

## [Skip to Main Content](#)



SEARCH IN COURSE

Search

- - Mohammad Azri Harahap
- 1. [Structuring Machine Learning Projects](#)
- 2. [Module 1](#)
- 3. Bird Recognition in the City of Peacetopia (Case Study)

### PreviousNext

- Introduction to ML Strategy
- Connect with your Mentors and Fellow Learners on Discourse!
- Setting Up your Goal
- Comparing to Human-level Performance
- Lecture Notes (Optional)
- Machine Learning Flight Simulator



**Reading:** [ReadingMachine Learning Flight Simulator](#)

. Duration: 2 minutes2 min



**Quiz:** [Bird Recognition in the City of Peacetopia \(Case Study\)](#)

15 questions

- Heroes of Deep Learning (Optional)

# Bird Recognition in the City of Peacetopia (Case Study)

Quiz 1 hour 15 minutes • 1h 15m

**Submit your assignment**

Due May 21, 11:59 PM WIB May 21, 11:59 PM WIB

Attempts 3 every 24 hours

Try again

**Receive grade**

To Pass 80% or higher

**Your grade**

53.33%

View Feedback

We keep your highest score

Like

Dislike

Report an issue

[Back](#)

# Bird Recognition in the City of Peacetopia (Case Study)

Graded Quiz. • 1h 15m

Due May 21, 11:59 PM WIB

**Try again once you are ready**

Grade received 53.33%

Latest Submission Grade 53.33%

To pass 80% or higher

Try again

**1.**

Question 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 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.

Note: Having three evaluation metrics makes it harder for you to quickly choose between two different algorithms, and will slow down the speed with which your team can iterate. True/False?



True



False

1 / 1 point

Expand

Correct

2.

Question 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?

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

0 / 1 point

Expand

Incorrect

No. This violates their use case because it is possible that the model won't fit on the target devices, or it may not provide alerts in a timely manner.

3.

Question 3

The essential difference between an optimizing metric and satisficing metrics is the priority assigned by the stakeholders. True/False?

- ☒ True
- ☐ False

0 / 1 point

Expand

Incorrect

No. Stakeholders must define thresholds for satisficing metrics, leaving the optimizing metric unbounded.

4.

Question 4

You propose a 95/2.5%/2.5% for train/dev/test splits to the City Council. They ask for your reasoning. Which of the following best justifies your proposal?

- ☒ With a dataset comprising 10M individual samples, 2.5% represents 250k samples, which should be more than enough for dev and testing to evaluate bias and variance.
- ☐ The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics.
- ☐ The emphasis on the training set will allow us to iterate faster.

- ☐ The most important goal is achieving the highest accuracy, and that can be done by allocating the maximum amount of data to the training set.

1 / 1 point

Expand

**Correct**

Yes. The purpose of dev and test sets is fulfilled even with smaller percentages of the data.

5.

Question 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?



True



False

0 / 1 point

Expand

**Incorrect**

No. Adding this data to the training set will change the training set distribution. However, it is not a problem to have different training and dev distributions. In contrast, it would be very problematic to have different dev and test set distributions.

6.

Question 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:

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

1 / 1 point

Expand

**Correct**

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.

Question 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?

☐

No, because there is insufficient information to tell.

☐

No, because this shows your variance is higher than your bias.

☒

Yes, because having a 4.0% training error shows you have a high bias.

☐

Yes, because this shows your bias is higher than your variance.

0 / 1 point

Expand

Incorrect

8.

Question 8

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

☐

The performance of their best ornithologist (0.3%).

☐

The average of all the numbers above (0.66%).

☒

The average performance of all their ornithologists (0.5%).

☐

The average of regular citizens of Peacetopia (1.2%).

0 / 1 point

Expand

Incorrect

No. The average reflects a range of skills, not the best.

9.

Question 9

A learning algorithm's performance can be better than human-level performance but it can never be better than Bayes error. True/False?

☐

True.

☒

False.

0 / 1 point

Expand

**Incorrect**

No. By definition, human level error is worse than Bayes error.

**10.**

Question 10

Which of the following best expresses how to evaluate the next steps in your project when your results for human-level performance, train, and dev set error are 0.1%, 2.0%, and 2.1% respectively?

- ☐ Evaluate the test set to determine the magnitude of the variance.
- ☐ Keep tuning until the train set accuracy is equal to human-level performance because it is the optimizing metric.
- ☒ Based on differences between the three levels of performance, prioritize actions to decrease bias and iterate.
- ☐ Port the code to the target devices to evaluate if your model meets or exceeds the satisficing metrics.

**1 / 1 point**

Expand

**Correct**

Yes. Always choose the area with the biggest opportunity for improvement.

**11.**

Question 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)

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

**Correct**

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

- ☐ You have underfitted to the dev set.

**1 / 1 point**

Expand

**Correct**

Great, you got all the right answers.

**12.**

Question 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%. Which of the following are true? (Check all that apply.)

- ☐ 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.

☒ All or almost all of the avoidable bias has been accounted for.

**Correct**

Yes. Exceeding human performance makes the identification of avoidable bias very challenging.

☒ You are close to Bayes error and possible overfitting.

**Correct**

Yes. By definition, Bayes error cannot be exceeded except for overfitting.

**1 / 1 point**

Expand

**Correct**

Great, you got all the right answers.

**13.**

Question 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?

☐ Apply regularization to minimize the false negative rate.

☐ Brainstorm with your team to refine the optimizing metric to include false negatives as they further develop the model.

☒ 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.

**0 / 1 point**

Expand

**Incorrect**

No. This choice also points to the incorrect target.

**14.**

Question 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?

- ☐ Put the 1,000 images into the training set so as to try to do better on these birds.
- ☐ Try data augmentation/data synthesis to get more images of the new type of bird.
- ☐ 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.

1 / 1 point

Expand

Correct

15.

Question 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. You have a huge dataset of 100,000,000 cat images. Training on this data takes about two weeks. Which of the statements do you agree with? (Check all that agree.)

- ☒ Lowering the number of images will reduce training time and likely allow for an acceptable tradeoff between iteration speed and accuracy.

Correct

Yes. There is a sweet spot that allows development at a reasonable rate without significant accuracy loss.

- ☐ Reducing the model complexity will allow the use of the larger data set but preserve accuracy.

☒ This significantly impacts iteration speed.

**Correct**

Yes. This training time is an absolute constraint on iteration.

**1 / 1 point**

Expand

**Correct**

Great, you got all the right answers.