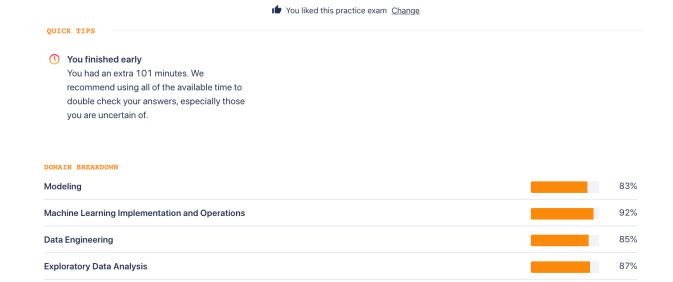
CloudGuru - Practice Exam - 3rd attempt

https://practice-exam.acloud.guru/aws-certified-machine-learning-specialty? courseld=aws-certified-machine-learning-specialty&_ga=2.151769904.849115380.1605014235-906779002.1602959981&_gac=1.251718395.1602959981.CjwKCAjwrKr8BRB_EiwA7eFaphu95ArJm-GaAsgiegqwqcDAUIPEYAwg5tU6E1-93RcNJU2Y23-VXRoCzacQAvD_BwE

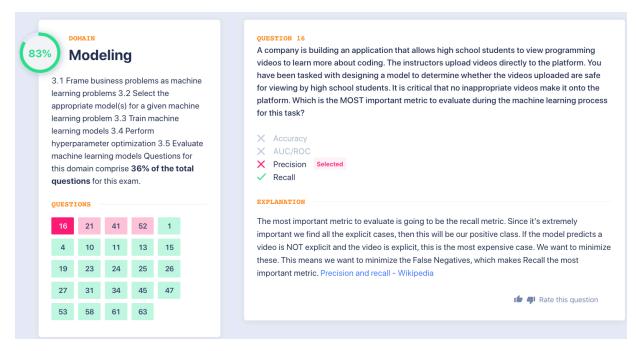


You scored 86%

You answered 56 of 65 questions correctly in 01:18:06



Questions I did not get right:



I put the question is: Is video safe? Y/N as the question refers to "safe". We want to limit false positives, where we say a video is safe but it is not => Precision

But the other way to ask the question I: IS the video explicit? Y/N In that case, we want to limit false negatives, where we way a video is not explicit but it is => Recall

So it all comes down the the formulation of the binary classifier problem

You are consulting with a large educational organization on a ML model using the built-in BlazingText SageMaker algorithm. They have asked for help deciding which metric to use in an automatic model tuning job. What can you recommend to help them get started?

- The metrics to use for optimization can vary depending on how you are using the algorithm.
- X Metrics with a prefix of validation: are the ones to always use for optimization jobs. Selected
- X The proper metric for BlazingText optimization is validation:accuracy.
- X Metrics with a prefix of train: are the ones to always use for optimization jobs.
- X BlazingText is not supported with hyperparameter tuning.

EXPLANATION

BlazingText optimization is an example of where we would use different metrics depending on how we are using the algorithm. You should consult the documentation for the algorithm you are using to determine the proper metric. Tune a BlazingText Model - Amazon SageMaker

I agree with that as I knew we could use BlazingText in Word2Vec mode or Text Classification.

But what I selected is correct too.

You are trying to follow an old family recipe for making sourdough bread. The recipe states that you should add enough water to the flour such that it is slightly sticky. Which if the following best characterizes this instruction?

- X Algorithm Selected
- X Supervised Learning
- ✓ Heuristic
- X Unsupervised Learning
- X Reinforcement Learning
- × Metric

EXPLANATION

A heuristic technique, often called simply a heuristic, is any approach to problem solving or self-discovery that employs a practical method, not guaranteed to be optimal, perfect, or rational, but instead sufficient for reaching an immediate goal. Where finding an optimal solution is impossible or impractical, heuristic methods can be used to speed up the process of finding a satisfactory solution. A heuristic is an educated guess or intuition that usually does not ensure an optimal outcome. By asking the chef to evaluate the quantity of water needed based on individual perception best embodies a heuristic. ALGORITHMS AND HEURISTICS

A binary classification model has been created to sort parts on an assembly line into acceptable or unacceptable, based on a complex array of readings. The model incorrectly decides that some flawed parts are acceptable when they should have been marked as unacceptable. Which of the following correctly defines this type of result?

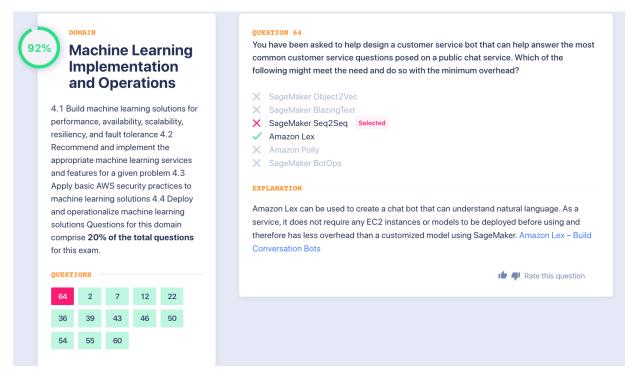
- False Negative
- X False Positive Selected
- X Type I Error Selected
- X True Negative
- ✓ Type II Error
- X True Positive

