Congratulations! You passed!

Grade received 100% **To pass** 80% or higher

Go to next item

1. Suppose that you have a very accurate model for a social app that uses several features to predict whether a user is a spammer or not. You trained the model with a particular idea of what a spammer was, for example, a user who sends ten messages in one minute. Over time, the app grew and became more popular, but the outcome of the predictions has drastically changed. As people are chatting and messaging more, now sending ten messages in a minute becomes normal and not something that only spammers do. What kind of drift causes this spam-detection model's predictive ability to decay?

1 / 1 point

- ○ Data Drift
- ◉ Concept Drift

Correct

Nailed it! The model's original idea about what it means to be a spammer has changed, and now sending ten messages in a minute becomes normal. In other words, the concept of spammers has drifted. Consequently, since you haven't updated your model, it will predict these non-spammers as spammers (a false positive).

2. When mitigating data drift, if you don't have enough recently labeled data to create an entirely new dataset, how can you determine what data from your previous training dataset is still valid? (Select all that apply)

1 / 1 point

- [x] Using a statistical measure like the Kullback–Leibler divergence.

Correct
Nice job! Kullback–Leibler divergence, also called relative entropy, measures how much the probability distribution has drifted from the original position.

- [ ] Using a human Data-Labeling service.
- [x] Using an unsupervised learning technique like clustering.

Correct
That's right! Clustering identifies relevant portions of the original dataset to retrain a model's functionality that has become dispersed. It is a form of restructuring, and hence it could be a way of direct preventative maintenance.

3. True or False: Having restructured a new training dataset after detecting model decay, you must start over by resetting and completely retraining your model. There is no way around it.

1 / 1 point

- ○ True
- ◉ False

Correct
Exactly! You have two choices for how to train your model. Using your new data, you can either start over or continue training your model and fine-tuning it from the last checkpoint.

4. If you can't automate the process of finding out when model retraining is required, what are some good ideas for policies to establish when to do it? (Select all that apply)

1 / 1 point

- [x] Retrain on schedule.

Correct
Keep it up! You could also just retrain your model according to a schedule. In practice, this is what many people do because it's simple to understand and works reasonably well in many domains. You should be careful, though, not to retrain the model too often or not often enough.

- [x] Retrain on new data availability.

Correct
You've got it! When the availability of new training data limits you, you may be forced to retain as much of your old training data for as long as possible and avoid fully retraining your model.

- [x] Retrain on demand.

Correct

Way to go! You can retrain your model whenever it seems necessary, including situations like noticing drift or needing to change (add or remove) class labels or features, for example.

- [ ] Retrain on performance-degradation detection.