



- Expert Verified, Online, Free.

MENU



Google Discussions



Exam Professional Machine Learning Engineer All Questions

View all questions & answers for the Professional Machine Learning Engineer exam

Go to Exam

EXAM PROFESSIONAL MACHINE LEARNING ENGINEER TOPIC 1 QUESTION 108 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 108

Topic #: 1

[\[All Professional Machine Learning Engineer Questions\]](#)

You are experimenting with a built-in distributed XGBoost model in Vertex AI Workbench user-managed notebooks. You use BigQuery to split your data into training and validation sets using the following queries:

```
CREATE OR REPLACE TABLE 'myproject.mydataset.training' AS
(SELECT * FROM 'myproject.mydataset.mytable' WHERE RAND() <= 0.8);
```

```
CREATE OR REPLACE TABLE 'myproject.mydataset.validation' AS
(SELECT * FROM 'myproject.mydataset.mytable' WHERE RAND() <= 0.2);
```

After training the model, you achieve an area under the receiver operating characteristic curve (AUC ROC) value of 0.8, but after deploying the model to production, you notice that your model performance has dropped to an AUC ROC value of 0.65. What problem is most likely occurring?

- A. There is training-serving skew in your production environment.
- B. There is not a sufficient amount of training data.
- C. The tables that you created to hold your training and validation records share some records, and you may not be using all the data in your initial table.
- D. The RAND() function generated a number that is less than 0.2 in both instances, so every record in the validation table will also be in the training table.

Show Correct Answer

by  mymy9418 at Dec. 17, 2022, 11:47 a.m.

Comments

Type your comment...


Submit

  **eico** 1 month, 4 weeks ago

Selected Answer: C

Answer C

   upvoted 1 times

  **M25** 1 year, 5 months ago

Selected Answer: C

- Excluding D as RAND() samples 80% for ".training" & 20% for ".validaton":
<https://stackoverflow.com/questions/42115968/how-does-rand-works-in-bigquery>;
- Could be that those 2 samplings share some records since pseudo-randomly sampled over the same ".mytable", & therefore might not be using all of its data, thus C seems valid;
- Excluding B as there is no indication otherwise of insufficient amount of training data, after training AUC ROC was 0.8, that we know;
- There could be a training-serving skew occurring in Prod, but "most likely occurring" is C as a result of the selective information presented: https://developers.google.com/machine-learning/guides/rules-of-ml#training-serving_skew

   upvoted 4 times

  **formazioneQI** 1 year, 6 months ago

Selected Answer: C

Answer C


   upvoted 2 times

  **Yajnas_arpohc** 1 year, 7 months ago

Selected Answer: C

C seems closest here



   upvoted 1 times

  **TNT87** 1 year, 7 months ago

Selected Answer: C

Answer C

   upvoted 1 times

  **ailiba** 1 year, 8 months ago

Selected Answer: C

since we are calling rand twice it might be that data that was in training set ends up in testing set too. If we had called it just once I would say D.

   upvoted 2 times

  **Ahmades** 1 year, 9 months ago

Selected Answer: D

Hesitated between C and D, but D looks more precise

   upvoted 1 times

  **pshemol** 1 year, 8 months ago

If there were one RAND() in front of those two queries it would be true. There are two separate RAND() and "every record in the validation table will also be in the training table" is not true.

   upvoted 2 times

  **hiromi** 1 year, 10 months ago

Selected Answer: C

C (not sure)

   upvoted 4 times

  **mymy9418** 1 year, 10 months ago

Selected Answer: C

the rand is generated twice

   upvoted 2 times

Start Learning for free



Social Media

[Facebook](#) , [Twitter](#)

[YouTube](#) , [Reddit](#)

[Pinterest](#)



We are the biggest and most updated IT certification exam material website.

Using our own resources, we strive to strengthen the IT professionals community for free.



© 2024 ExamTopics

ExamTopics doesn't offer Real Microsoft Exam Questions. ExamTopics doesn't offer Real Amazon Exam Questions. ExamTopics Materials do not

contain actual questions and answers from Cisco's Certification Exams.

CFA Institute does not endorse, promote or warrant the accuracy or quality of ExamTopics. CFA® and Chartered Financial Analyst® are registered trademarks owned by CFA Institute.