EXAMTOPICS

- 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 269 DISCUSSI...**

Actual exam question from Google's Professional Machine Learning Engineer
Question #: 269
Topic #: 1

[All Professional Machine Learning Engineer Questions]

You are developing a model to help your company create more targeted online advertising campaigns. You need to create a dataset that you will use to train the model. You want to avoid creating or reinforcing unfair bias in the model. What should you do? (Choose two.)

    A. Include a comprehensive set of demographic features

    B. Include only the demographic groups that most frequently interact with advertisements

    C. Collect a random sample of production traffic to build the training dataset

    D. Collect a stratified sample of production traffic to build the training dataset

    E. Conduct fairness tests across sensitive categories and demographics on the trained model

Show Suggested Answer

by 👤 pikachu007 at *Jan. 13, 2024, 4:20 p.m.*

Comments

Type your comment...

Submit

wences 1 month, 1 week ago

Selected Answer: D

From my statistical point of view, D and E will mitigate the effect of bias.

👍 ↩ 🚩 upvoted 1 times

AzureDP900 3 months, 2 weeks ago

D and E is right answer, question asks us to select 2 right answers
• To avoid creating or reinforcing unfair bias in the model, you should collect a representative and diverse dataset (option D) that includes a stratified sample of production traffic. This ensures that your training data is inclusive and accurately represents the diversity of your target audience.
• Once you have collected your training dataset, you should conduct fairness tests across sensitive categories and demographics on the trained model (option E). This involves evaluating whether the model treats different demographic groups fairly and without bias. If biases are detected, you can take steps to mitigate them and ensure that your model is fair and accurate.

👍 ↩ 🚩 upvoted 1 times

---

dija123 3 months, 3 weeks ago

Selected Answer: D

Agree with D and E

👍 ↩ 🚩 upvoted 1 times

---

omermahgoub 6 months, 1 week ago

Selected Answer: D

D. Stratified sampling to ensure the different demographic groups or categories are proportionally represented in the training data. This helps mitigate bias that might arise if certain groups are under-represented.
E. Fairness tests can reveal disparities in how the model treats different populations, allowing you to identify and address potential biases.

👍 ↩ 🚩 upvoted 3 times

MultiCloudIronMan 6 months, 2 weeks ago

Selected Answer: D

D and E is the two answers. Two selections are required

👍 ↩ 🚩 upvoted 2 times

> **pinimichele01** 6 months, 2 weeks ago
>
> why not D and A?
>
> 👍 ↩ 🚩 upvoted 2 times

CHARLIE2108 8 months ago

Selected Answer: D

I went D, E

👍 ↩ 🚩 upvoted 2 times

---

guilhermebutzke 8 months, 1 week ago

Selected Answer: D

DE

D. Collect a stratified sample of production traffic to build the training dataset: This ensures that the training data represents the diverse demographics that will be targeted by the advertising campaigns. Random sampling might unintentionally underrepresent certain groups, leading to biased model outputs.

E. Conduct fairness tests across sensitive categories and demographics on the trained model: This allows you to identify and address any potential biases that may have emerged during the training process. Evaluating the model's performance on different groups helps ensure fair and responsible deployment.

👍 ↩ 🚩 upvoted 1 times

---

daidai75 9 months ago

Selected Answer: D

I go for D & E:
A stratified sample ensures that the training data represents the distribution of the target population across relevant demographics or other sensitive categories. This helps mitigate bias arising from underrepresented groups in the data. Regularly testing the model for fairness across sensitive categories helps identify and address potential bias issues before deploying the model in production. This can involve metrics like precision, recall, and F1 score for different demographic groups.

👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **b1a8fae** 9 months ago

Selected Answer: D

D E. ChatGPT explanation below (but I think makes quite a lot of sense)

Collect a Stratified Sample (Option D): Stratified sampling involves dividing the population into subgroups (strata) and then randomly sampling from each subgroup. This ensures that the training dataset represents the diversity of the population, helping to avoid biases. By collecting a stratified sample of production traffic, you are more likely to have a balanced representation of different demographic groups, reducing the risk of biased model outcomes.

Conduct Fairness Tests (Option E): After training the model, it's crucial to conduct fairness tests to evaluate its performance across different sensitive categories and demographics. This involves measuring the model's predictions and outcomes for various groups to identify any disparities. Fairness tests help you assess and address biases that may have been inadvertently introduced during the training process.

👍 ↩ 🏳 upvoted 3 times

⊟ 👤 **shadz10** 9 months, 1 week ago

Selected Answer: C

C, D - Conducting fairness tests across sensitive categories and demographics on the trained model is indeed important. However, this option focuses on post-training analysis rather than dataset creation. While it's a crucial step for ensuring fairness, it doesn't directly address how to create a training dataset to avoid bias.
Hence C,D

👍 ↩ 🏳 upvoted 1 times

   ⊟ 👤 **tavva_prudhvi** 8 months, 1 week ago

   Check b1a8fae comment on why D is better than C!

   👍 ↩ 🏳 upvoted 1 times

⊟ 👤 **pikachu007** 9 months, 1 week ago

Selected Answer: D

D. Stratified Sampling: Randomly sampling your data might not accurately represent the diversity of your target audience, potentially introducing bias by over- or under-representing certain demographics. Stratified sampling ensures your training dataset reflects the distribution of sensitive features (e.g., age, gender, income) observed in your production traffic, helping mitigate bias during model training.

E. Fairness Testing: Simply collecting unbiased data isn't enough. Regularly testing your trained model for fairness across sensitive categories is crucial. This involves measuring and analyzing metrics like accuracy, precision, recall, and F1 score for different demographic groups. Identifying disparities in performance can trigger further investigation and potential re-training to address bias.

👍 ↩ 🏳 upvoted 1 times

Start Learning for free

EXAMTOPICS

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.