CART 451: Edge Impulse Exercise Report

November 1st, 2022

Amanda Clément

ID 40136259

DESCRIPTION OF DATA SET

My data set is composed of 156 items: 50 images of a mini music box [Figure 1], 50 images of a Clipper lighter [Figure 2] and 56 images of random objects which are categorized as noise [Figures 3, 4, 5, 6 and 7]. I took the photographs from varying angles and zoom levels to try to capture as many physical characteristics of the objects as possible. I chose the music box in particular since, despite its small size, it is composed of many distinctive parts that I believed would make it an interesting candidate for analysis. For my second item, I chose a Clipper lighter as it has an iridescent quality that makes it interesting to look at from different perspectives. For the noise images, I photographed two precious stones, an acorn, a metal figurine and an analog clock. These noise objects were chosen as they are small, detailed and were conveniently placed within arms reach from my desk.

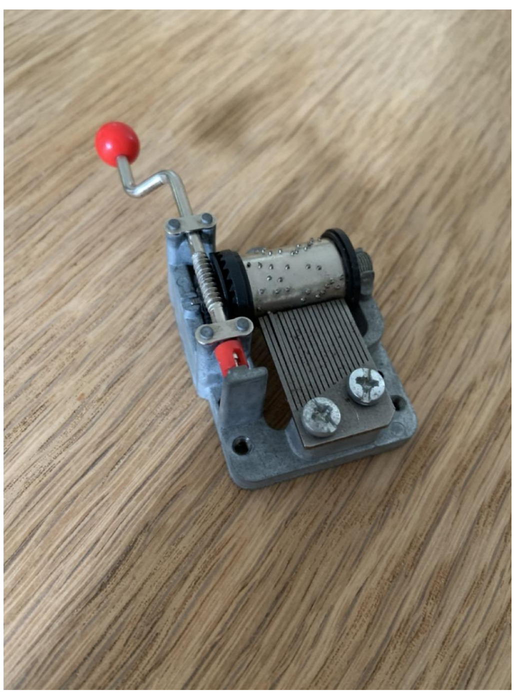

Figure 1. Music box

Figure 2. Clipper lighter

Figure 3. Crystal

Figure 4. Crystal

Figure 5. Acorn

Figure 6. Metal figurine

Figure 7. Analog clock

PURPOSE OF TASK

After being introduced to machine learning and artificial intelligence, we were asked to upload our collection of photographs to Edge Impulse to create, train and test an image recognition model to subsequently analyze its performance. The idea was to use machine learning to analyze patterns in the dataset, which helps us understand, but also reflect on the processes behind machine learning technologies and algorithms. For instance, we must decide how to handle outliers in the dataset that may conflict with the desired outcome, performance or accuracy of the model. By removing these outliers, we are intervening with the model's "natural" performance and have to therefore question our own bias. By leaving these outliers, we may be sacrificing accuracy in the model's performance. As such, this exercise encouraged us to reflect on the decisions one must make when creating and training machine learning models as well as the impacts these decisions may have on the technology's performance.

STEPS TAKEN TO COMPLETE THE TASK

The first step was to upload our collections of photographs, and to group and label them appropriately. In my case, I split them into three groups: music box images, Clipper lighter images, and noise images. Edge Impulse then automatically determined that 120 of my 156 items (77%) would be used for training the model, while the remaining 36 samples (23%) would be used for testing. We then visited the Data Explorer tab in the Data Acquisition section to see a complete view of all the data, where we could clear, inspect or change labels of the individual data items (I personally did not make any modifications). We then created a new impulse on the Impulse Design page, added an Image Processing block to preprocess and normalize the data, and added a Transfer Learning block to use a pre-trained image classification model on the image data. In other words, the raw data was taken, signal processing was used to extract features from it, and finally a learning block was used to classify the new data. Next, we visited the Transfer Learning page to train the model and observe its performance in terms of accuracy. We then went to the Model Testing page to analyze the model's testing accuracy and to see which images the model identified correctly, which images it misidentified, and which images it was uncertain about. Lastly, we deployed and ran the impulse to test it out using live data.

