuracy is 80%, while the training accuracy is 90% It is known that human-level performance for this type of image classification is around 90% What should the Specialist consider to fix this issue1?

A. Making the network larger

B. Using some form of regularization

C. A longer training time

D. Using a different optimizer

Answer: B (LEAVE A REPLY)

NEW QUESTION: 88

This graph shows the training and validation loss against the epochs for a neural network The network being trained is as follows

* Two dense layers one output neuron
* 100 neurons in each layer
* 100 epochs
* Random initialization of weights

Which technique can be used to improve model performance in terms of accuracy in the validation set?

A. Adding another layer with the 100 neurons

B. Random initialization of weights with appropriate seed

C. Early stopping

D. Increasing the number of epochs

Answer: A (LEAVE A REPLY)

NEW QUESTION: 89

A company has collected customer comments on its products, rating them as safe or unsafe, using decision trees. The training dataset has the following features: id, date, full review, full review summary, and a binary safe/unsafe tag. During training, any data sample with missing features was dropped. In a few instances, the test set was found to be missing the full review text field.

For this use case, which is the most effective course of action to address test data samples with missing features?

A. Drop the test samples with missing full review text fields, and then run through the test set.

B. Copy the summary text fields and use them to fill in the missing full review text fields, and then run through the test set.

C. Use an algorithm that handles missing data better than decision trees.

D. Generate synthetic data to fill in the fields that are missing data, and then run through the test set.

Answer: B (LEAVE A REPLY)

In this case, a full review summary usually contains the most descriptive phrases of the entire review and is a valid stand-in for the missing full review text field.

NEW QUESTION: 90
A Machine Learning Specialist needs to create a data repository to hold a large amount of time-based training data for a new model. In the source system, new files are added every hour Throughout a single 24-hour period, the volume of hourly updates will change significantly. The Specialist always wants to train on the last 24 hours of the data Which type of data repository is the MOST cost-effective solution?
A. An Amazon S3 data lake with hourly object prefixes
B. An Amazon EMR cluster with hourly hive partitions on Amazon EBS volumes
C. An Amazon RDS database with hourly table partitions
D. An Amazon EBS-backed Amazon EC2 instance with hourly directories
Answer: A (LEAVE A REPLY)

NEW QUESTION: 91
When submitting Amazon SageMaker training jobs using one of the built-in algorithms, which common parameters MUST be specified? (Select THREE.)
A. The output path specifying where on an Amazon S3 bucket the trained model will persist.
B. Hyperparameters in a JSON array as documented for the algorithm used.
C. The 1AM role that Amazon SageMaker can assume to perform tasks on behalf of the users.
D. The validation channel identifying the location of validation data on an Amazon S3 bucket.
E. The training channel identifying the location of training data on an Amazon S3 bucket.
F. The Amazon EC2 instance class specifying whether training will be run using CPU or GPU.
Answer: A,C,F (LEAVE A REPLY)

NEW QUESTION: 92
Machine Learning Specialist is working with a media company to perform classification on popular articles from the company's website. The company is using random forests to classify how popular an article will be before it is published. A sample of the data being used is below.

Article_Title	Author	Top_Keywords	Day_Of_Week	URL_of_Article	Page_Views
Building a Big Data Platform	Jane Doe	Big Data, Spark, Hadoop	Tuesday	http://examplecorp.com/data_platform.html	1300456
Getting Started with Deep Learning	John Doe	Deep Learning, Machine Learning, Spark	Tuesday	http://examplecorp.com/started_deep_learning.html	1230661
MXNet ML Guide	Jane Doe	Machine Learning, MXNet, Logistic Regression	Thursday	http://examplecorp.com/mxnet_guide.html	937291
Intro to NoSQL Databases	Mary Major	NoSQL, Operations, Database	Monday	http://examplecorp.com/nosql_intro_guide.html	407812

Given the dataset, the Specialist wants to convert the Day_Of_Week column to binary values.
What technique should be used to convert this column to binary values?

A. Normalization transformation

B. One-hot encoding

C. Tokenization

D. Binarization

Answer: B (LEAVE A REPLY)

NEW QUESTION: 93

A Machine Learning Specialist is assigned to a Fraud Detection team and must tune an XGBoost model, which is working appropriately for test dat a. However, with unknown data, it is not working as expected. The existing parameters are provided as follows.

```
param = {
      'eta': 0.05,  # the training step for each iteration
      'silent': 1,  # logging mode - quiet
      'n_estimators':2000,
      'max_depth':30,
      'min_child_weight': 3,
      'gamma': 0,
      'subsample': 0.8,
      'objective': 'multi:softprob',  # error evaluation for multiclass training
      'num_class': 201}  # the number of classes that exist in this dataset
num_round = 60  # the number of training iterations
```

Which parameter tuning guidelines should the Specialist follow to avoid overfitting?

A. Update the objective to binary:logistic.

B. Lower the min_child_weight parameter value.

C. Lower the max_depth parameter value.

D. Increase the max_depth parameter value.

Answer: (SHOW ANSWER)

NEW QUESTION: 94
A Machine Learning Specialist is planning to create a long-running Amazon EMR cluster. The EMR cluster will have 1 master node, 10 core nodes, and 20 task nodes. To save on costs, the Specialist will use Spot Instances in the EMR cluster.
Which nodes should the Specialist launch on Spot Instances?
A. Any of the core nodes
B. Both core and task nodes
C. Any of the task nodes
D. Master node
Answer: (SHOW ANSWER)

NEW QUESTION: 95
A Machine Learning Specialist is required to build a supervised image-recognition model to identify a cat. The ML Specialist performs some tests and records the following results for a neural network-based image classifier:
Total number of images available = 1,000 Test set images = 100 (constant test set) The ML Specialist notices that, in over 75% of the misclassified images, the cats were held upside down by their owners.
Which techniques can be used by the ML Specialist to improve this specific test error?
A. Increase the number of layers for the neural network.
B. Increase the dropout rate for the second-to-last layer.
C. Increase the number of epochs for model training.
D. Increase the training data by adding variation in rotation for training images.
Answer: (SHOW ANSWER)

NEW QUESTION: 96
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.

NEW QUESTION: 97

A Machine Learning Specialist needs to move and transform data in preparation for training Some of the data needs to be processed in near-real time and other data can be moved hourly There are existing Amazon EMR MapReduce jobs to clean and feature engineering to perform on the data Which of the following services can feed data to the MapReduce jobs? (Select TWO)

A. AWSDMS

B. Amazon Kinesis

C. AWS Data Pipeline

D. Amazon Athena

E. Amazon ES

Answer: B,C (LEAVE A REPLY)

https://aws.amazon.com/jp/emr/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc

NEW QUESTION: 98

An insurance company is developing a new device for vehicles that uses a camera to observe drivers' behavior and alert them when they appear distracted The company created approximately 10,000 training images in a controlled environment that a Machine Learning Specialist will use to train and evaluate machine learning models During the model evaluation the Specialist notices that the training error rate diminishes faster as the number of epochs increases and the model is not accurately inferring on the unseen test images Which of the following should be used to resolve this issue? (Select TWO)

A. Perform data augmentation on the training data

B. Add vanishing gradient to the model

C. Make the neural network architecture complex.

D. Add L2 regularization to the model

E. Use gradient checking in the model

Answer: (SHOW ANSWER)

NEW QUESTION: 99

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

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

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

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

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

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

Answer: A (LEAVE A REPLY)

https://aws.amazon.com/big-data/real-time-analytics-featured-partners/

NEW QUESTION: 100

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: B (LEAVE A REPLY)
https://docs.aws.amazon.com/sagemaker/latest/dg/sagemaker-dg.pdf
If you plan to use GPU devices, make sure that your containers are nvidia-docker compatible. Only the CUDA toolkit should be included on containers. Don't bundle NVIDIA drivers with the image.
For more information about nvidia-docker, see NVIDIA/nvidia-docker.

