

Lesson 19: Troubleshooting Models

45 minutes

Overview

This is the fourth of a five-day sequence of lessons that prepare students for the final project. In this lesson, students examine survey data from other members of the student team and analyze why their models are not working correctly. In examining the data, students develop strategies for avoiding these issues in the future and strategies for coping with these issues should they happen again. These are skills students will use in the final project as they develop their own surveys and collect data.

Question of the Day: What are strategies to make sure our data generates an accurate model?

Standards

Full Course Alignment

AI4K12 National Guidelines 2021

- ▶ **BI-3** - Computers can learn from data

CSTA K-12 Computer Science Standards (2017)

- ▶ **AP** - Algorithms & Programming
- ▶ **IC** - Impacts of Computing

Agenda

Warm Up (5 minutes)

Journal

Activity (35 minutes)

Wrap Up (5 minutes)

Journal

Objectives

Students will be able to:

- Explain how choices in data collection can lead to issues when training a machine learning model

Preparation

- Check the "**Teacher's Lounge**" forum for verified teachers to find additional strategies or resources shared by fellow teachers

Links

Heads Up! Please make a copy of any documents you plan to share with students.

For the teachers

- **Troubleshooting Models** - Slides

▼ Make a Copy

For the students

- **Troubleshooting Models - Activity**

Guide ▼ Make a Copy

Teaching Guide

Warm Up (5 minutes)

Journal

 **Display:** Show the slide that shows Zoey explaining that they've tried training their model, but they can never get higher than a 60% accurate model.

 **Prompt:** *What are some reasons Zoey might not be able to train an accurate model?*

Discussion Goal: This prompt prepares students for today's main activity - strategies for what to do when your data doesn't lead to an accurate model. Students should draw on their experiences from the unit when they've trained their own models. Some answers may include:

- Zoey hasn't chosen enough features - if they select more features, it might become more accurate
- Zoey is selecting features that aren't strongly related to their label. They should use the data visualizations to get a better idea of what features will be best for the label.
- Zoey may not have enough data - if they collect more data, they might get better results.

Remarks

Even after creating our survey and gathering our data, it's still possible that we might not be able to train an accurate model from our data. There are a lot of reasons this could happen - today, we're going to talk about reasons this might happen and some strategies for how to adapt if our data doesn't lead to good results.

 **Question of the Day:** What are strategies to make sure our data generates an accurate model?

Activity (35 minutes)

 **Code Studio:** Have students log in to Code Studio and go to the first level of this lesson. Students will see several datasets they can choose from.

 1

Club Datasets

Distribute: Pass out the **Troubleshooting Models - Activity Guide**. Students will record data from their investigation here.

 Assessment Opportunity 

The responses in the activity guide can be used to formatively assess how well students have met the objectives of the lesson.

Display: Show students the slide explaining the first part of today's task - Kim, Nico, Isaac, and Zoey are having trouble training accurate models and the class needs to help troubleshoot. Students will investigate each student and try to describe what is causing the issue, and a recommendation for how to solve it or how to avoid it next time.

 **Do This:** Investigate each person's data in AI Lab and record your findings on your activity guide.

Circulate: Monitor students as they complete this task. Students should read the description from each student for clues as to why their data may not be working, then explore the data and results screens. Students should experiment with different features and testing the models to see where the issues may be.

 Teaching Tip 

Troubleshooting Models: Here are some tips to help students discover the issues with these models:

- **Zoey's Data:** Zoey's models will always have a 0% or 33% accuracy because she didn't collect enough data - students should notice the dataset only has 25 rows. Students might suggest that Zoey collects more data before trying to train her model again.

- **Isaac's Data:** Isaac's models will always be low because none of his features have a strong relationship to his label - students should notice that none of the Cross Tab charts have strong hot spots with any question. If students look at the questions themselves, they may notice that they don't seem to have a strong relationship to the label Isaac is trying to predict. Students may recommend Isaac choose different questions and collecting his data again.
- **Kim's Data:** Kim's models can be very accurate, but when students test the model they may find the questions and answer choices very specific and not inclusive - for example, the "favorite lunch food" question only contains meat options. Students may suggest that, even though the model is accurate, Kim may want to choose different questions that are more broad and inclusive rather than so specific.
- **Nico's Data:** Nico's models can be very accurate, but when students test the model they may find that the majority of the recommendation is for the "Band Club" no matter what they pick. Students may need some prompting to notice this, but the data selection screen shows that Nico collected data primarily from people in a band, which biased his model towards recommending the Band club. Students may recommend Nico can avoid this next time by making sure he has a wide variety of people complete his survey.

 **Share out:** Invite students to share their responses. As they do, you may decide to pull up the data in front of the class and emphasize what students discovered about the model.

 **Prompt:** *Hawa hasn't started collecting data yet. What are three strategies you would offer her so she doesn't make some of the mistakes as her peers?*

Have students record their answers on their activity guide. They can also share with a partner to help generate more ideas.

Discussion Goal: Students should make suggestions related to gathering data from diverse sources, making sure she has enough data, and making sure the survey is representative and general enough to get interesting results.

Teaching Tip

Survey Planning: This discussion and focus on planning a survey is similar to part of the project that students will complete next week - they will need to plan how they will collect data to ensure their data can generate an accurate model. It may be helpful to cue students to remember this when they start their own project next week.

Display: Show the next slide, which has the following text: *Isaac is worried he has time to generate a whole new survey and ask all new people, but he still wants to make an app. What are some other options he can consider?**

Have students consider this individually and share with a partner, then continue to the next slide.

Display: Show the next slide, which shows Isaac referring back to his planning guide and remembering a different idea he had for the project - an isolation predictor.

Remarks

Since Isaac's initial plan didn't work out, one strategy he could try is to look back at his initial brainstorm ideas and see if there are any other apps he might be able to make that could still address this issue. Even though Isaac may not be able to collect his own data, he may be able to use someone else's data to help solve his problem.

 **Code Studio:** Have students continue to the next level in Code Studio, where they will see several datasets available to them.

Do This: Isaac has found some additional student survey datasets from a public website that he thinks he can use to create the Loneliness Score app from his brainstorm. Isaac has already selected "Loneliness" as the label, but each dataset has different features he could choose from (like music or movies). Choose one of the datasets to investigate and see if you can create an accurate model to predict how lonely someone might be.

Circulate: Monitor students as they look through the datasets. They may try different datasets depending on their interests.

Teaching Tip

Using Stock Datasets: This activity represents a situation students may find themselves in during the project next week: they may want to use one of the example datasets to investigate their issue, especially if they aren't able to collect enough data from their survey or their survey doesn't lead to an accurate model. In this situation, finding an alternative dataset is a viable option for students as long as they document this choice in their model card.

Remarks

Even when we don't get an accurate model, this doesn't mean it's the end of the world. Sometimes we can find an alternative dataset to use. Or - remember that we can think of ourselves as scientists: we're experimenting and investigating how our data is related to help make a model. And, just like scientists, sometimes an experiment fails - and that's okay! Sometimes talking about why the experiment failed is just as important as talking about why an experiment succeeded! Let's see how Isaac could do this with a Model Card from his original data.

Wrap Up (5 minutes)

Journal

Prompt: *If Isaac were to still use his data to create an app or use this example dataset, what do you think he should put in the Intended Uses section? What should go in the Limitations section?*

Discussion Goal: Students should realize that when using these alternative datasets, they should still be careful about documenting their decisions in their model cards.



This work is available under a **Creative Commons License (CC BY-NC-SA 4.0)**.

If you are interested in licensing Code.org materials for commercial purposes **[contact us](#)**.