ACCURACY, PRECISION AND RECALL

My model's overall accuracy is 77.78%, with the music box's accuracy at 55.6%, the ligher's accuracy at 86.7% and the noise accuracy at 83.3%. The music box image grouping received an F1 score of

0.67, with precision at 0.83 and recall at 0.56. As such, the model is expected to quite frequently fail to detect music boxes, especially due to Type 2 errors (false negatives) where it gets a sample of a music box but does not classify it as such (the model is uncertain). The model is also expected to make some Type 1 errors (false positives) related to music boxes, although much less frequently than Type 2 errors, where it predicts that something that is not a music box is a music box. The Clipper lighter image grouping received an F1 score of 0.93, with precision at 1 and recall at 0.87. So, it is expected that it occasionally makes Type 2 errors where it predicts that a lighter is not a lighter. The noise image grouping received an F1 score of 0.91, with precision at 1 and recall at 0.87. Hence, it is expected that it will make some Type 2 errors where it gets a sample of noise but is uncertain on how to categorize it. Overall, I consider my model to be mostly accurate, but most problematic when it comes to music box samples.

ANALYZING THE GRAPHS

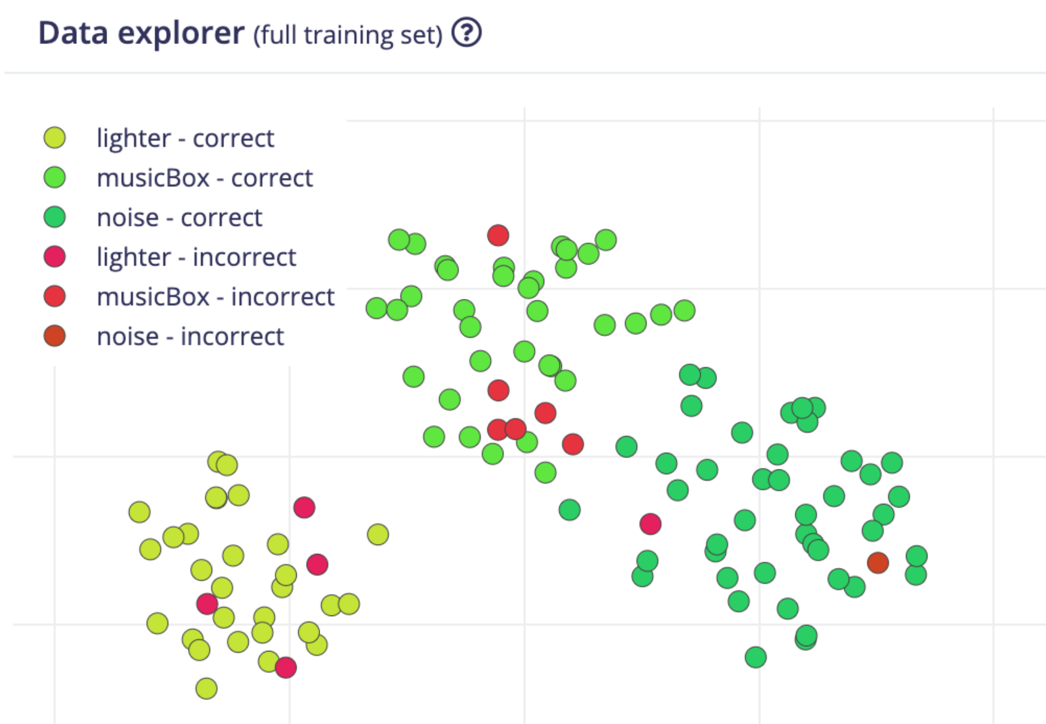

Figure 8. Screenshot of Data Explorer for the training classification set

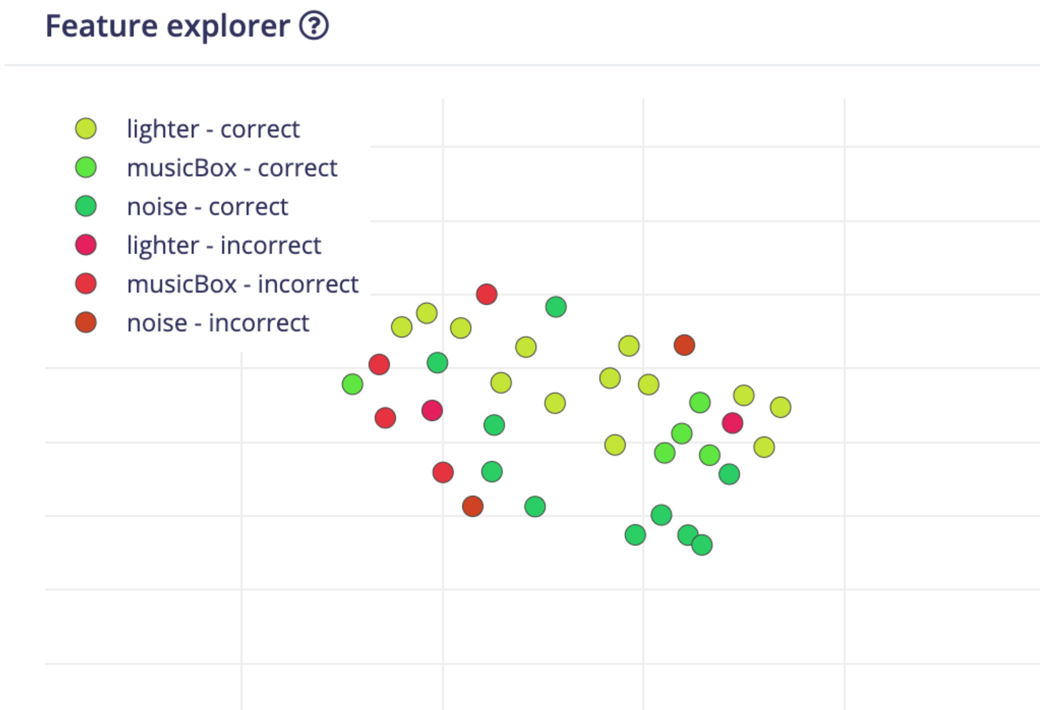

Figure 9. Screenshot of Feature Explorer for the testing classification set

These graphs display the entire dataset as classified by the neural network, which is the mechanism responsible for taking the input data and outputting a probability score for the likelihood that this input data belongs to a particular class. The explorers attempt to extract significant features from