NEW QUESTION: 101
A retail company wants to update its customer support system. The company wants to implement automatic routing of customer claims to different queues to prioritize the claims by category. Currently, an operator manually performs the category assignment and routing. After the operator classifies and routes the claim, the company stores the claim's record in a central database. The claim's record includes the claim's category.
The company has no data science team or experience in the field of machine learning (ML). The company's small development team needs a solution that requires no ML expertise.
Which solution meets these requirements?
A. Export the database to a .csv file with two columns: claim_label and claim_text. Use Amazon Comprehend custom classification and the .csv file to train the custom classifier. Develop a service in the application to use the Amazon Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.
B. Export the database to a .csv file with one column: claim_text. Use the Amazon SageMaker Latent Dirichlet Allocation (LDA) algorithm and the .csv file to train a model. Use the LDA algorithm to detect labels automatically. Use SageMaker to deploy the model to an inference endpoint. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
C. Export the database to a .csv file with two columns: claim_label and claim_text. Use the Amazon SageMaker Object2Vec algorithm and the .csv file to train a model. Use SageMaker to deploy the model to an inference endpoint. Develop a service in the application to use the inference endpoint to process incoming claims, predict the labels, and route the claims to the appropriate queue.
D. Use Amazon Textract to process the database and automatically detect two columns: claim_label and claim_text. Use Amazon Comprehend custom classification and the extracted information to train the custom classifier. Develop a service in the application to use the Amazon

Comprehend API to process incoming claims, predict the labels, and route the claims to the appropriate queue.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 102
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. Increase the randomization of training data in the mini-batches used in training
B. Allocate a higher proportion of the overall data to the training dataset
C. Reduce the number of layers and units (or neurons) from the deep learning network
D. Apply L1 or L2 regularization and dropouts to the training
Answer: C (LEAVE A REPLY)

NEW QUESTION: 103
A company has set up and deployed its machine learning (ML) model into production with an endpoint using Amazon SageMaker hosting services. The ML team has configured automatic scaling for its SageMaker instances to support workload changes. During testing, the team notices that additional instances are being launched before the new instances are ready. This behavior needs to change as soon as possible.
How can the ML team solve this issue?
A. Set up Amazon API Gateway and AWS Lambda to trigger the SageMaker inference endpoint.
B. Decrease the cooldown period for the scale-in activity. Increase the configured maximum capacity of instances.
C. Increase the cooldown period for the scale-out activity.
D. Replace the current endpoint with a multi-model endpoint using SageMaker.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 104
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. Multi-category classification
B. Streaming classification
C. Regression classification

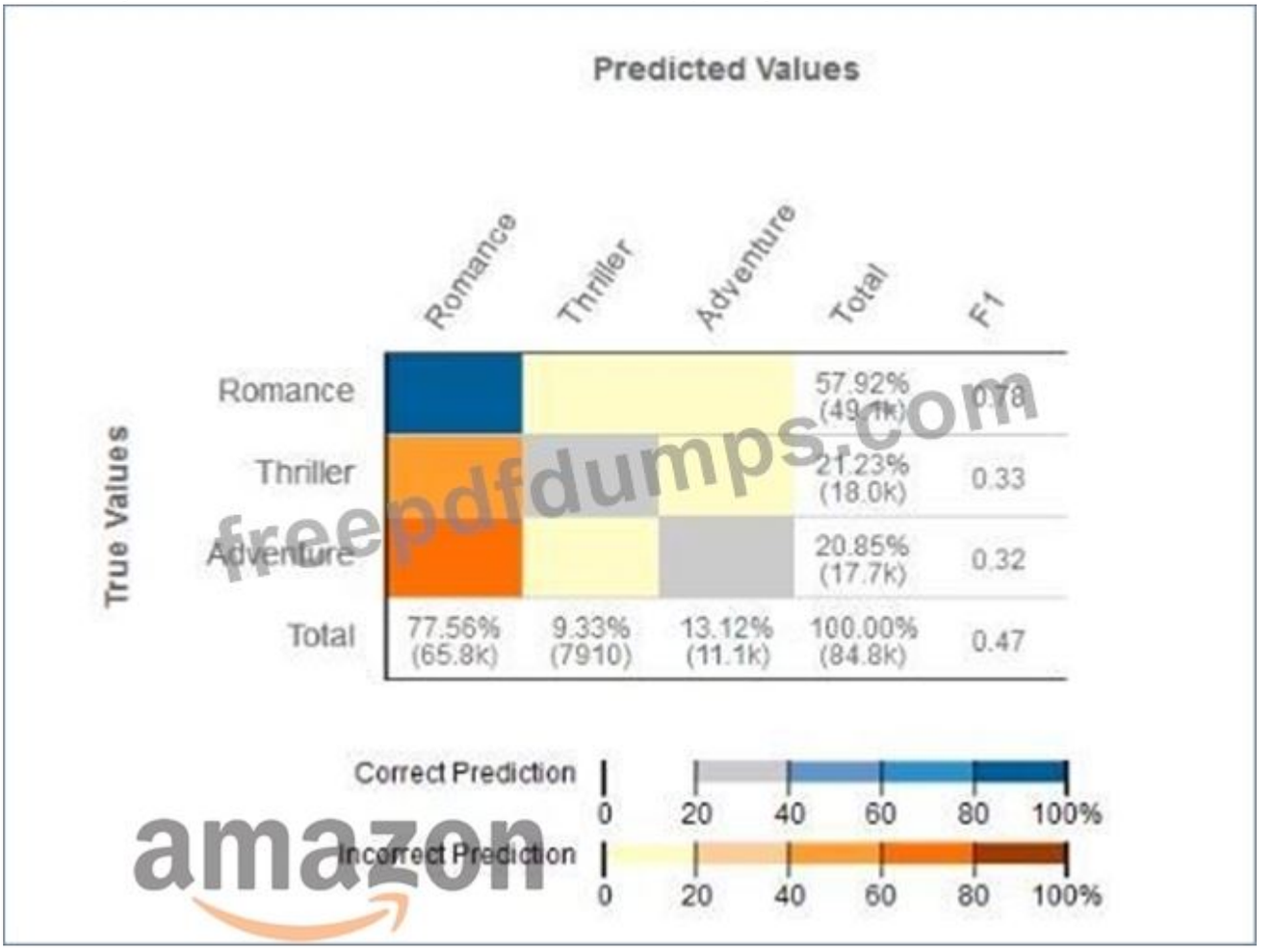
D. Binary classification
Answer: A (LEAVE A REPLY)

NEW QUESTION: 105
A large consumer goods manufacturer has the following products on sale
* 34 different toothpaste variants
* 48 different toothbrush variants
* 43 different mouthwash variants
The entire sales history of all these products is available in Amazon S3 Currently, the company is using custom-built autoregressive integrated moving average (ARIMA) models to forecast demand for these products The company wants to predict the demand for a new product that will soon be launched Which solution should a Machine Learning Specialist apply?
A. Train a custom ARIMA model to forecast demand for the new product.
B. Train an Amazon SageMaker DeepAR algorithm to forecast demand for the new product
C. Train a custom XGBoost model to forecast demand for the new product
D. Train an Amazon SageMaker k-means clustering algorithm to forecast demand for the new product.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 106
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 on an M5 instance. Attach Amazon Elastic Inference to the instance.
B. Deploy the model onto an Amazon Elastic Container Service (Amazon ECS) cluster using a P3 instance.
C. Redeploy the model on a P3dn instance.
D. Redeploy the model as a batch transform job on an M5 instance.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 107

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



A. The true class frequency for Romance is 57.92% and the predicted class frequency for Adventure is
1312%

B. The true class frequency for Romance is 77.56% * 0.78 and the predicted class frequency for Adventure is 20 85% ' 0.32

C. The true class frequency for Romance is 77.56% and the predicted class frequency for Adventure is 20
85%

D. The true class frequency for Romance is 0 78 and the predicted class frequency for Adventure is (0 47 -
0.32).

Answer: C (LEAVE A REPLY)

NEW QUESTION: 108

A Machine Learning Specialist is building a logistic regression model that will predict whether or not a person will order a pizza. The Specialist is trying to build the optimal model with an ideal classification threshold.
What model evaluation technique should the Specialist use to understand how different classification thresholds will impact the model's performance?
A. Receiver operating characteristic (ROC) curve
B. Misclassification rate
C. Root Mean Square Error (RMSE)
D. L1 norm
Answer: A (LEAVE A REPLY)
https://docs.aws.amazon.com/machine-learning/latest/dg/binary-model-insights.html

