

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 85 DISCUSSIO..

Actual exam question from Google's Professional Machine Learning Engineer

Question #: 85

Topic #: 1

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

You work for a large social network service provider whose users post articles and discuss news. Millions of comments are posted online each day, and more than 200 human moderators constantly review comments and flag those that are inappropriate. Your team is building an ML model to help human moderators check content on the platform. The model scores each comment and flags suspicious comments to be reviewed by a human. Which metric(s) should you use to monitor the model's performance?

- A. Number of messages flagged by the model per minute
- B. Number of messages flagged by the model per minute confirmed as being inappropriate by humans.
- C. Precision and recall estimates based on a random sample of 0.1% of raw messages each minute sent to a human for review
- D. Precision and recall estimates based on a sample of messages flagged by the model as potentially inappropriate each minute

Show Suggested Answer

by [ares81](#) at Dec. 12, 2022, 8:32 a.m.

Comments

Type your comment...

Submit

  **hiromi** Highly Voted  1 year, 10 months ago

Selected Answer: D

D

- <https://cloud.google.com/natural-language/automl/docs/beginners-guide>

- <https://cloud.google.com/vertex-ai/docs/text-data/classification/evaluate-model>

   upvoted 11 times

  **andresvelasco** Highly Voted  1 year, 1 month ago

Selected Answer: C

A. Number of messages flagged by the model per minute => NO, no measure of model performance

B. Number of messages flagged by the model per minute confirmed as being inappropriate by humans.=> DONT THINK SO, because we need the total number of messages (flagged?)

C. Precision and recall estimates based on a random sample of 0.1% of raw messages each minute sent to a human for review. => I think YES, because as I understand it that would be based on a sample of ALL messages not just the ones that have been flagged.

D. Precision and recall estimates based on a sample of messages flagged by the model as potentially inappropriate each minute => I think NO, because the sample includes only flagged messages, meaning positives, so you cannot really measure recall.

   upvoted 7 times

  **tavva_prudhvi** 11 months, 2 weeks ago

The main issue with option C is that it uses a random sample of only 0.1% of raw messages. This random sample might not contain enough examples of inappropriate content to accurately assess the model's performance. Since the majority of messages on the platform are likely appropriate, the random sample may not capture enough inappropriate content for a robust evaluation.

   upvoted 4 times

  **amene** Most Recent  3 weeks, 5 days ago

Selected Answer: B

I went with B.

Remember how to calculate Recall: $TP/(TP+FN)$. Since "sample of messages flagged by the model" are only P cases, you won't have your F cases reviewed by a human, therefore you won't have FN, therefore it's not D.

I also believe that 0.1% of raw messages is going to have too little P cases, therefore not C.

And then we remain with option B, which is not optimal, but it is the best we can do in this situation.

   upvoted 1 times

  **baimus** 1 month, 2 weeks ago

Selected Answer: C

It is absolutely not possible to calculate recall with D because we only have positives in the sample we need false negatives. Because of the high quantity of total data, 0.1% is fine, the answer is C

   upvoted 1 times

  **ludovikush** 7 months ago

Selected Answer: D

Precision and recall are critical metrics for evaluating the performance of classification models, especially in contexts where both the accuracy of positive predictions (precision) and the ability to identify all positive instances (recall) are important. In this case:

Precision (the proportion of messages flagged by the model as inappropriate that were actually inappropriate) helps ensure that the model minimizes the burden on human moderators by not flagging too many false positives, which could overwhelm them.

Recall (the proportion of actual inappropriate messages that were correctly flagged by the model) ensures that the model is effective at catching as many inappropriate messages as possible, reducing the risk of harmful content being missed.

   upvoted 3 times

  **etienne0** 7 months, 1 week ago

Selected Answer: C

I go with C

   upvoted 1 times

  **pmle_nintendo** 7 months, 3 weeks ago

Selected Answer: D

Let's consider below hypothetical scenario:

Total number of comments per minute: 10,000

Comments actually inappropriate: 500

If we use a random sample of only 0.1% of raw messages (10 comments) for evaluation, there's a high chance that this small sample may not include any or only a few inappropriate comments. As a result, the precision and recall estimates based on this sample may be skewed, leading to unreliable assessments of the model's performance. Thus, C is ruled out.


   upvoted 1 times

  **Werner123** 7 months, 3 weeks ago

Selected Answer: D

C does not make sense to me since it is a very small random sample. It is also only messages that have been sent to humans for review meaning that there is bias in that result set.

   upvoted 1 times

  **b1a8fae** 9 months, 3 weeks ago

D only caring for observations flagged by the model means we don't control for false negatives (approved actually inappropriate messages). B seems like a better option to me: the wording confuses me a bit, but I understand it as the true and false positives (human flagged comments and their modelled label)

   upvoted 1 times

  **Mickey321** 11 months, 1 week ago

Selected Answer: D

In favor of D

   upvoted 1 times

  **pico** 11 months, 1 week ago

Selected Answer: C

Given the context of content moderation, a balanced approach is often preferred. Therefore, option C, precision and recall estimates based on a random sample of raw messages, is a good choice. It provides a holistic view of the model's performance, taking into account both false positives (precision) and false negatives (recall), and it reflects how well the model is handling the entire dataset.

   upvoted 1 times

  **Krish6488** 11 months, 2 weeks ago

Selected Answer: D

A --> Conveys model's activity levels but not accuracy

B --> Accuracy to some extent but won't give full picture as it does not account False negatives

C --> Using a random sample of the raw messages allows you to estimate precision and recall for the overall activity, not just the flagged content.

D --> Specifically measures on the subset of data that it flagged

Both C & D work well in this case, but the specificity is higher in option D and hence will go with D

   upvoted 1 times

  **MultipleWorkerMirroredStrategy** 11 months, 4 weeks ago

Selected Answer: C

Google Cloud used to have a service called "continuous evaluation", where human labelers classify data to establish a ground truth. Thinking along those lines, the answer is C as it's the logical equivalent of that service.

<https://cloud.google.com/ai-platform/prediction/docs/continuous-evaluation>

   upvoted 1 times

  **PST21** 1 year, 3 months ago

Question is to measure model performance so has to be precision & recall, hence D.

   upvoted 2 times

  **Voyager2** 1 year, 4 months ago

Selected Answer: D

D. Precision and recall estimates based on a sample of messages flagged by the model as potentially inappropriate each minute

You will need precision and recall to identify false positives and false negatives. A very small random sample doesn't help specially because probably you will have skewed data. So D.

   upvoted 1 times

  **M25** 1 year, 5 months ago

Selected Answer: D

Went with D

   upvoted 1 times

  **lucaluca1982** 1 year, 6 months ago

Selected Answer: D

we need to monitor the model, so D

   upvoted 1 times

[Load full discussion...](#)

[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.