

1.

Pertanyaan #1

Problem Statement

This example is adapted from a real production application, but with details disguised to protect confidentiality.



You are a famous researcher in the City of Peacetopia. The people of Peacetopia have a common characteristic: they are afraid of birds. To save them, you have **to build an algorithm that will detect any bird flying over Peacetopia** and alert the population.

The City Council gives you a dataset of 10,000,000 images of the sky above Peacetopia, taken from the city's security cameras. They are labeled:

- $y = 0$: There is no bird on the image
- $y = 1$: There is a bird on the image

Your goal is to build an algorithm able to classify new images taken by security cameras from Peacetopia.

There are a lot of decisions to make:

- What is the evaluation metric?
- How do you structure your data into train/dev/test sets?

Metric of success

The City Council tells you the following that they want an algorithm that

1. Has high accuracy.
2. Runs quickly and takes only a short time to classify a new image.
3. Can fit in a small amount of memory, so that it can run in a small processor that the city will attach to many different security cameras.

You are delighted because this list of criteria will speed development and provide guidance on how to evaluate two different algorithms. True/False?

- ☐ False
- ☒ True:

 **Luaskan**

 **Salah**

Salah
No. The goal is to have one metric that focuses the development effort and increases iteration velocity.

2.

Pertanyaan #2


The city asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria?

2. The city asks for your help in further defining the criteria for accuracy, runtime, and memory. How would you suggest they identify the criteria?

1 / 1 poin

- ☐ Suggest that they purchase more infrastructure to ensure the model runs quickly and accurately.
- ☐ Suggest to them that they focus on whichever criterion is important and then eliminate the other two.
- ☒ Suggest to them that they define which criterion is most important. Then, set thresholds for the other two.

 **Luaskan**

 **Benar**
Yes. The thresholds provide a way to evaluate models head to head.

Benar
Yes. The thresholds provide a way to evaluate models head to head.

3.

Pertanyaan #3

Based on the city’s requests, which of the following would you say is true?

3. Based on the city’s requests, which of the following would you say is true?

1 / 1 poin

- ☐ Accuracy, running time and memory size are all satisfying metrics because you have to do sufficiently well on all three for your system to be acceptable.
- ☐ Accuracy is a satisfying metric; running time and memory size are an optimizing metric.
- ☐ Accuracy, running time and memory size are all optimizing metrics because you want to do well on all three.
- ☒ Accuracy is an optimizing metric; running time and memory size are satisfying metrics.

↗ Luaskan

✓ Benar

Benar

4.

Pertanyaan #4
Structuring your data

Before implementing your algorithm, you need to split your data into train/dev/test sets. Which of these do you think is the best choice?

4. Structuring your data

1 / 1 poin

Before implementing your algorithm, you need to split your data into train/dev/test sets. Which of these do you think is the best choice?

- ☒

Train	Dev	Test
9,500,000	250,000	250,000
- ☐

Train	Dev	Test
6,000,000	1,000,000	3,000,000
- ☐

Train	Dev	Test
3,333,334	3,333,334	3,333,334
- ☐

Train	Dev	Test
6,000,000	3,000,000	1,000,000

↗ Luaskan

✓ Benar
Yes.

Benar

Yes.

5.

Pertanyaan #5

Now that you’ve set up your train/dev/test sets, the City Council comes across another 1,000,000 images from social media and offers them to you. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. You should add the citizens’ data to the training set. True/False?

5. Now that you've set up your train/dev/test sets, the City Council comes across another 1,000,000 images from social media and offers them to you. These images are different from the distribution of images the City Council had originally given you, but you think it could help your algorithm. You should add the citizens' data to the training set. True/False?

1 / 1 poin

- ☐ False
- ☒ True

 Luaskan

 Benar

Yes. This will cause the training and dev/test set distributions to become different, however as long as dev/test distributions are the same you are aiming at the same target.

Benar

Yes. This will cause the training and dev/test set distributions to become different, however as long as dev/test distributions are the same you are aiming at the same target.

6.