NEW QUESTION: 109
A Data Scientist is working on an application that performs sentiment analysis. The validation accuracy is poor and the Data Scientist thinks that the cause may be a rich vocabulary and a low average frequency of words in the dataset Which tool should be used to improve the validation accuracy?
A. Scikit-learn term frequency-inverse document frequency (TF-IDF) vectorizers
B. Amazon SageMaker BlazingText allow mode
C. Natural Language Toolkit (NLTK) stemming and stop word removal
D. Amazon Comprehend syntax analysts and entity detection
Answer: (SHOW ANSWER)

NEW QUESTION: 110
A Machine Learning Specialist is using an Amazon SageMaker notebook instance in a private subnet of a corporate VPC. The ML Specialist has important data stored on the Amazon SageMaker notebook instance's Amazon EBS volume, and needs to take a snapshot of that EBS volume. However the ML Specialist cannot find the Amazon SageMaker notebook instance's EBS volume or Amazon EC2 instance within the VPC.
Why is the ML Specialist not seeing the instance visible in the VPC?
A. Amazon SageMaker notebook instances are based on AWS ECS instances running within AWS service accounts.
B. Amazon SageMaker notebook instances are based on the Amazon ECS service within customer accounts.
C. Amazon SageMaker notebook instances are based on the EC2 instances within the customer account but they run outside of VPCs.
D. Amazon SageMaker notebook instances are based on EC2 instances running within AWS service accounts.
Answer: (SHOW ANSWER)

NEW QUESTION: 111

A data scientist is developing a pipeline to ingest streaming web traffic dat a. The data scientist needs to implement a process to identify unusual web traffic patterns as part of the pipeline. The patterns will be used downstream for alerting and incident response. The data scientist has access to unlabeled historic data to use, if needed.
The solution needs to do the following:
Calculate an anomaly score for each web traffic entry.
Adapt unusual event identification to changing web patterns over time.
Which approach should the data scientist implement to meet these requirements?
A. Collect the streaming data using Amazon Kinesis Data Firehose. Map the delivery stream as an input source for Amazon Kinesis Data Analytics. Write a SQL query to run in real time against the streaming data with the k-Nearest Neighbors (kNN) SQL extension to calculate anomaly scores for each record using a tumbling window.
B. Collect the streaming data using Amazon Kinesis Data Firehose. Map the delivery stream as an input source for Amazon Kinesis Data Analytics. Write a SQL query to run in real time against the streaming data with the Amazon Random Cut Forest (RCF) SQL extension to calculate anomaly scores for each record using a sliding window.
C. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker Random Cut Forest (RCF) built-in model. Use an Amazon Kinesis Data Stream to process the incoming web traffic data. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the RCF model to calculate the anomaly score for each record.
D. Use historic web traffic data to train an anomaly detection model using the Amazon SageMaker built-in XGBoost model. Use an Amazon Kinesis Data Stream to process the incoming web traffic data. Attach a preprocessing AWS Lambda function to perform data enrichment by calling the XGBoost model to calculate the anomaly score for each record.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 112
A retail company uses a machine learning (ML) model for daily sales forecasting. The company's brand manager reports that the model has provided inaccurate results for the past 3 weeks.
At the end of each day, an AWS Glue job consolidates the input data that is used for the forecasting with the actual daily sales data and the predictions of the model. The AWS Glue job stores the data in Amazon S3. The company's ML team is using an Amazon SageMaker Studio notebook to gain an understanding about the source of the model's inaccuracies.
What should the ML team do on the SageMaker Studio notebook to visualize the model's degradation MOST accurately?
A. Create a histogram of the daily sales over the last 3 weeks. In addition, create a histogram of the daily sales from before that period.
B. Create a histogram of the model errors over the last 3 weeks. In addition, create a histogram of the model errors from before that period.
C. Create a line chart with the weekly mean absolute error (MAE) of the model.

D. Create a scatter plot of daily sales versus model error for the last 3 weeks. In addition, create a scatter plot of daily sales versus model error from before that period.
Answer: (SHOW ANSWER)

NEW QUESTION: 113
During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?
A. Dataset shuffling is disabled
B. The learning rate is very high
C. The batch size is too big
D. The class distribution in the dataset is imbalanced
Answer: (SHOW ANSWER)

NEW QUESTION: 114
A Machine Learning Specialist has created a deep learning neural network model that performs well on the training data but performs poorly on the test data.
Which of the following methods should the Specialist consider using to correct this? (Select THREE.)
A. Increase feature combinations.
B. Decrease dropout.
C. Increase regularization.
D. Increase dropout.
E. Decrease feature combinations.
F. Decrease regularization.
Answer: A,D,F (LEAVE A REPLY)

NEW QUESTION: 115
A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes Which function will produce the desired output?
A. Rectified linear units (ReLU)
B. Dropout
C. Smooth L1 loss
D. Softmax
Answer: D (LEAVE A REPLY)

NEW QUESTION: 116

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: 117
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. Organize the Docker container's file structure to execute on GPU instances.
B. Bundle the NVIDIA drivers with the Docker image.
C. Set the GPU flag in the Amazon SageMaker CreateTrainingJob request body
D. Build the Docker container to be NVIDIA-Docker compatible.
Answer: A (LEAVE A REPLY)

NEW QUESTION: 118
A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.
How should the Data Science team configure the notebook instance placement to meet these requirements?
A. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
B. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Use IAM policies to grant access to Amazon S3 and Amazon SageMaker.
C. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.

D. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker.

Answer: C (LEAVE A REPLY)

We must use the VPC endpoint (either Gateway Endpoint or Interface Endpoint)to comply with this requirement "Data communication traffic must stay within the AWS network".
https://docs.aws.amazon.com/sagemaker/latest/dg/notebook-interface-endpoint.html

NEW QUESTION: 119

A Data Scientist uses logistic regression to build a fraud detection model. While the model accuracy is 99%, 90% of the fraud cases are not detected by the model.
What action will definitively help the model detect more than 10% of fraud cases?

A. Using undersampling to balance the dataset

B. Decreasing the class probability threshold

C. Using regularization to reduce overfitting

D. Using oversampling to balance the dataset

Answer: B (LEAVE A REPLY)

Decreasing the class probability threshold makes the model more sensitive and, therefore, marks more cases as the positive class, which is fraud in this case. This will increase the likelihood of fraud detection. However, it comes at the price of lowering precision.

NEW QUESTION: 120

A company is using Amazon Polly to translate plaintext documents to speech for automated company announcements However company acronyms are being mispronounced in the current documents How should a Machine Learning Specialist address this issue for future documents'?

A. Use Amazon Lex to preprocess the text files for pronunciation

B. Output speech marks to guide in pronunciation

C. Convert current documents to SSML with pronunciation tags

D. Create an appropriate pronunciation lexicon.

Answer: C (LEAVE A REPLY)

NEW QUESTION: 121

This graph shows the training and validation loss against the epochs for a neural network The network being trained is as follows

* Two dense layers one output neuron
* 100 neurons in each layer
* 100 epochs
* Random initialization of weights

Which technique can be used to improve model performance in terms of accuracy in the validation set?

