# Production ML Systems: Test Your Knowledge

Let's do a quick test! You must answer at least 4 questions correctly to pass this quiz.

1. You are using machine learning to build a classification model that predicts unicorn appearances. Your dataset details 10,000 unicorn appearances and 10,000 unicorn non-appearances. The dataset contains the location, time of day, elevation, temperature, humidity, tree cover, presence of a rainbow, and several other features.

After launching your unicorn appearance predictor, you will need to keep your model fresh by retraining on new data. Because you are gathering too much new data to train on, you decide to limit the training data by sampling the new data over a window of time. You also need to account for daily and annual patterns in unicorn appearances. What window of time do you choose?

- ○ One day, because a larger window would result in lots of data and your model would take too long to train.
- ○ One week, so that your dataset is not too large but you can still smooth out patterns.
- ◉ One year, to ensure that your model is not biased by yearly patterns.

  ✓ Correct! You should choose a representative dataset so that your model learns to predict across all scenarios.

2. You launch your unicorn appearance predictor. It's working well! You go on vacation and return after three weeks to find that your model quality has dropped significantly. Assume that unicorn behavior is unlikely to change significantly in three weeks. What is the most likely explanation for the decrease in quality?

- ◉ Training-serving skew: the format of the serving data gradually changed at some point after the model started serving.

  ✓ While unicorn behavior probably didn't change, perhaps the underlying data reporting or data formatting changed in the serving data after the training data was collected. You can detect potential training-serving skew by checking the serving data against the data schema of the training data.
- ○ You used accuracy as a metric during training.
- ○ Your model is stale.
- ○ None of the above.

3. You review the model's predictions for Antarctica, and discover the model has been making poor predictions there since the model was released into production. Which of the following could be the source of the problem?

- ◉ You didn't have enough training examples for Antarctica.

  ✓ Because your model didn't include sufficient training examples for Antarctica, it didn't learn the unique patterns of unicorn behavior on the continent, and was unable to make accurate predictions there.
- ○ You used dynamic training instead of static training.
- ○ Your model has become stale.
- ○ All of the above.

4. Your unicorn appearance predictor has operated for a year. You've fixed many problems, and quality is now high. However, you notice a small but persistent problem. Your model quality has drifted slightly lower in urban areas. What might be the cause?

- ◉ The high quality of your predictions lead users to easily find unicorns, affecting unicorn appearance behavior itself.

  ✓ Unicorns responded to increased attention by changing their behavior in urban areas. As your model's predictions adapt to the changing behavior, unicorns continue to change their behavior. Such a situation, where your model's behavior affects the training data itself, is called a feedback loop. You should try modifying your training-serving skew detection to detect changes in serving data that correspond to changes in unicorn behavior.
- ○ Urban areas are difficult to model.
- ○ Unicorn appearances are reported multiple times in heavily populated areas, skewing your training data.

5. Through all your troubleshooting, you've greatly improved the quality of the unicorn model's predictions, and as a result, usage has increased tenfold. However, users are now complaining that the model is extremely slow; inference requests typically take more than 30 seconds to return predictions. Which of the following changes could help solve this problem?

- ○ Switch the model from dynamic training to static training.
- ◉ Switch the model from dynamic inference to static inference.

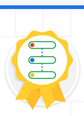
  ✓ Switching to static inference means the model's predictions will be cached, so they can be served to users with very low latency. However, a model that uses static inference can only return these cached predictions, so it may not be able to provide predictions for uncommon inputs it hasn't seen before.
- ○ Validate the model quality before serving.
- ○ None of the above solutions would help.

## Results

You scored 5 out of 5. Congratulations! You have passed this quiz.

You've unlocked a new milestone:
**Machine Learning Crash Course: Production ML systems!**

Sign in to claim your badge.



Share

Sign in

Was this helpful?