Pertanyaan #6

One member of the City Council knows a little about machine learning and thinks you should add the 1,000,000 citizens' data images proportionately to the train/dev/test sets. You object because:

6. One member of the City Council knows a little about machine learning and thinks you should add the 1,000,000 citizens' data images proportionately to the train/dev/test sets. You object because:

1 / 1 poin

- ☒ If we add the images to the test set then it won't reflect the distribution of data expected in production.
- ☐ The 1,000,000 citizens' data images do not have a consistent x-->y mapping as the rest of the data.
- ☐ The training set will not be as accurate because of the different distributions.
- ☐ The additional data would significantly slow down training time.

 Luaskan

 Benar

Yes. Using the data in the training set could be beneficial, but you wouldn't want to include such images in your test set as they are not from the expected distribution of data you'll see in production.

Benar

Yes. Using the data in the training set could be beneficial, but you wouldn't want to include such images in your test set as they are not from the expected distribution of data you'll see in production.

7.

Pertanyaan #7

You train a system, and its errors are as follows (error = 100%-Accuracy):

Training set error	4.0%
Dev set error	4.5%

This suggests that one good avenue for improving performance is to train a bigger network so as to drive down the 4.0% training error. Do you agree?

7. You train a system, and its errors are as follows (error = 100%-Accuracy):

1 / 1 poin

Training set error	4.0%
Dev set error	4.5%

This suggests that one good avenue for improving performance is to train a bigger network so as to drive down the 4.0% training error. Do you agree?

- ☒ No, because there is insufficient information to tell.
- ☐ No, because this shows your variance is higher than your bias.
- ☐ Yes, because this shows your bias is higher than your variance.
- ☐ Yes, because having a 4.0% training error shows you have a high bias.

 Luaskan

 Benar

Benar

8.

Pertanyaan #8


You want to define what human-level performance is to the city council. Which of the following is the best answer?

8. You want to define what human-level performance is to the city council. Which of the following is the best answer?

1 / 1 poin

- ☐ The average performance of all their ornithologists (0.5%).
- ☒ The performance of their best ornithologist (0.3%).
- ☐ The average of all the numbers above (0.66%).
- ☐ The average of regular citizens of Peacetopia (1.2%).

 Luaskan

 Benar
Yes. The best human performance is closest to Bayes' error.

Benar

Yes. The best human performance is closest to Bayes' error.

9.

Pertanyaan #9

Which of the following statements do you agree with?

9. Which of the following statements do you agree with?

1 / 1 poin

- ☐ A learning algorithm's performance can be better than human-level performance and better than Bayes error.
- ☐ A learning algorithm's performance can never be better than human-level performance nor better than Bayes error.
- ☐ A learning algorithm's performance can never be better than human-level performance but it can be better than Bayes error.
- ☒ A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error.

 Luaskan

 Benar

Benar

10.

Pertanyaan #10

After working on your algorithm you have to decide the next steps. Currently, human-level performance is 0.1%, training is at 2.0% and the dev set is at 2.1%. Which statement below best describes your thought process?

10. After working on your algorithm you have to decide the next steps. Currently, human-level performance is 0.1%, training is at 2.0% and the dev set is at 2.1%. Which statement below best describes your thought process?

1 / 1 poin

- ☒ Decrease regularization to boost smaller signals.
- ☐ Decrease variance via regularization so training and dev sets have similar performance.
- ☐ Get a bigger training set to reduce variance.
- ☒ Address bias first through a larger model to get closest to human level error.

 Correct

Yes. Bias is higher than variance.

 Correct

Yes. Selecting the largest difference from (train set error - human level error) and (dev set error - train set error) and reducing bias or variance accordingly is the most productive step.

 Luaskan

 Benar

Great, you got all the right answers.

Benar

Great, you got all the right answers.

11.

Pertanyaan #11

After running your model with the test set you find it is a 7.0% error compared to a 2.1% error for the dev set and 2.0% for the training set. What can you conclude? (Choose all that apply)

11. After running your model with the test set you find it is a 7.0% error compared to a 2.1% error for the dev set and 2.0% for the training set. What can you conclude? (Choose all that apply)

1 / 1 poin

- ☐ You have underfitted to the dev set.
- ☒ You have overfitted to the dev set.

✓

Correct

Yes. The dev set performance versus the test set indicates it is overfitting.

- ☒ You should try to get a bigger dev set.

✓

Correct

Yes. The dev set performance versus the test set indicates it is overfitting.

- ☐ Try decreasing regularization for better generalization with the dev set.

↶ ↷

Luaskan

✓

Benar

Great, you got all the right answers.