A. Early stopping

B. Adding another layer with the 100 neurons

C. Increasing the number of epochs

D. Random initialization of weights with appropriate seed

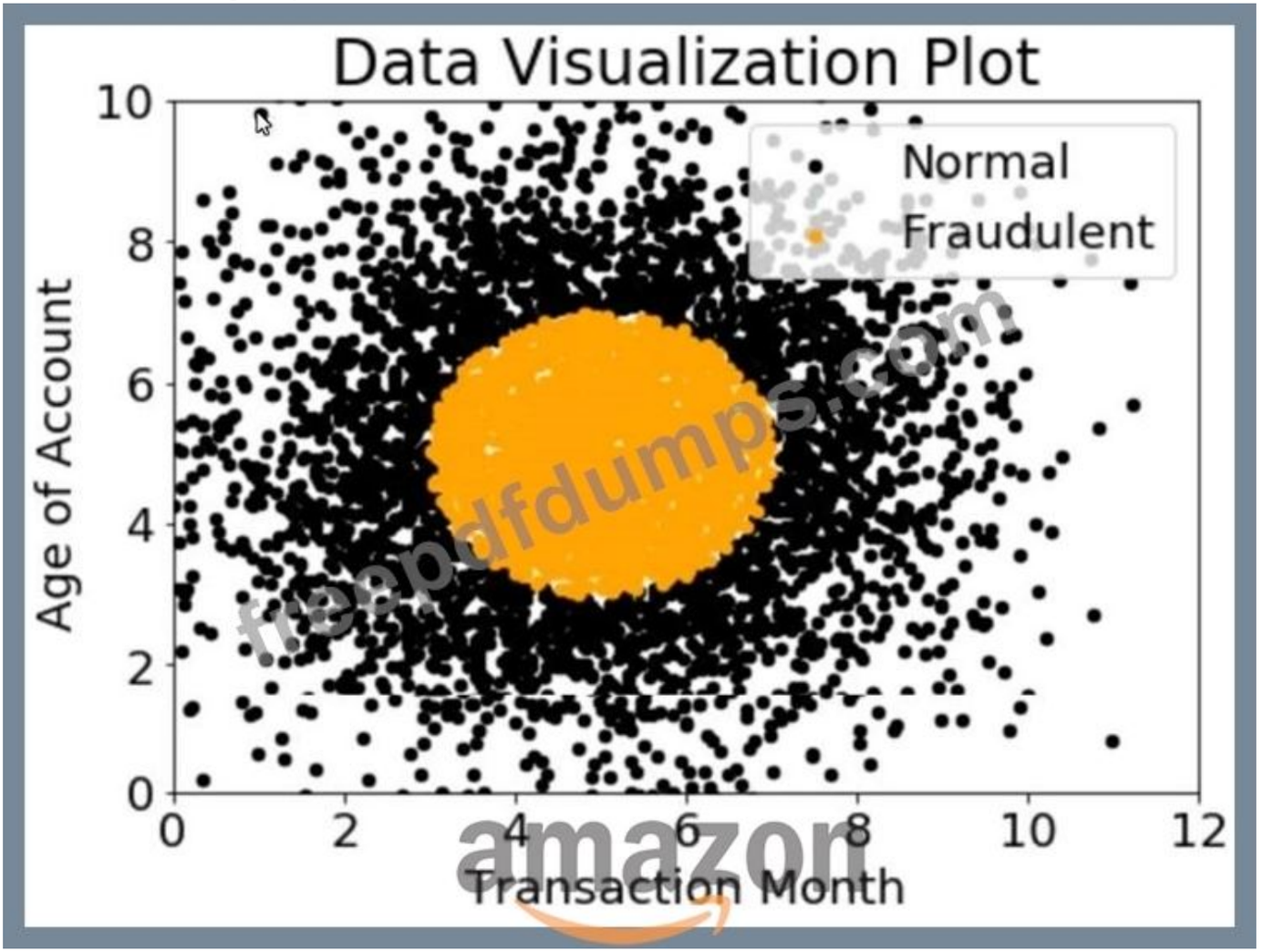
Answer: B (LEAVE A REPLY)

NEW QUESTION: 122

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

NEW QUESTION: 123
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information, which model would have the HIGHEST recall with respect to the fraudulent class?

A. Decision tree

B. Linear support vector machine (SVM)

C. Single Perceptron with sigmoidal activation function

D. Naive Bayesian classifier

Answer: D (LEAVE A REPLY)

NEW QUESTION: 124

Which of the following metrics should a Machine Learning Specialist generally use to compare/evaluate machine learning classification models against each other?

A. Recall

B. Misclassification rate

C. Mean absolute percentage error (MAPE)

D. Area Under the ROC Curve (AUC)

Answer: A (LEAVE A REPLY)

Explanation/Reference: https://docs.aws.amazon.com/machine-learning/latest/dg/multiclass-model-insights.html

NEW QUESTION: 125

The displayed graph is from a foresting 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 the seasonality well, but not the trend.

B. The model does not predict the trend or the seasonality well.

C. The model predicts both the trend and the seasonality well.

D. The model predicts the trend well, but not the seasonality.

Answer: B (LEAVE A REPLY)

NEW QUESTION: 126
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 regression forest where the number of trees is set equal to the number of product categories
B. An XGBoost model where the objective parameter is set to multi: softmax
C. A DeepAR forecasting model based on a recurrent neural network (RNN)
D. A deep convolutional neural network (CNN) with a softmax activation function for the last layer
Answer: (SHOW ANSWER)

NEW QUESTION: 127
An agency collects census information within a country to determine healthcare and social program needs by province and city. The census form collects responses for approximately 500 questions from each citizen Which combination of algorithms would provide the appropriate insights? (Select TWO)
A. The factorization machines (FM) algorithm
B. The Latent Dirichlet Allocation (LDA) algorithm
C. The principal component analysis (PCA) algorithm
D. The k-means algorithm g The Random Cut Forest (RCF) algorithm
Answer: (SHOW ANSWER)

NEW QUESTION: 128
A Machine Learning Specialist is working for a credit card processing company and receives an unbalanced dataset containing credit card transactions. It contains 99,000 valid transactions and 1,000 fraudulent transactions The Specialist is asked to score a model that was run against the dataset The Specialist has been advised that identifying valid transactions is equally as important as identifying fraudulent transactions What metric is BEST suited to score the model?
A. Area Under the ROC Curve (AUC)
B. Root Mean Square Error (RMSE)
C. Precision
D. Recall
Answer: (SHOW ANSWER)

NEW QUESTION: 129
A Machine Learning Specialist is building a convolutional neural network (CNN) that will classify 10 types of animals. The Specialist has built a series of layers in a neural network that will take an

input image of an animal, pass it through a series of convolutional and pooling layers, and then finally pass it through a dense and fully connected layer with 10 nodes The Specialist would like to get an output from the neural network that is a probability distribution of how likely it is that the input image belongs to each of the 10 classes Which function will produce the desired output?
A. Softmax
B. Rectified linear units (ReLU)
C. Smooth L1 loss
D. Dropout
Answer: B (LEAVE A REPLY)

NEW QUESTION: 130
A Machine Learning Specialist is developing a custom video recommendation model for an application. The dataset used to train this model is very large with millions of data points and is hosted in an Amazon S3 bucket. The Specialist wants to avoid loading all of this data onto an Amazon SageMaker notebook instance because it would take hours to move and will exceed the attached 5 GB Amazon EBS volume on the notebook instance.
Which approach allows the Specialist to use all the data to train the model?
A. Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to train the full dataset.
B. Use AWS Glue to train a model using a small subset of the data to confirm that the data will be compatible with Amazon SageMaker. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
C. Load a smaller subset of the data into the SageMaker notebook and train locally. Confirm that the training code is executing and the model parameters seem reasonable. Initiate a SageMaker training job using the full dataset from the S3 bucket using Pipe input mode.
D. Launch an Amazon EC2 instance with an AWS Deep Learning AMI and attach the S3 bucket to the instance. Train on a small amount of the data to verify the training code and hyperparameters.
Go back to Amazon SageMaker and train using the full dataset
Answer: C (LEAVE A REPLY)

NEW QUESTION: 131
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. Last observation carried forward
B. Multiple imputation
C. Mean substitution
D. Listwise deletion

Answer: (SHOW ANSWER)

NEW QUESTION: 132
A Data Scientist is working on an application that performs sentiment analysis. The validation accuracy is poor, and the Data Scientist thinks that the cause may be a rich vocabulary and a low average frequency of words in the dataset.
Which tool should be used to improve the validation accuracy?
A. Amazon Comprehend syntax analysis and entity detection
B. Amazon SageMaker BlazingText cbowmode
C. Natural Language Toolkit (NLTK) stemming and stop word removal
D. Scikit-leam term frequency-inverse document frequency (TF-IDF) vectorizer
Answer: D (LEAVE A REPLY)
Explanation/Reference: https://monkeylearn.com/sentiment-analysis/

NEW QUESTION: 133
A data scientist uses an Amazon SageMaker notebook instance to conduct data exploration and analysis. This requires certain Python packages that are not natively available on Amazon SageMaker to be installed on the notebook instance.
How can a machine learning specialist ensure that required packages are automatically available on the notebook instance for the data scientist to use?
A. Install AWS Systems Manager Agent on the underlying Amazon EC2 instance and use Systems Manager Automation to execute the package installation commands.
B. Create a Jupyter notebook file (.ipynb) with cells containing the package installation commands to execute and place the file under the /etc/init directory of each Amazon SageMaker notebook instance.
C. Use the conda package manager from within the Jupyter notebook console to apply the necessary conda packages to the default kernel of the notebook.
D. Create an Amazon SageMaker lifecycle configuration with package installation commands and assign the lifecycle configuration to the notebook instance.
Answer: (SHOW ANSWER)
Explanation/Reference: https://towardsdatascience.com/automating-aws-sagemaker-notebooks-2dec62bc2c84

NEW QUESTION: 134
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. Decrease the XGBoost max_depth parameter because the model is currently overfitting the data.

