Free Questions for MLS-C01 by dumpssheet

Shared by Ferguson on 12-12-2023

For More Free Questions and Preparation Resources

Check the Links on Last Page

Question 1

Question Type: MultipleChoice

A machine learning specialist is developing a regression model to predict rental rates from rental listings. A variable named Wall_Color represents the most prominent exterior wall color of the property. The following is the sample data, excluding all other variables:

Property_ID	Wall_Color
1000	Red
1001	White
1002	Green

The specialist chose a model that needs numerical input data.

Which feature engineering approaches should the specialist use to allow the regression model to learn from the Wall_Color data? (Choose two.)

Options:

A- Apply integer transformation and set Red = 1, White = 5, and Green = 10.

B- Add new columns that store one-hot representation of colors.

C- Replace the color name string by its length.

D- Create three columns to encode the color in RGB format.

E- Replace each color name by its training set frequency.

Answer:

A, D

Question 2

Question Type: MultipleChoice

A company is building a demand forecasting model based on machine learning (ML). In the development stage, an ML specialist uses an Amazon SageMaker notebook to perform feature engineering during work hours that consumes low amounts of CPU and memory resources. A data engineer uses the same notebook to perform data preprocessing once a day on average that requires very high memory and completes in only 2 hours. The data preprocessing is not configured to use GPU. All the processes are running well on an ml.m5.4xlarge notebook instance.

The company receives an AWS Budgets alert that the billing for this month exceeds the allocated budget.

Which solution will result in the MOST cost savings?

Options:

A- Change the notebook instance type to a memory optimized instance with the same vCPU number as the ml.m5.4xlarge instance has.

Stop the notebook when it is not in use. Run both data preprocessing and feature engineering development on that instance.

B- Keep the notebook instance type and size the same. Stop the notebook when it is not in use. Run data preprocessing on a P3 instance type with the same memory as the ml.m5.4xlarge instance by using Amazon SageMaker Processing.

C- Change the notebook instance type to a smaller general purpose instance. Stop the notebook when it is not in use. Run data preprocessing on an ml.r5 instance with the same memory size as the ml.m5.4xlarge instance by using Amazon SageMaker Processing.

D- Change the notebook instance type to a smaller general purpose instance. Stop the notebook when it is not in use. Run data preprocessing on an R5 instance with the same memory size as the ml.m5.4xlarge instance by using the Reserved Instance option.

Answer:

B

Question 3

Question Type: **MultipleChoice**

A company wants to use automatic speech recognition (ASR) to transcribe messages that are less than 60 seconds long from a voicemail-style application. The company requires the correct identification of 200 unique product names, some of which have unique spellings or pronunciations.

The company has 4,000 words of Amazon SageMaker Ground Truth voicemail transcripts it can use to customize the chosen ASR model. The company needs to ensure that everyone can update their customizations multiple times each hour.

Which approach will maximize transcription accuracy during the development phase?

Options:

A- Use a voice-driven Amazon Lex bot to perform the ASR customization. Create customer slots within the bot that specifically identify each of the required product names. Use the Amazon Lex synonym mechanism to provide additional variations of each product name as mis-transcriptions are identified in development.

B- Use Amazon Transcribe to perform the ASR customization. Analyze the word confidence scores in the transcript, and automatically create or update a custom vocabulary file with any word that has a confidence score below an acceptable threshold value. Use this updated custom vocabulary file in all future transcription tasks.

C- Create a custom vocabulary file containing each product name with phonetic pronunciations, and use it with Amazon Transcribe to perform the ASR customization. Analyze the transcripts and manually update the custom vocabulary file to include updated or additional entries for those names that are not being correctly identified.

D- Use the audio transcripts to create a training dataset and build an Amazon Transcribe custom language model. Analyze the transcripts and update the training dataset with a manually corrected version of transcripts where product names are not being transcribed correctly. Create an updated custom language model.

Answer:

A

Question 4

Question Type: MultipleChoice

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?

Options:

A- Autoregressive Integrated Moving Average (AIRMA)

B- Exponential Smoothing (ETS)

C- Convolutional Neural Network - Quantile Regression (CNN-QR)

D- Prophet

Answer:

B

Question 5

Question Type: MultipleChoice

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.)

Options:

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:

C, D, F

Explanation:

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%20

https://arxiv.org/abs/2101.04530

Question 6

Question Type: MultipleChoice

A data scientist is training a text classification model by using the Amazon SageMaker built-in BlazingText algorithm. There are 5 classes in the dataset, with 300 samples for category A, 292 samples for category B, 240 samples for category C, 258 samples for category D, and 310 samples for category E.

The data scientist shuffles the data and splits off 10% for testing. After training the model, the data scientist generates confusion matrices for the training and test sets.

Training data confusion matrix

		Predicted class					
		A	**B**	**C**	**D**	**E**	**Total**
True class	**A**	270	0	0	0	0	270
	B	1	260	0	0	2	263
	C	0	0	111	100	5	216
	D	4	3	132	92	1	232
	E	0	0	2	3	274	279
	Total	275	263	245	195	282	1260

Test data confusion matrix

		Predicted class					
		A	**B**	**C**	**D**	**E**	**Total**
True class	**A**	9	1	0	0	0	10
	B	2	25	0	2	0	29
	C	10	2	11	10	1	34
	D	1	0	12	14	0	27
	E	9	1	4	1	25	40
	Total	31	29	27	27	26	140

What could the data scientist conclude form these results?

Options:

A- Classes C and D are too similar.

B- The dataset is too small for holdout cross-validation.

C- The data distribution is skewed.

D- The model is overfitting for classes B and E.

Answer:

B

Question 7

Question Type: MultipleChoice

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.)

Options:

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

Explanation:

https://towardsdatascience.com/image-processing-techniques-for-computer-vision-11f92f511e21

https://docs.aws.amazon.com/rekognition/latest/customlabels-dg/training-model.html