

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 159 DISCUSSI...

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 159

Topic #: 1

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

Your organization manages an online message board. A few months ago, you discovered an increase in toxic language and bullying on the message board. You deployed an automated text classifier that flags certain comments as toxic or harmful. Now some users are reporting that benign comments referencing their religion are being misclassified as abusive. Upon further inspection, you find that your classifier's false positive rate is higher for comments that reference certain underrepresented religious groups. Your team has a limited budget and is already overextended. What should you do?

- A. Add synthetic training data where those phrases are used in non-toxic ways.
- B. Remove the model and replace it with human moderation.
- C. Replace your model with a different text classifier.
- D. Raise the threshold for comments to be considered toxic or harmful.

[Show Suggested Answer](#)

by  PST21 at July 20, 2023, 4:10 p.m.

Comments

Type your comment...

[Submit](#)

  **Dirtie_Sinkie** 1 month ago

Selected Answer: A

Gonna go with A on this one. Some toxic comments will still make it through if you choose D, whereas A addresses the problem fully and directly. Therefore I think A is a more complete answer than D.

   upvoted 1 times

  **Dirtie_Sinkie** 1 month ago

Even though in the question it says "Your team has a limited budget and is already overextended" I still think A is the better answer because it doesn't take much effort to create synthetic data and add it to train. The outcome will be more accurate than D.

   upvoted 1 times

  **baimus** 1 month, 1 week ago

Selected Answer: A

A is better than D, because D means that more genuinely toxic comments will make it through. A will teach the model to acknowledge the small subset of mislabelled comments, without exposing the customers to additional toxicity.

   upvoted 1 times

  **AzureDP900** 4 months ago

option A (Add synthetic training data where those phrases are used in non-toxic ways) directly addresses the specific issue of bias and improves the model's accuracy by providing more contextually relevant training examples. This approach is more targeted and has a lower risk of introducing new biases or negatively impacting other aspects of comment moderation.

I hope this additional explanation helps clarify why option D might not be the best choice in this scenario!

   upvoted 2 times

  **AzureDP900** 4 months ago

Raising the threshold would mean increasing the minimum score required for a comment to be classified as toxic or harmful. This could potentially reduce the number of false positives (benign comments being misclassified as toxic) by making it harder for the model to classify a comment as toxic.

   upvoted 1 times

  **Simple_shreedhar** 4 months, 4 weeks ago

A option directly addresses the bias issue without incurring significant ongoing costs or burdening the moderation team. By augmenting the training dataset with synthetic examples where phrases related to underrepresented religious groups are used in non-toxic ways, the classifier can learn to distinguish between toxic and benign comments more accurately.


   upvoted 2 times

  **gscharly** 6 months ago

Selected Answer: D

agree with daidai75

   upvoted 1 times

  **pinimichele01** 6 months, 1 week ago

Selected Answer: D

Your team has a limited budget and is already overextended

   upvoted 2 times

  **7cb0ab3** 6 months, 2 weeks ago

Selected Answer: A

I went for A because it directly tackles the issue of misclassification and improving the model's understanding of religious references. B and C don't make sense.

D would generally reduce the number of comments flagged as toxic, which could decrease the false positive rate. However, this approach risks allowing genuinely harmful comments to go unflagged. It addresses the symptom (high false positive rate) rather than the underlying cause

   upvoted 2 times

  **edoo** 7 months, 2 weeks ago

Selected Answer: A

B and C are non sense, I don't want to risk potentially increasing the FNR by reducing the FPR (Raise the threshold). Thus A.

   upvoted 1 times

  **daidai75** 8 months, 3 weeks ago

Selected Answer: D

Your team has a limited budget and is already overextended, that means the re-training is hardly possible.

   unvoted 2 times

  **tavva_prudhvi** 1 year, 2 months ago

In the long run, usually we go with A, but Option D could be a temporary solution to reduce false positives, while being aware that it may allow some genuinely toxic comments to go unnoticed. However, this may be a necessary trade-off until your team has the resources to improve the classifier or find a better solution.

   upvoted 1 times

  **powerby35** 1 year, 3 months ago

Selected Answer: D

"Your team has a limited budget and is already overextended"

   upvoted 2 times

  **[Removed]** 1 year, 3 months ago

Selected Answer: D

By raising the threshold for comments to be considered toxic or harmful, you will decrease the number of false positives.

B is wrong because we are taking a Google MLE exam :) A and C are wrong because both of them involve a good amount of additional work, either for extending the dataset or training/experimenting with a new model. Considering your team is already over the budget and has too many tasks on their plate (overextended), these two options are not available for you.

   upvoted 2 times

  **tavva_prudhvi** 11 months, 1 week ago

But, by raising the threshold, we might be allowing some genuinely toxic comments to pass through without being flagged. This could potentially lead to an increase in the false negative rate, right?

   upvoted 1 times

  **PST21** 1 year, 3 months ago

Selected Answer: A

A. Add synthetic training data where those phrases are used in non-toxic ways.

In this situation, where your automated text classifier is misclassifying benign comments referencing certain underrepresented religious groups as toxic or harmful, adding synthetic training data where those phrases are used in non-toxic ways can be a cost-effective solution to improve the model's performance.

   upvoted 1 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.