B. Increase the XGBoost max_depth parameter because the model is currently underfitting the data.

C. Change the XGBoost eval_metric parameter to optimize based on AUC instead of error.

D. Change the XGBoost eval_metric parameter to optimize based on rmse instead of error.

E. Increase the XGBoost scale_pos_weight parameter to adjust the balance of positive and negative weights.

Answer: C,E (LEAVE A REPLY)

NEW QUESTION: 135

A company is running a machine learning prediction service that generates 100 TB of predictions every day. A Machine Learning Specialist must generate a visualization of the daily precision-recall curve from the predictions, and forward a read-only version to the Business team.

Which solution requires the LEAST coding effort?

A. Run a daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3. Visualize the arrays in Amazon QuickSight, and publish them in a dashboard shared with the Business team.

B. Generate daily precision-recall data in Amazon ES, and publish the results in a dashboard shared with the Business team.

C. Run daily Amazon EMR workflow to generate precision-recall data, and save the results in Amazon S3.

Give the Business team read-only access to S3.

D. Generate daily precision-recall data in Amazon QuickSight, and publish the results in a dashboard shared with the Business team.

Answer: A (LEAVE A REPLY)

NEW QUESTION: 136

A data scientist is working on a public sector project for an urban traffic system. While studying the traffic patterns, it is clear to the data scientist that the traffic behavior at each light is correlated, subject to a small stochastic error term. The data scientist must model the traffic behavior to analyze the traffic patterns and reduce congestion.

How will the data scientist MOST effectively model the problem?

A. The data scientist should obtain the optimal equilibrium policy by formulating this problem as a single-agent reinforcement learning problem.
B. The data scientist should obtain a correlated equilibrium policy by formulating this problem as a multi-agent reinforcement learning problem.
C. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using unlabeled simulated data representing the new traffic patterns in the city and applying an unsupervised learning approach.
D. Rather than finding an equilibrium policy, the data scientist should obtain accurate predictors of traffic flow by using historical data through a supervised learning approach.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 137
A Data Engineer needs to build a model using a dataset containing customer credit card information.
How can the Data Engineer ensure the data remains encrypted and the credit card information is secure?
A. Use AWS KMS to encrypt the data on Amazon S3
B. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VPC. Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.
C. Use a custom encryption algorithm to encrypt the data and store the data on an Amazon SageMaker instance in a VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.
D. Use an IAM policy to encrypt the data on the Amazon S3 bucket and Amazon Kinesis to automatically discard credit card numbers and insert fake credit card numbers.
Answer: (SHOW ANSWER)

NEW QUESTION: 138

A company is observing low accuracy while training on the default built-in image classification algorithm in Amazon SageMaker. The Data Science team wants to use an Inception neural network architecture instead of a ResNet architecture.
Which of the following will accomplish this? (Choose two.)
A. Download and apt-get installthe inception network code into an Amazon EC2 instance and use this instance as a Jupyter notebook in Amazon SageMaker.
B. Bundle a Docker container with TensorFlow Estimator loaded with an Inception network and use this for model training.
C. Customize the built-in image classification algorithm to use Inception and use this for model training.
D. Create a support case with the SageMaker team to change the default image classification algorithm to Inception.
E. Use custom code in Amazon SageMaker with TensorFlow Estimator to load the model with an Inception network, and use this for model training.
Answer: C,E (LEAVE A REPLY)

NEW QUESTION: 139
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? (Select 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 evaljnetric parameter to optimize based on AUC instead of error.
Answer: C,E (LEAVE A REPLY)

NEW QUESTION: 140

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 needs to be completely re-engineered because it is unable to handle 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's hyperparameters should be periodically updated to prevent drift.
D. The model should be periodically retrained using the original training data plus new data as product inventory changes.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 141
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 lo 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. 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
B. 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
C. Insert an Amazon Kinesis Data Analytics stream downstream of the Kinesis Data Firehouse stream that transforms raw record attributes into simple transformed values using SQL.
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: A (LEAVE A REPLY)

NEW QUESTION: 142
A data engineer at a bank is evaluating a new tabular dataset that includes customer dat a. The data engineer will use the customer data to create a new model to predict customer behavior. After creating a correlation matrix for the variables, the data engineer notices that many of the 100 features are highly correlated with each other.
Which steps should the data engineer take to address this issue? (Choose two.)
A. Use a linear-based algorithm to train the model.
B. Apply principal component analysis (PCA).

C. Remove a portion of highly correlated features from the dataset.
D. Apply min-max feature scaling to the dataset.
E. Apply one-hot encoding category-based variables.
Answer: B,D (LEAVE A REPLY)
Reference:
https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html

NEW QUESTION: 143
A Machine Learning Specialist is using Apache Spark for pre-processing training data As part of the Spark pipeline, the Specialist wants to use Amazon SageMaker for training a model and hosting it Which of the following would the Specialist do to integrate the Spark application with SageMaker? (Select THREE)
A. Compress the training data into a ZIP file and upload it to a pre-defined Amazon S3 bucket.
B. Download the AWS SDK for the Spark environment
C. Install the SageMaker Spark library in the Spark environment.
D. Convert the DataFrame object to a CSV file, and use the CSV file as input for obtaining inferences from SageMaker.
E. Use the sageMakerModel. transform method to get inferences from the model hosted in SageMaker
F. Use the appropriate estimator from the SageMaker Spark Library to train a model.
Answer: A,D,E (LEAVE A REPLY)

NEW QUESTION: 144
A Data Scientist is developing a binary classifier to predict whether a patient has a particular disease on a series of test results. The Data Scientist has data on 400 patients randomly selected from the population. The disease is seen in 3% of the population.
Which cross-validation strategy should the Data Scientist adopt?
A. A stratified k-fold cross-validation strategy with k=5
B. A k-fold cross-validation strategy with k=5
C. An 80/20 stratified split between training and validation
D. A k-fold cross-validation strategy with k=5 and 3 repeats
Answer: A (LEAVE A REPLY)

NEW QUESTION: 145
A Machine Learning Specialist is building a prediction model for a large number of features using linear models, such as linear regression and logistic regression During exploratory data analysis the Specialist observes that many features are highly correlated with each other This may make the model unstable What should be done to reduce the impact of having such a large number of features?
A. Create a new feature space using principal component analysis (PCA)
B. Apply the Pearson correlation coefficient

C. Use matrix multiplication on highly correlated features.
D. Perform one-hot encoding on highly correlated features
Answer: (SHOW ANSWER)

NEW QUESTION: 146
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. An XGBoost model where the objective parameter is set to multi: softmax
C. A deep convolutional neural network (CNN) with a softmax activation function for the last layer
D. A regression forest where the number of trees is set equal to the number of product categories
Answer: C (LEAVE A REPLY)