Benar
Great, you got all the right answers.

12.
Pertanyaan #12
After working on this project for a year, you finally achieve:

Human-level performance	0.10%
Training set error	0.05%
Dev set error	0.05%

What can you conclude? (Check all that apply.)

12. After working on this project for a year, you finally achieve:

1 / 1 poin

Human-level performance	0.10%
Training set error	0.05%
Dev set error	0.05%

What can you conclude? (Check all that apply.)

- ☒ It is now harder to measure avoidable bias, thus progress will be slower going forward.

✓

Correct

- ☒ If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is ≤ 0.05

✓

Correct

- ☐ With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- ☐ This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.

↶ ↷

Luaskan

✓

Benar

Great, you got all the right answers.

Benar
Great, you got all the right answers.

13.

Pertanyaan #13

It turns out Peacetopia has hired one of your competitors to build a system as well. You and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! Still, when Peacetopia tries out both systems, they conclude they like your competitor’s system better because, even though you have higher overall accuracy, you have more false negatives (failing to raise an alarm when a bird is in the air). What should you do?

13. It turns out Peacetopia has hired one of your competitors to build a system as well. You and your competitor both deliver systems with about the same running time and memory size. However, your system has higher accuracy! Still, when Peacetopia tries out both systems, they conclude they like your competitor’s system better because, even though you have higher overall accuracy, you have more false negatives (failing to raise an alarm when a bird is in the air). What should you do?

1 / 1 poin

- ☒ Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.
- ☐ Apply regularization to minimize the false negative rate.
- ☐ Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ☐ Ask your team to take into account both accuracy and false negative rate during development.

[Luaskan](#)



Benar

Yes. The target has shifted so an updated metric is required.

Benar

Yes. The target has shifted so an updated metric is required.

14.

Pertanyaan #14

You’ve handily beaten your competitor, and your system is now deployed in Peacetopia and is protecting the citizens from birds! But over the last few months, a new species of bird has been slowly migrating into the area, so the performance of your system slowly degrades because your data is being tested on a new type of data.



You have only 1,000 images of the new species of bird. The city expects a better system from you within the next 3 months. Which of these should you do first?

14. You've handily beaten your competitor, and your system is now deployed in Peacetopia and is protecting the citizens from birds! But over the last few months, a new species of bird has been slowly migrating into the area, so the performance of your system slowly degrades because your data is being tested on a new type of data.

1 / 1 poin



You have only 1,000 images of the new species of bird. The city expects a better system from you within the next 3 months. Which of these should you do first?

- ☐ Try data augmentation/data synthesis to get more images of the new type of bird.
- ☐ Put the 1,000 images into the training set so as to try to do better on these birds.
- ☐ Add the 1,000 images into your dataset and reshuffle into a new train/dev/test split.
- ☒ Use the data you have to define a new evaluation metric (using a new dev/test set) taking into account the new species, and use that to drive further progress for your team.

[Luaskan](#)

✓ Benar

Benar

15.

Pertanyaan #15

The City Council thinks that having more Cats in the city would help scare off birds. They are so happy with your work on the Bird detector that they also hire you to build a Cat detector. (Wow Cat detectors are just incredibly useful, aren't they?) Because of years of working on Cat detectors, you have such a huge dataset of 100,000,000 cat images that training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.)

15. The City Council thinks that having more Cats in the city would help scare off birds. They are so happy with your work on the Bird detector that they also hire you to build a Cat detector. (Wow Cat detectors are just incredibly useful, aren't they?) Because of years of working on Cat detectors, you have such a huge dataset of 100,000,000 cat images that training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.)

1 / 1 poin

☒ Needing two weeks to train will limit the speed at which you can iterate.

✓ Correct

☒ If 100,000,000 examples is enough to build a good enough Cat detector, you might be better off training with just 10,000,000 examples to gain a $\approx 10\times$ improvement in how quickly you can run experiments, even if each model performs a bit worse because it's trained on less data.

✓ Correct

☒ Buying faster computers could speed up your teams' iteration speed and thus your team's productivity.

✓ Correct

☐ Having built a good Bird detector, you should be able to take the same model and hyperparameters and just apply it to the Cat dataset, so there is no need to iterate.

↗ Luaskan

✓ Benar

Great, you got all the right answers.

Benar

Great, you got all the right answers.