Lesson 7: Algorithmic Bias

Overview

In this lesson, students will practice cropping images to uncover the bias underlying the Twitter cropping algorithm. Then, students will read and watch a video about the discovery of this biased algorithm. Finally, students will discuss ways to recognize and reduce bias along with analyzing Twitter's response to the allegations of bias in their cropping algorithm.

Standards

Full Course Alignment

CSP Conceptual Framework

▶ **IOC-1** - While computing innovations are typically designed to achieve a specific purpose, they may have unintended consequences.

Agenda

Warm Up (5 minutes)

Activity (30 minutes)

Cropping Widget (10 minutes)

Twitter Cropping Algorithm (20 minutes)

Wrap Up (10 minutes)

Objectives

Students will be able to:

- Reason about which types of tasks are should not be completed by an algorithm

Preparation

- Complete the cropping widget activity yourself so you can see what your students will experience with the widget will be and what questions or issues they may bring up.
- Watch the parts of the video **"Are We Automating Racism?"** referenced in the lesson plan.
- Read the two articles from the links section below.

Links

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

For the teachers

- **CSP Unit 9 - Data** - Slides
- **Sharing learnings about our image cropping algorithm (Twitter)** - Resource

For the students

- **Twitter says its image-cropping algorithm was biased, so it's ditching it (CNN)** - Resource (**Download**)

Teaching Guide

Warm Up (5 minutes)

Discuss: *Think about a time when you shared a picture with your friends or on social media. Have you ever had to crop a photo you shared? How did you decide what to crop out of the picture?*

Discussion Goal: Use this discussion to show that users usually have a reason for sharing photos and the focus depends on that reason. Students will most likely bring up a wide range of examples that can be used to highlight how there isn't a one size fits all approach to sharing photos and cropping them.

Remarks

> Today we are going to try our own hands at cropping and then closely examine a real-world situation where training a model to crop photos leads to algorithmic bias.

Activity (30 minutes)

Cropping Widget (10 minutes)

Do This: Tell students to use the widget to crop the images as if they were uploading them to a social media account. They can also upload their own images and crop them as well.

Pair students up for the discussion. Have them go through the images together and discuss their responses to the questions.

Discuss: Which images were challenging for you to decide how to crop? Which images did you and your partner crop the same? Which images did you and your partner crop differently?

Discussion Goal: Use this discussion to highlight the different experiences and perspectives everyone brings to deciding how to crop an image. Direct students to consider not only what they decided to keep in the picture, but also what they decided to crop out of the picture. What happens when we crop something out of a picture? It never gets seen by the audience. Look back at the photos you cropped. What or who got left out? You may choose to guide the class toward describing general characteristics of images that were challenging to crop.

Remarks

> Now imagine we recorded how everyone in this room cropped the images from this widget. We could use that information to train a model, just like we saw in AI for Oceans. This model could be used to crop other photos being uploaded to the social media site, just like the model we trained yesterday was used to identify certain types of fish and sea creatures.

Discuss: How might the cropping data from our classroom be biased? What are some ways we could address the biases in our data?

Discussion Goal: Help students see if this were indeed training a model, bias could arise from a few different places.

- The limited set of pictures included in the app that are most likely not representative of the photos being shared on social media.
- The bias of the students themselves. Are teenagers more likely to focus on certain things when cropping? Is this representative of the whole population of social media users?

1 **Cropping Widget**

Teaching Tip

When discussing bias, connect back to the training students did in the previous lesson when deciding what an “angry” or “fun” fish looks like, pushing students to consider how their opinions might lead to problems when training an AI model that would be widely used.

Twitter Cropping Algorithm (20 minutes)

Remarks

> It turns out that the issue of how to crop an image is something social media platforms have been working on for some time. We are going to watch a video that shows the issues that arose when Twitter used machine learning to train an algorithm to do this cropping.

Do This: Watch the video **“Are We Automating Racism?”** Watch from 0:00-4:05 and 9:30-13:46.

Remarks

> In the video, the primary reporter Jos asks her two colleagues, “So do you think this machine is racist?” We might all have different definitions of the term racism, but for the purposes of this discussion we are going to use this definition of racism - prejudice, discrimination, or antagonism directed against a person or people on the basis of their membership in a particular racial or ethnic group, typically one that is a minority or marginalized.

Discuss: Using this definition of racism, what aspect(s) of racism did the Twitter cropping algorithm do: prejudice, discrimination, or antagonism?

Discussion Goal: Use this discussion to highlight the evidence brought up in the video both from the reporters and the tweets they mention from other users showing how the algorithm was biased or prejudiced towards lighter faces compared with darker faces.

Teaching Tip

This discussion is framed around the Twitter cropping algorithm and the evidence highlighted in the video. You may choose to open this discussion up to other examples based on your students’ experiences.

Discuss: How was the Twitter cropping algorithm trained? According to the video, where is a potential source of bias when training similar cropping algorithms?

Discussion Goal: The most important thing for students to come away from this discussion with is the fact that the machine is not racist. The algorithm that the programmers wrote is biased towards lighter faces. Use this discussion to ensure students understand how the model used the saliency datasets discussed in the video and how some of those datasets were trained on photo libraries containing images that do not represent the population as a whole.

Teaching Tip

The rest of this video dives deeper into machine learning and algorithmic bias, including other examples of how it can lead to harm for certain groups of people. If you’d like to show the entire video to your students, consider watching the parts not listed at the end of class.

Journal: If you were the CEO of Twitter and found evidence of this bias in your cropping algorithm, how would you respond? What steps would you take?

You may choose to have students share their responses if there is time, otherwise move to reading the article.

Remarks

> Now that you've had time to consider how you might respond, we are going to read about how this actually played out and Twitter's response when presented with evidence of bias in their cropping algorithm.

Do This: Have students read the article **"Twitter says its image-cropping algorithm was biased, so it's ditching it (CNN)"**. After they are finished reading they should mark up the text with the following:

- **Highlight / Underline:** Any information in this article that you want to know more about
- **At The End:** Write a 10-word summary of the article

Teaching Tip

To Print or Not To Print? This lesson is written assuming that you have printed out the article and have it physically available for students to write on, even though it is also possible to have students interact with this text digitally. If students read the article digitally, it is most important that they still follow the active reading strategies outlined in this lesson - highlighting the text, writing in the margins, and summarizing. This may require some additional time & instruction to teach students your preferred tools of digital annotation, and may require some additional adjustments to some of the later annotation strategies in this lesson.

Wrap Up (10 minutes)

Discuss: Have students discuss their reactions to the quotes pulled out from Rumman Chowdry, a software engineering director for Twitter's machine learning ethics, transparency, and accountability team.

- Do you think Twitter's response was appropriate?
- What are the risks and potential harms of systems, such as social media platforms like Twitter, becoming too dependent on machine learning?

Remarks

> The story of Twitter and its photo cropping algorithm shows how something that's built with one intention, in this case to better highlight photos, can lead to unintended effects, in this case, bias of white faces over black faces, which can also have harmful impacts on society or culture. And, this case provides an example of a software company that has taken actionable steps to address algorithmic bias by establishing this team (machine learning ethics, transparency, and accountability team) lead by Rumman Chowdry.

Assessment: Check For Understanding

Check For Understanding Question(s) and solutions can be found in each lesson on Code Studio. These questions can be used for an exit ticket.

Question: Which of the following was an unintended effect of the use of the Twitter cropping algorithm?

2 3