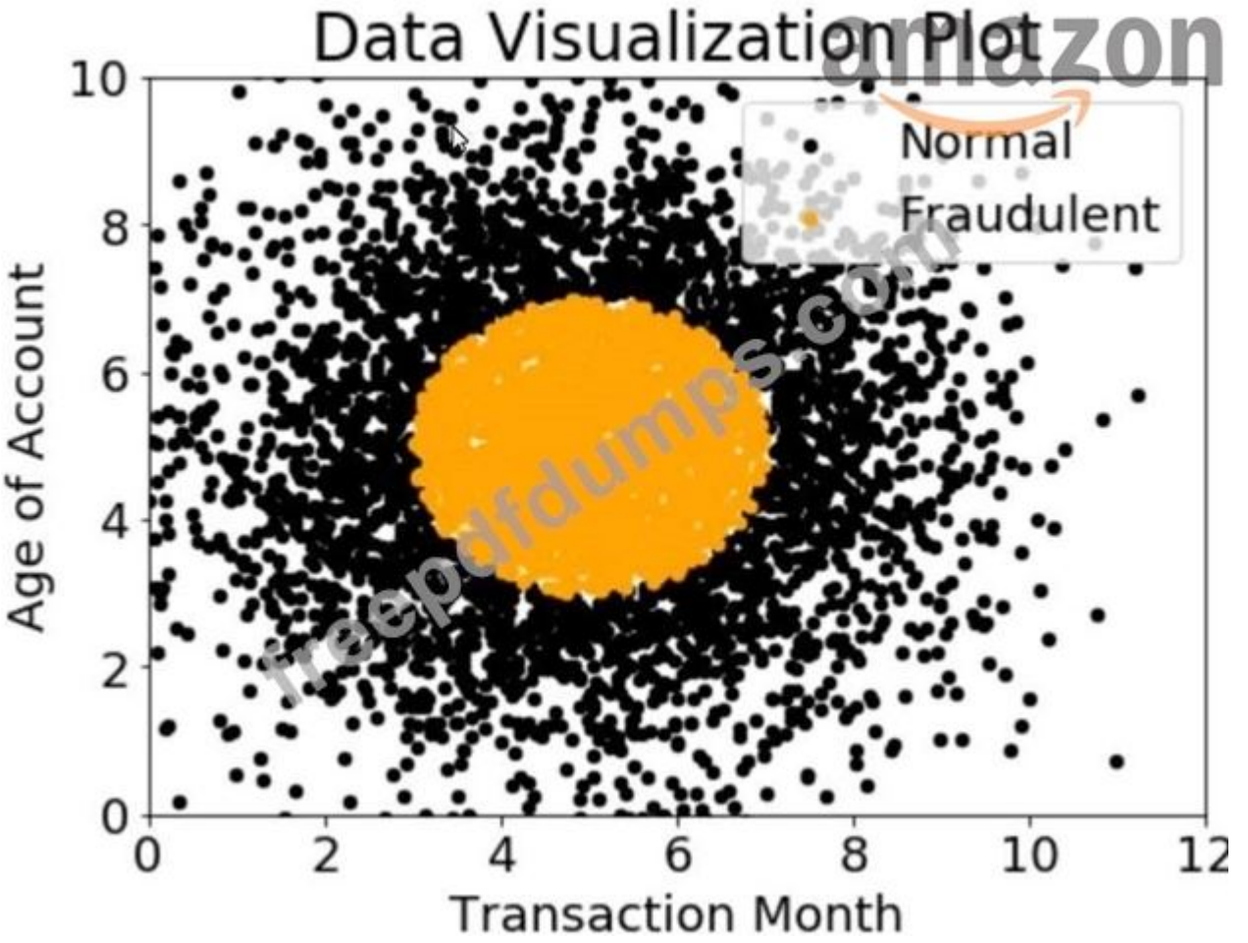
NEW QUESTION: 147
A Data Engineer needs to build a model using a dataset containing customer credit card information.
How can the Data Engineer ensure the data remains encrypted and the credit card information is secure?
A. Use a custom encryption algorithm to encrypt the data and store the data on an Amazon SageMaker instance in a VPC. Use the SageMaker DeepAR algorithm to randomize the credit card numbers.
B. Use an Amazon SageMaker launch configuration to encrypt the data once it is copied to the SageMaker instance in a VPC Use the SageMaker principal component analysis (PCA) algorithm to reduce the length of the credit card numbers.
C. Use AWS KMS to encrypt the data on Amazon S3 and Amazon SageMaker, and redact the credit card numbers from the customer data with AWS Glue.
D. Use an 1AM policy to encrypt the data on the Amazon S3 bucket and Amazon Kinesis to automatically discard credit card numbers and insert fake credit card numbers.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 148
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
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 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 and replace the last fully connected layer
Answer: C (LEAVE A REPLY)

NEW QUESTION: 149
A bank wants to launch a low-rate credit promotion. The bank is located in a town that recently experienced economic hardship. Only some of the bank's customers were affected by the crisis, so the bank's credit team must identify which customers to target with the promotion. However, the credit team wants to make sure that loyal customers' full credit history is considered when the decision is made.
The bank's data science team developed a model that classifies account transactions and understands credit eligibility. The data science team used the XGBoost algorithm to train the model. The team used 7 years of bank transaction historical data for training and hyperparameter tuning over the course of several days.
The accuracy of the model is sufficient, but the credit team is struggling to explain accurately why the model denies credit to some customers. The credit team has almost no skill in data science.
What should the data science team do to address this issue in the MOST operationally efficient manner?
A. Use Amazon SageMaker Studio to rebuild the model. Create a notebook that uses the XGBoost training container to perform model training. Deploy the model at an endpoint. Enable Amazon SageMaker Model Monitor to store inferences. Use the inferences to create Shapley values that help explain model behavior. Create a chart that shows features and SHapley Additive exPlanations (SHAP) values to explain to the credit team how the features affect the model outcomes.
B. Use Amazon SageMaker Studio to rebuild the model. Create a notebook that uses the XGBoost training container to perform model training. Activate Amazon SageMaker Debugger, and configure it to calculate and collect Shapley values. Create a chart that shows features and SHapley Additive exPlanations (SHAP) values to explain to the credit team how the features affect the model outcomes.
C. Create an Amazon SageMaker notebook instance. Use the notebook instance and the XGBoost library to locally retrain the model. Use the plot_importance() method in the Python XGBoost interface to create a feature importance chart. Use that chart to explain to the credit team how the features affect the model outcomes.
D. Use Amazon SageMaker Studio to rebuild the model. Create a notebook that uses the XGBoost training container to perform model training. Deploy the model at an endpoint. Use Amazon SageMaker
Answer: C (LEAVE A REPLY)
Processing to post-analyze the model and create a feature importance explainability chart automatically for the credit team.

NEW QUESTION: 150
A company wants to classify user behavior as either fraudulent or normal. Based on internal research, a Machine Learning Specialist would like to build a binary classifier based on two features: age of account and transaction month. The class distribution for these features is illustrated in the figure provided.



Based on this information, which model would have the HIGHEST accuracy?
A. Single perceptron with tanh activation function
B. Logistic regression
C. Long short-term memory (LSTM) model with scaled exponential linear unit (SELU)
D. Support vector machine (SVM) with non-linear kernel
Answer: D (LEAVE A REPLY)

NEW QUESTION: 151
A Machine Learning Specialist trained a regression model, but the first iteration needs optimizing. The Specialist needs to understand whether the model is more frequently overestimating or underestimating the target.
What option can the Specialist use to determine whether it is overestimating or underestimating the target value?
A. Residual plots
B. Root Mean Square Error (RMSE)

C. Area under the curve
D. Confusion matrix
Answer: B (LEAVE A REPLY)

NEW QUESTION: 152
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 histogram showing whether the most important input feature is Gaussian.
B. A scatter plot with points colored by target variable that uses (-Distributed Stochastic Neighbor Embedding (I-SNE) to visualize the large number of input variables in an easier-to-read dimension.
C. A scatter plot showing the correlation between maximum tree depth and the objective metric.
D. A scatter plot showing (he performance of the objective metric over each training iteration
Answer: (SHOW ANSWER)

NEW QUESTION: 153
A Machine Learning Specialist at a company sensitive to security is preparing a dataset for model training. The dataset is stored in Amazon S3 and contains Personally Identifiable Information (PII).
The dataset:
* Must be accessible from a VPC only.
* Must not traverse the public internet.
How can these requirements be satisfied?
A. Create a VPC endpoint and use security groups to restrict access to the given VPC endpoint and an Amazon EC2 instance

