



- 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 102 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 102

Topic #: 1

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

You have successfully deployed to production a large and complex TensorFlow model trained on tabular data. You want to predict the lifetime value (LTV) field for each subscription stored in the BigQuery table named subscription. subscriptionPurchase in the project named my-fortune500-company-project.

You have organized all your training code, from preprocessing data from the BigQuery table up to deploying the validated model to the Vertex AI endpoint, into a TensorFlow Extended (TFX) pipeline. You want to prevent prediction drift, i.e., a situation when a feature data distribution in production changes significantly over time. What should you do?

- A. Implement continuous retraining of the model daily using Vertex AI Pipelines.
- B. Add a model monitoring job where 10% of incoming predictions are sampled 24 hours.
- C. Add a model monitoring job where 90% of incoming predictions are sampled 24 hours.
- D. Add a model monitoring job where 10% of incoming predictions are sampled every hour.

Show Suggested Answer

by hiromi at Dec. 20, 2022, 11:36 a.m.

## Comments

Type your comment...



[Submit](#)

  **MultiCloudIronMan** 7 months ago

**Selected Answer: B**

You need to monitor it first and foremost to see if there is a drift and if there is then a measure can be devised. training every date is an over kill.

   upvoted 2 times

  **pico** 1 year, 1 month ago

**Selected Answer: A**

Continuous Retraining: Continuously retraining the model allows it to adapt to changes in the data distribution, helping to mitigate prediction drift. Daily retraining provides a good balance between staying up-to-date and avoiding excessive retraining.

Options B, C, and D involve model monitoring but do not address the issue of keeping the model updated with the changing data distribution. Monitoring alone can help you detect drift, but it does not actively prevent it. Retraining the model is necessary to address drift effectively.



   upvoted 3 times

  **maukaba** 1 year ago

Option A can prevent drift prediction. All the other options can only detect.

Therefore the correct answer is A unless it is possible to monitor drifts and then remediate without retrainings.

   upvoted 1 times

  **Nish1729** 10 months ago

Follow me on X (twitter): @nbcodes for more useful tips.

I think you're slightly missing the point, the answer should be B, let me explain why..

The whole point of this question is to come up with a PREVENTATIVE way of handling prediction drift so you need to find a way to DETECT the drift before it occurs, this is exactly what solution B does and ensures it's done in a way that is not too frequent i.e D and not too resource intensive with the large sample i.e C remember if sampling is done well you don't need 90% of the data to detect drift.

Solution A suggests retraining every day which is a CRAZY proposal, why would you retrain every day even if you don't know if your data is drifting?? Huge waste of resources and time.

   upvoted 2 times

  **M25** 1 year, 6 months ago

**Selected Answer: B**

Went with B

   upvoted 1 times

  **tavva\_prudhvi** 1 year, 7 months ago

**Selected Answer: B**

Continuous retraining (option A) is not necessarily the best solution for preventing prediction drift, as it can be time-consuming and expensive. Instead, monitoring the performance of the model in production is a better approach. Option B is a good choice because it samples a small percentage of incoming predictions and checks for any significant changes in the feature data distribution over a 24-hour period. This allows you to detect any drift and take appropriate action to address it before it affects the model's performance. Options C and D are less effective because they either sample too many or too few predictions and/or at too frequent intervals.

   upvoted 4 times

  **andresvelasco** 1 year, 1 month ago

I am just not sure why sampling too few (10%) is important. Is this a costly service?

   upvoted 1 times

  **tavva\_prudhvi** 12 months ago

Model monitoring, especially at a large scale, can consume significant computational resources. Sampling a smaller percentage of predictions (like 10%) helps manage these resource demands and associated costs. The more predictions you sample, the more storage, computation, and network resources you'll need to analyze the data, potentially increasing the cost.

In many cases, a 10% sample of the data can provide statistically significant insights into the model's performance and the presence of drift. It's a balancing act between getting enough data to make informed decisions and not overburdening the system.

In some datasets, especially large ones, a lot of the data might be redundant or not particularly informative. Sampling a

smaller fraction can help filter out noise and focus on the most relevant information.



   upvoted 1 times

  **pico** 11 months, 3 weeks ago

Neither B,C or D have a step to prevent the prediction drift.

The question says: "you want to prevent prediction drift"

   upvoted 1 times

  **TNT87** 1 year, 8 months ago

**Selected Answer: B**

Answer B

   upvoted 1 times

  **John\_Pongthorn** 1 year, 9 months ago

**Selected Answer: B**

B , I got it from Machine Learning in the Enterprise course for google partnet skillboost  
you can watch carefully on video "Model management using Vertex AI"  
I imply that it is default setting on typical case.



   upvoted 3 times

  **behzadsw** 1 year, 10 months ago

**Selected Answer: D**

Using 10% of hourly requests would yield a better distribution and faster feed back loop

   upvoted 1 times

  **hargur** 1 year, 10 months ago

I think it is B, we can say 10% to be a sample but not 90%



   upvoted 2 times

  **mymy9418** 1 year, 10 months ago

**Selected Answer: B**

I guess 10% of 24 hours should be good enough?

   upvoted 3 times

  **hiromi** 1 year, 10 months ago

**Selected Answer: B**

B (not sure)

- <https://cloud.google.com/vertex-ai/docs/model-monitoring/overview>

- <https://cloud.google.com/vertex-ai/docs/model-monitoring/using-model-monitoring#drift-detection>

   upvoted 2 times

**Start Learning for free**



**Social Media**

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

[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.