Module #4: Identifying Biases and Limitations in ML Datasets

Module Overview

In this module, we will identify strategies for improving our machine learning models. To do this, we will look closely at the data used to train our machine learning models to identify potential biases. Biased training data is often due to an unintended pattern in the data. We will also utilize some of the more advanced features in Google's Teachable Machine to help us evaluate our models. Finally, we will recognize limitations of our machine learning models by showing that they cannot interpolate data when presented with broken teeth.

Driving Question

How can we improve our model's performance?

Primary Learning Objectives

- Identify bias in model input data.
- Interpret internal evaluation metrics provided by Google's Teachable Machine.
- Recognize model limitations by testing performance on incomplete teeth.
- Discuss human biases in the development of ML models.

Materials

- Module 4 PPT
- Computer
- 3D printed megalodon tooth
- Broken megalodon tooth
- Edible megalodon tooth (optional)

In-class Lesson Guide

Activity 1: Identifying Bias in the Training Data

- Use this pre-made model that includes intentional **bias** in the training data.
- From this page, select "Open an existing project from Drive."
- Once you load the data into Teachable Machine, look at the data within each class and look for unintended patterns in the training data.
    - What color backgrounds are used for each class?
    - What direction are the teeth facing in the images?
    - Are there any other objects in the images besides the teeth?
- Any information in the image besides the tooth itself could potentially bias the model. This is called **extraneous data**.
- Any variation exhibited by the object of interest, such as the orientation or color, could also bias the model.
- There are two solutions to address this bias: 1) create a really large dataset that accounts for the wide range of variation; or 2) reduce the amount of unnecessary variation in your training data, so that it focuses solely on the object of interest.
- Ask students why it could be bad to have bias in your training data.

- Lead a discussion on the risk of human bias in technology. (Examples are provided in the Module 4 PPT.)

Activity 2: Improve Your Previous Models

- From this page, select "Open an existing project from Drive."
- Choose one of the models created in Module 3 or use the pre-made model from the previous activity.
- Click "Train Model" to generate your machine learning model.
- After the model has trained, click "Advanced" and then open "Under the Hood"
- Click "Calculate accuracy per class" (Google's Teachable Machine will randomly select 15% of your model's training data to test the accuracy of each class.)
- Click "Calculate confusion matrix" (This graph shows you what is being misidentified.)
- With this information, you know which classes need to be improved. Now you can revisit the training data to improve your model.

Activity 3: Recognize Model Limitations

- From this page, select "Open an existing project from Drive."
- Choose one of the models created in Module 3 or use your model from the previous activity.
- Test your model's performance on a selection of broken teeth.
    - You can use these pre-selected images or find others on the internet.
- Ask students to explain why the model could or could not identify the broken teeth.
    - The model likely will not perform very well on the broken examples, even if the identifications are obvious to you because your training data did not include examples of broken teeth.
    - However, in some cases it may correctly predict the identification if enough relevant features are present.
- How much of the tooth do you think needs to be present to get an accurate identification?
- Does it matter which portion of the tooth is present?