B. Create a VPC endpoint and use Network Access Control Lists (NACLs) to allow traffic between only the given VPC endpoint and an Amazon EC2 instance.
C. Create a VPC endpoint and apply a bucket access policy that restricts access to the given VPC endpoint and the VPC.
D. Create a VPC endpoint and apply a bucket access policy that allows access from the given VPC endpoint and an Amazon EC2 instance.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 154
A power company wants to forecast future energy consumption for its customers in residential properties and commercial business properties. Historical power consumption data for the last 10 years is available. A team of data scientists who performed the initial data analysis and feature selection will include the historical power consumption data and data such as weather, number of individuals on the property, and public holidays.
The data scientists are using Amazon Forecast to generate the forecasts.
Which algorithm in Forecast should the data scientists use to meet these requirements?
A. Exponential Smoothing (ETS)
B. Convolutional Neural Network - Quantile Regression (CNN-QR)
C. Autoregressive Integrated Moving Average (AIRMA)
D. Prophet
Answer: (SHOW ANSWER)

NEW QUESTION: 155
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 trfese requirements?
A. 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.
B. 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.
C. Use an AWS Glue crawler to crawl S3 data, an AWS Lambda function to trigger an AWS Glue ETL job, and an AWS Glue Data catalog to search and discover metadata.
D. Use an AWS Glue crawler to crawl S3 data, an Amazon CloudWatch alarm to trigger an AWS Batch job, and an AWS Glue Data Catalog to search and discover metadata.
Answer: (SHOW ANSWER)

NEW QUESTION: 156

During mini-batch training of a neural network for a classification problem, a Data Scientist notices that training accuracy oscillates What is the MOST likely cause of this issue?
A. The batch size is too big
B. The learning rate is very high
C. The class distribution in the dataset is imbalanced
D. Dataset shuffling is disabled
Answer: B (LEAVE A REPLY)

NEW QUESTION: 157
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. Root Mean Square Error (RMSE)
B. Confusion matrix
C. Residual plots
D. Area under the curve
Answer: D (LEAVE A REPLY)

NEW QUESTION: 158
A retail company is using Amazon Personalize to provide personalized product recommendations for its customers during a marketing campaign. The company sees a significant increase in sales of recommended items to existing customers immediately after deploying a new solution version, but these sales decrease a short time after deployment. Only historical data from before the marketing campaign is available for training.
How should a data scientist adjust the solution?
A. Implement a new solution using the built-in factorization machines (FM) algorithm in Amazon SageMaker.
B. Use the event tracker in Amazon Personalize to include real-time user interactions.
C. Add event type and event value fields to the interactions dataset in Amazon Personalize.
D. Add user metadata and use the HRNN-Metadata recipe in Amazon Personalize.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 159
A library is developing an automatic book-borrowing system that uses Amazon Rekognition. Images of library members' faces are stored in an Amazon S3 bucket. When members borrow books, the Amazon Rekognition CompareFaces API operation compares real faces against the stored faces in Amazon S3.
The library needs to improve security by making sure that images are encrypted at rest. Also, when the images are used with Amazon Rekognition. they need to be encrypted in transit. The

library also must ensure that the images are not used to improve Amazon Rekognition as a service.
How should a machine learning specialist architect the solution to satisfy these requirements?
A. Enable server-side encryption on the S3 bucket. Submit an AWS Support ticket to opt out of allowing images to be used for improving the service, and follow the process provided by AWS Support.
B. Enable client-side encryption on the S3 bucket. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.
C. Switch to using the AWS GovCloud (US) Region for Amazon S3 to store images and for Amazon Rekognition to compare faces. Set up a VPN connection and only call the Amazon Rekognition API operations through the VPN.
D. Switch to using an Amazon Rekognition collection to store the images. Use the IndexFaces and SearchFacesByImage API operations instead of the CompareFaces API operation.
Answer: D (LEAVE A REPLY)

NEW QUESTION: 160
An Machine Learning Specialist discover the following statistics while experimenting on a model.

What can the Specialist from the experiments?
A. The model in Experiment 1 had a high bias error that was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal variance error in Experiment 1
B. The model in Experiment 1 had a high random noise error that was reduced in Expenment 3 by regularization Expenment 2 shows that random noise cannot be reduced by increasing layers and neurons in the model
C. The model In Experiment 1 had a high variance error lhat was reduced in Experiment 3 by regularization Experiment 2 shows that there is minimal bias error in Experiment 1
D. The model in Experiment 1 had a high bias error and a high variance error that were reduced in Experiment 3 by regularization Experiment 2 shows thai high bias cannot be reduced by increasing layers and neurons in the model
Answer: D (LEAVE A REPLY)

NEW QUESTION: 161
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 histogram showing whether the most important input feature is Gaussian.
B. 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.
C. A scatter plot showing the performance of the objective metric over each training iteration.
D. A scatter plot showing the correlation between maximum tree depth and the objective metric.
Answer: B (LEAVE A REPLY)
https://medium.com/all-things-ai/in-depth-parameter-tuning-for-random-forest-d67bb7e920d

NEW QUESTION: 162
A medical imaging company wants to train a computer vision model to detect areas of concern on patients' CT scans. The company has a large collection of unlabeled CT scans that are linked to each patient and stored in an Amazon S3 bucket. The scans must be accessible to authorized users only. A machine learning engineer needs to build a labeling pipeline.
Which set of steps should the engineer take to build the labeling pipeline with the LEAST effort?
A. Create a workforce with AWS Identity and Access Management (IAM). Build a labeling tool on Amazon EC2 Queue images for labeling by using Amazon Simple Queue Service (Amazon SQS). Write the labeling instructions.
B. Create an Amazon Mechanical Turk workforce and manifest file. Create a labeling job by using the built-in image classification task type in Amazon SageMaker Ground Truth. Write the labeling instructions.
C. Create a private workforce and manifest file. Create a labeling job by using the built-in bounding box task type in Amazon SageMaker Ground Truth. Write the labeling instructions.
D. Create a workforce with Amazon Cognito. Build a labeling web application with AWS Amplify. Build a labeling workflow backend using AWS Lambda. Write the labeling instructions.
Answer: C (LEAVE A REPLY)
https://docs.aws.amazon.com/sagemaker/latest/dg/sms-workforce-private.html

NEW QUESTION: 163
A company offers an online shopping service to its customers. The company wants to enhance the site's security by requesting additional information when customers access the site from locations that are different from their normal location. The company wants to update the process to call a machine learning (ML) model to determine when additional information should be requested.
The company has several terabytes of data from its existing ecommerce web servers containing the source IP addresses for each request made to the web server. For authenticated requests, the records also contain the login name of the requesting user.

Which approach should an ML specialist take to implement the new security feature in the web application?

A. Use Amazon SageMaker Ground Truth to label each record as either a successful or failed access attempt. Use Amazon SageMaker to train a binary classification model using the factorization machines (FM) algorithm.

B. Use Amazon SageMaker to train a model using the Object2Vec algorithm. Schedule updates and retraining of the model using new log data nightly.

C. Use Amazon SageMaker Ground Truth to label each record as either a successful or failed access attempt. Use Amazon SageMaker to train a binary classification model using the IP Insights algorithm.

D. Use Amazon SageMaker to train a model using the IP Insights algorithm. Schedule updates and retraining of the model using new log data nightly.

Answer: (SHOW ANSWER)

NEW QUESTION: 164

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. 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.

C. 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.

D. 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.

Answer: (SHOW ANSWER)

Explanation

NEW QUESTION: 165

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 word embedding factors that store edit distance with every other word.