EXPLANATION

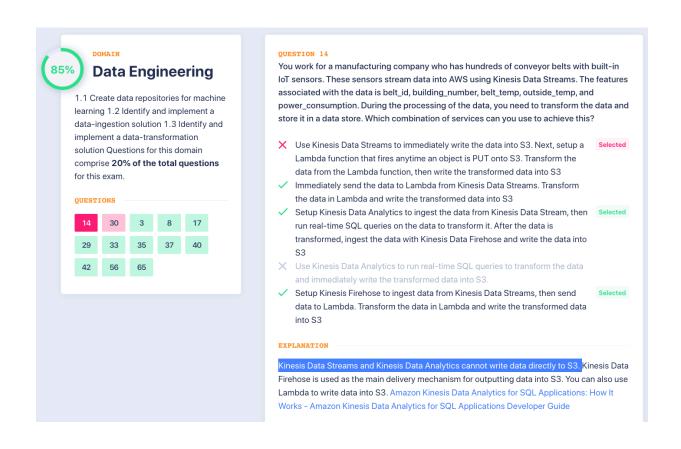
A false negative is when a model does not properly flag something that is should have. In this case, the flawed parts that got through as good parts were incorrectly identified as ok when they were bad. A false negative is also known as a Type II error. Type I and type II errors - Wikipedia

Same here, formatting the question is key. It should be: is a part "unacceptable"? In that case Type II makes sense with False negatives (not this part is acceptable when it is not)

I had formulated as: is a part "acceptable?



That one was obvious - I misread it and should have focused on "bot"



A machine learning model is being created using Amazon's Factorization Machines algorithm to help make click predictions and item recommendations for new customers. Which of the following would be candidates during the training process?

- Creating a binary classification model where the testing dataset is scored using Binary Cross Entropy (Log Loss), Accuracy, and F1 Score.
- X Using sparse data in CSV format as training data.
- X Making inferences to the model in application/csv format.
- Using sparse data in recordIO-protobuf format with Float32 tensors as training data.

Selected

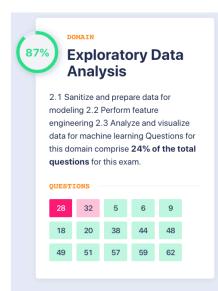
- Creating a regression model where the testing dataset is scored using Root Mean Square Error (RMSE).
- Selected

Selected

Creating a multi-classification model where the testing dataset is scored using Area Under The Curve (AUC).

EXPLANATION

The factorization machine algorithm can be run in either in binary classification mode or regression mode. In regression mode, the testing dataset is scored using Root Mean Square Error (RMSE). In binary classification mode, the test dataset is scored using Binary Cross Entropy (Log Loss), Accuracy (at threshold=0.5) and F1 Score (at threshold=0.5). For training, the factorization machines algorithm currently supports only the recordIO-protobuf format with Float32 tensors. CSV format is not a good candidate. For inference, factorization machines support the application/json and x-recordio-protobuf formats. Factorization Machines Algorithm - Amazon SageMaker



OUESTION 28

You have been tasked with creating a labeled dataset by classifying text data into different categories depending on the summary of the corpus. You plan to use this data with a particular machine learning algorithm within AWS. Your goal is to make this as streamlined as possible with minimal amount of setup from you and your team. What tool can be used to help label your dataset with the minimum amount of setup?

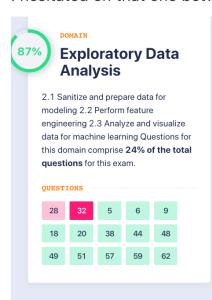
- ★ Marketplace AMI for NLP problems
- X Amazon Neural Topic Modeling (NTM) built-in algorithm
- X Amazon Latent Dirichlet Allocation (LDA) algorithm
- × AWS Comprehend sentiment analysis
- AWS SageMaker GroundTruth text classification job
- X AWS Comprehend entity detection Selected

EXPLANATION

You can use SageMaker Ground Truth to create ground truth datasets by creating labeling jobs. When you create a text classification job, workers group text into the categories that you define. You can define multiple categories but the worker can apply only one category to the text. Use the instructions to guide your workers to make the correct choice. Always define a generic class in addition to your specific classes. Giving your workers a generic option helps to minimize inaccurately classified text. Amazon SageMaker Ground Truth - Amazon SageMaker

Rate this question

I hesitated on that one between the 2



QUESTION 32

You are preparing plain text corpus data to use in a NLP process. Which of the following is/are one of the important step(s) to pre-process the text in NLP based projects?

- ✓ Stemming Selected
- ✓ Word standardization
- X Congregate all of the plain text corpus data into a single document
- X Add random text noise
- ✓ Stop word removal Selected
- X One-hot encode ordinal n-gram values Selected

EXPLANATION

Stemming is a rudimentary rule-based process of stripping the suffixes ("ing", "ly", "es", "s" etc) from a word. Stop words are those words which will have no relevance to the context of the data for example is/am/are. Object Standardization is also one of the good ways to preprocess the text by removing things like acronyms, hashtags with attached words, and colloquial slang that typically are not recognized by search engines and models. Natural language processing - Wikipedia

Rate this question