

🔗 Google Discussions



Exam Professional Data Engineer All Questions

View all questions & answers for the Professional Data Engineer exam

Go to Exam

📄 EXAM PROFESSIONAL DATA ENGINEER TOPIC 1 QUESTION 2 DISCUSSION

Actual exam question from Google's Professional Data Engineer

Question #: 2

Topic #: 1

[\[All Professional Data Engineer Questions\]](#)

You are building a model to make clothing recommendations. You know a user's fashion preference is likely to change over time, so you build a data pipeline to stream new data back to the model as it becomes available. How should you use this data to train the model?

- A. Continuously retrain the model on just the new data.
- B. Continuously retrain the model on a combination of existing data and the new data.
- C. Train on the existing data while using the new data as your test set.
- D. Train on the new data while using the existing data as your test set.

Show Suggested Answer

by RP123 at March 9, 2020, 9:40 a.m.

Comments

Type your comment...

Submit

🗨️ **serg3d** Highly Voted 4 years, 11 months ago

I think it should be B because we have to use a combination of old and new test data as well as training data

👍 👎 🗳️ upvoted 27 times

upvoted 37 times

dambilwa 4 years, 10 months ago

Yes - The training set should be shuffled well to represent data across all scenarios

upvoted 4 times

jagadamba **Highly Voted** 4 years, 10 months ago

B, as we need to train the data with new data, so that it will keep learning, and as well as used for test

upvoted 11 times

Ahamada **Most Recent** 2 months, 1 week ago

Selected Answer: B

Answer is B, need to take account both data to train

upvoted 1 times

cqrm3n 3 months, 3 weeks ago

Selected Answer: B

We should continuously retrain existing data with latest data to balance the need to adapt to new changes while not overriding historical knowledge. This is called Continuous Retraining where the model is periodically updated with latest data so that recommendations remain accurate over time.

upvoted 1 times

jaimecalderon 4 months, 1 week ago

Selected Answer: B

Continuously retrain the model on a combination of existing data and the new data.

upvoted 1 times

SamuelTsch 6 months, 2 weeks ago

Selected Answer: B

From my point of view, we should take both datasets.

upvoted 1 times

rocky48 1 year, 6 months ago

Selected Answer: B

Option A is not recommended because retraining the model on just new data will cause the model to lose the information it has learned from the historical data.

Option C and D are not recommended because they are using the new data as test set and this approach will lead to a model that is overfitting and not generalize well to new users.

So answer is B

upvoted 2 times

rajkinz 1 year, 6 months ago

Answer is C. It is time sensitive data so latest data should be used for testing.

Reference: <https://cloud.google.com/automl-tables/docs/prepare#ml-use>

upvoted 2 times

rtcpst 1 year, 6 months ago

Selected Answer: B

This approach allows the model to benefit from both the historical data (existing data) and the new data, ensuring that it adapts to changing preferences while retaining knowledge from the past. By combining both types of data, the model can learn to make recommendations that are up-to-date and relevant to users' evolving preferences.

upvoted 1 times

Websurfer 1 year, 9 months ago

Selected Answer: B

train on old and new data

upvoted 1 times

AmmarFasih 1 year, 11 months ago

Selected Answer: B

Option B is the right answer. Since the questions states the models needs to be updated since the clothing preference changes. Hence we need the new data to be utilized for training/ updating model.

upvoted 1 times

bha11111 2 years, 1 month ago

Selected Answer: B

Have verified this

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 jin0 2 years, 2 months ago

there are two point first when retraining second what data. I think retraining should be occur when the model could not predict well in this case there is monitoring metric should be needed first but no one said, second what data? in this case I think the answer is A. because when the model could not predict well it means the data variance and bias are changed so, it's no make sense what is combination new data with old data because the data being not be changed is not necessary anymore..

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 jin0 2 years, 2 months ago

And the questions should explain in detail.. whether it's deep learning or tree based machine learning model.. and how large of new dataset is.. I think

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 Morock 2 years, 2 months ago

Selected Answer: C

The trend keep changing, so must mix new and old data...

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 samdhimal 2 years, 3 months ago

Selected Answer: B

B. Continuously retrain the model on a combination of existing data and the new data.

This approach will help to ensure that the model remains up-to-date with the latest fashion preferences of the users, while also leveraging the historical data to provide context and improve the accuracy of the recommendations. Retraining the model on a combination of existing and new data will help to prevent the model from being overly influenced by the new data and losing its ability to generalize to users with different preferences.

Option A is not recommended because retraining the model on just new data will cause the model to lose the information it has learned from the historical data.

Option C and D are not recommended because they are using the new data as test set and this approach will lead to a model that is overfitting and not generalize well to new users.

👍 ↩ 🚩 upvoted 6 times

🗨️ 👤 rocky48 1 year, 6 months ago

Nice explanation bro.

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 korntewin 2 years, 3 months ago

The answer can be A, if we implement online learning! But for regular model which can't implement online learning (everything with no gradient descent) the answer should be B.

👍 ↩ 🚩 upvoted 1 times

🗨️ 👤 testoneAZ 2 years, 3 months ago

Correct answer is B

👍 ↩ 🚩 upvoted 1 times

[Load full discussion...](#)



Platform

> Home

> Examtopics PRO

> All Exams

> Training Courses

