

## [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

Retake the quiz in 22h 10m

**Receive grade**

To Pass 80% or higher

**Your grade**

73.33%

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 73.33%

Latest Submission Grade 33.33%

To pass 80% or higher

Retake the assignment in 22h 10m

1.

Question 1

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 meet with them and ask for just one evaluation metric. True/False?

☒ True:

☐ False

0 / 1 point

Expand

Incorrect

You answered when the time elapsed.

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

1 / 1 point

Expand

**Correct**

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

3.

Question 3

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

- ☐ False
- ☒ True

0 / 1 point

Expand

**Incorrect**

You answered when the time elapsed.

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?

- ☐ The emphasis on the training set provides the most accurate model, supporting the memory and processing satisficing metrics
- ☐ 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.
- ☒ 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 will allow us to iterate faster.

0 / 1 point

Expand

**Incorrect**

You answered when the time elapsed.

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. Which of the following is the best use of that additional data?



Add it to the dev set to evaluate how well the model generalizes across a broader set.



Add it to the training set.



Split it among train/dev/test equally.



Do not use the data. It will change the distribution of any set it is added to.

1 / 1 point

Expand

Correct

Yes. It is not a problem to have different training and dev distributions. Different dev and test distributions would be an issue.

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:



The 1,000,000 citizens' data images do not have a consistent  $x \rightarrow y$  mapping as the rest of the data.



If we add the images to the test set then it won't reflect the distribution of data expected in production.



The additional data would significantly slow down training time



The training set will not be as accurate because of the different distributions.

0 / 1 point

Expand

Incorrect

You answered when the time elapsed.

7.

Question 7

Human performance for identifying birds is  $< 1\%$ , training set error is  $5.2\%$  and dev set error is  $7.3\%$ . Which of the options below is the best next step?



Validate the human data set with a sample of your data to ensure the images are of sufficient quality.



Train a bigger network to drive down the  $> 4.0\%$  training error.



Get more data or apply regularization to reduce variance.



Try an ensemble model to reduce bias and variance.

0 / 1 point

Expand

**Incorrect**

No. We do not have test set accuracy.

8.

Question 8

If your goal is to have “human-level performance” be a proxy (or estimate) for Bayes error, how would you define “human-level performance”?



The best performance of a specialist (ornithologist) or possibly a group of specialists.



The performance of their volunteer amateur ornithologists.



The performance of the head of the City Council.



The performance of the average citizen of Peacetopia.

1 / 1 point

Expand

**Correct**

Yes. This is the peak of human performance in this task.

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?

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

0 / 1 point

Expand

**Incorrect**

You answered when the time elapsed.

**11.**

Question 11

You've now also run your model on the test set and find that it is a 7.0% error compared to a 2.1% error for the dev set. What should you do? (Choose all that apply)

- ☐ Get a bigger test set to increase its accuracy.
- ☒ Try increasing regularization to reduce overfitting to the dev set.

**Correct**

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

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

**This should not be selected**

No. Decreasing regularization will increase overfitting.

0 / 1 point

Expand

**Incorrect**

You didn't select all the correct 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%

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

- ☐ With only 0.05% further progress to make, you should quickly be able to close the remaining gap to 0%
- ☒ If the test set is big enough for the 0.05% error estimate to be accurate, this implies Bayes error is  $\leq 0.05 \leq 0.05$

**Correct**

- ☐ This is a statistical anomaly (or must be the result of statistical noise) since it should not be possible to surpass human-level performance.
- ☒ It is now harder to measure avoidable bias, thus progress will be slower going forward.

**Correct**

**1 / 1 point**

Expand

**Correct**

Great, you got all the right answers.

**13.**

Question 13

Your system is now very accurate but has a higher false negative rate than the City Council of Peacetopia would like. What is your best next step?

- ☐ Look at all the models you've developed during the development process and find the one with the lowest false negative error rate.
- ☒ Pick false negative rate as the new metric, and use this new metric to drive all further development.
- ☐ Reset your "target" (metric) for the team and tune to it.
- ☐ Expand your model size to account for more corner cases.

**0 / 1 point**

Expand

**Incorrect**

No. This choice also points to the incorrect target.

**14.**

Question 14

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

- ☐ Split them between dev and test and re-tune.
- ☐ Add pooling layers to downsample features to accommodate the new species.
- ☒ Augment your data to increase the images of the new bird.
- ☐ Put the new species' images in training data to learn their features.

**1 / 1 point**

Expand

**Correct**

Yes. A sufficient number of images is necessary to account for the new species.



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.)

- ☒ You could consider a tradeoff where you use a subset of the cat data to find reasonable performance with reasonable iteration pacing.

**Correct**

Yes. This is similar to satisficing metrics where "good enough" determines the size of the data.

- ☐ Given a significant budget for cloud GPUs, you could mitigate the training time.
- ☐ With the experience gained from the Bird detector you are confident to build a good Cat detector on the first try.

**This should not be selected**

No. Although you may have gained many insights that may reduce the number of iterations needed, a DeepLearning model requires multiple iterations when working in a new dataset.

- ☒ Accuracy should exceed the City Council's requirements but the project may take as long as the bird detector because of the two week training/iteration time.

**Correct**

Yes. The 10x size increase adds a small amount of accuracy but takes too much time.

0 / 1 point

Expand

**Incorrect**

You didn't select all the correct answers