B. Create one-hot word encoding vectors.

C. Produce a set of synonyms for every word using Amazon Mechanical Turk.

D. Download word embedding's pre-trained on a large corpus.
Answer: (SHOW ANSWER)

NEW QUESTION: 166
A Data Science team within a large company uses Amazon SageMaker notebooks to access data stored in Amazon S3 buckets. The IT Security team is concerned that internet-enabled notebook instances create a security vulnerability where malicious code running on the instances could compromise data privacy. The company mandates that all instances stay within a secured VPC with no internet access, and data communication traffic must stay within the AWS network.
How should the Data Science team configure the notebook instance placement to meet these requirements?
A. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has S3 VPC endpoints and Amazon SageMaker VPC endpoints attached to it.
B. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Place the Amazon SageMaker endpoint and S3 buckets within the same VPC.
C. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Use 1AM policies to grant access to Amazon S3 and Amazon SageMaker.
D. Associate the Amazon SageMaker notebook with a private subnet in a VPC. Ensure the VPC has a NAT gateway and an associated security group allowing only outbound connections to Amazon S3 and Amazon SageMaker
Answer: A (LEAVE A REPLY)

NEW QUESTION: 167
A manufacturing company wants to use machine learning (ML) to automate quality control in its facilities. The facilities are in remote locations and have limited internet connectivity. The company has 20 TB of training data that consists of labeled images of defective product parts. The training data is in the corporate on-premises data center.
The company will use this data to train a model for real-time defect detection in new parts as the parts move on a conveyor belt in the facilities. The company needs a solution that minimizes costs for compute infrastructure and that maximizes the scalability of resources for training. The

solution also must facilitate the company's use of an ML model in the low-connectivity environments.
Which solution will meet these requirements?
A. Train the model on premises. Upload the model to an Amazon S3 bucket. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.
B. Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Set up an edge device in the manufacturing facilities with AWS IoT Greengrass. Deploy the model on the edge device.
C. Move the training data to an Amazon S3 bucket. Train and evaluate the model by using Amazon SageMaker. Optimize the model by using SageMaker Neo. Deploy the model on a SageMaker hosting services endpoint.
D. Train and evaluate the model on premises. Upload the model to an Amazon S3 bucket. Deploy the model on an Amazon SageMaker hosting services endpoint.
Answer: C (LEAVE A REPLY)

NEW QUESTION: 168
A company that manufactures mobile devices wants to determine and calibrate the appropriate sales price for its devices. The company is collecting the relevant data and is determining data features that it can use to train machine learning (ML) models. There are more than 1,000 features, and the company wants to determine the primary features that contribute to the sales price.
Which techniques should the company use for feature selection? (Choose three.)
A. Data scaling with standardization and normalization
B. Correlation plot with heat maps
C. Data binning
D. Univariate selection
E. Feature importance with a tree-based classifier
F. Data augmentation
Answer: (SHOW ANSWER)
Reference:
https://towardsdatascience.com/feature-selection-using-python-for-classification-problem-b5f00a1c7028#:~:text=Univariate%20feature%20selection%20works%20by,analysis%20of%20variance%20(ANOVA).&text=That%20is%20why%20it%20is%20called%20'univariate'
https://arxiv.org/abs/2101.04530

NEW QUESTION: 169
A large mobile network operating company is building a machine learning model to predict customers who are likely to unsubscribe from the service. The company plans to offer an incentive for these customers as the cost of churn is far greater than the cost of the incentive.

The model produces the following confusion matrix after evaluating on a test dataset of 100 customers:

n= 100	PREDICTED CHURN	PREDICTED CHURN
	Yes	No
ACTUAL Churn Yes	10	4
Actual No	10	76

Based on the model evaluation results, why is this a viable model for production?

A. The model is 86% accurate and the cost incurred by the company as a result of false positives is less than the false negatives.

B. The precision of the model is 86%, which is less than the accuracy of the model.

C. The model is 86% accurate and the cost incurred by the company as a result of false negatives is less than the false positives.

D. The precision of the model is 86%, which is greater than the accuracy of the model.

Answer: B (LEAVE A REPLY)

NEW QUESTION: 170

A Machine Learning Specialist was given a dataset consisting of unlabeled data The Specialist must create a model that can help the team classify the data into different buckets What model should be used to complete this work?

A. Random Cut Forest (RCF)

B. XGBoost

C. K-means clustering

D. BlazingText

Answer: C (LEAVE A REPLY)

NEW QUESTION: 171

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. Setup up Amazon EMR with Hadoop Distributed File System (HDFS) to store the files, and restrict access to the EMR instances using IAM policies.

B. Use an Amazon S3-backed data lake to store the raw images, and set up the permissions using bucket policies.

C. Configure Amazon EFS with IAM policies to make the data available to Amazon EC2 instances owned by the IAM users.

D. 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.

Answer: A (LEAVE A REPLY)

NEW QUESTION: 172
A company is observing low accuracy while training on the default built-in image classification algorithm in Amazon SageMaker. The Data Science team wants to use an Inception neural network architecture instead of a ResNet architecture.
Which of the following will accomplish this? (Select TWO.)
A. Customize the built-in image classification algorithm to use Inception and use this for model training.
B. Create a support case with the SageMaker team to change the default image classification algorithm to Inception.
C. Bundle a Docker container with TensorFlow Estimator loaded with an Inception network and use this for model training.
D. Use custom code in Amazon SageMaker with TensorFlow Estimator to load the model with an Inception network and use this for model training.
E. Download and apt-get install the inception network code into an Amazon EC2 instance and use this instance as a Jupyter notebook in Amazon SageMaker.
Answer: D,E (LEAVE A REPLY)

NEW QUESTION: 173
A manufacturing company has structured and unstructured data stored in an Amazon S3 bucket. A Machine Learning Specialist wants to use SQL to run queries on this data.
Which solution requires the LEAST effort to be able to query this data?
A. Use AWS Lambda to transform the data and Amazon Kinesis Data Analytics to run queries.
B. Use AWS Batch to run ETL on the data and Amazon Aurora to run the queries.
C. Use AWS Glue to catalogue the data and Amazon Athena to run queries.
D. Use AWS Data Pipeline to transform the data and Amazon RDS to run queries.
Answer: (SHOW ANSWER)

NEW QUESTION: 174
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. Last observation carried forward
B. Listwise deletion
C. Mean substitution
D. Multiple imputation
Answer: D (LEAVE A REPLY)

NEW QUESTION: 175

A Machine Learning Specialist has built a model using Amazon SageMaker built-in algorithms and is not getting expected accurate results The Specialist wants to use hyperparameter optimization to increase the model's accuracy Which method is the MOST repeatable and requires the LEAST amount of effort to achieve this?

A. Create an AWS Step Functions workflow that monitors the accuracy in Amazon CloudWatch Logs and relaunches the training job with a defined list of hyperparameters

B. Launch multiple training jobs in parallel with different hyperparameters

C. Create a hyperparameter tuning job and set the accuracy as an objective metric.

D. Create a random walk in the parameter space to iterate through a range of values that should be used for each individual hyperparameter

Answer: A (LEAVE A REPLY)

NEW QUESTION: 176

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. Send Amazon CloudWatch Logs that were generated by Amazon SageMaker lo Amazon ES and use Kibana to query and visualize the log data.

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. Review SageMaker logs that have been written to Amazon S3 by leveraging Amazon Athena and Amazon OuickSight to visualize logs as they are being produced

D. Build custom Amazon CloudWatch Logs and then leverage Amazon ES and Kibana to query and visualize the data as it is generated by Amazon SageMaker

Answer: (SHOW ANSWER)

NEW QUESTION: 177

A Machine Learning Specialist has created a deep learning neural network model that performs well on the training data but performs poorly on the test data.

Which of the following methods should the Specialist consider using to correct this? (Select THREE.)

A. Decrease regularization.

B. Increase dropout.

C. Decrease dropout.

D. Decrease feature combinations.

E. Increase feature combinations.

F. Increase regularization.

Answer: (SHOW ANSWER)

Valid AWS-Certified-Machine-Learning-Specialty Dumps shared by Actual4test.com for Helping Passing AWS-Certified-Machine-Learning-Specialty Exam! Actual4test.com now offer the **newest AWS-Certified-Machine-Learning-Specialty exam dumps**, the Actual4test.com AWS-Certified-Machine-Learning-Specialty exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com AWS-Certified-Machine-Learning-Specialty dumps with Test Engine here: https://www.actual4test.com/AWS-Certified-Machine-Learning-Specialty_examcollection.html (**330** Q&As Dumps, **30%OFF** **Special Discount:** **Freepdfdumps**)