the data and then map this information to a 2D space. As such, they provide insight on how the machine learning model classifies the data. Each dot represents a sample item in the set, and those coloured green represent items that are classified correctly while the red ones represent items that are classified incorrectly.

In the Data Explorer, the lighter data (leftmost yellowish-green cluster) is mostly isolated from the rest of the data, whereas the music box data (upper-hand section of the larger right-hand cluster) and noise data (lower-hand section of the larger right-hand cluster) are more closely grouped. As such, the model appears to have an easier time separating the Clipper lighter samples from the rest of the data, but correlates the music box and noise samples more closely. Also, since the music box data has the most misclassified data in the training phase, it is expected that this object will be the most problematic in the model's testing phase and overall performance.

In the Feature Explorer, as expected, the music box grouping contains the most misclassified samples as well as the most scattered dots across the graph. This aligns with the Test Data list, which shows that four of the eight incorrect results are related to the model's inability to recognize the music box given a sample of it. The data dots from the Clipper lighter samples are more closely clustered and feature less red dots since the model has an easier time identifying it. In other words, the model is less likely to misinterpret or fail to identify the Clipper lighter in comparison to the music box. The same goes for the noise data.

IMPROVING PERFORMANCE

To improve the model's performance, I could provide it with more photos of the music box and the Clipper lighter to help it better differentiate them from each other. I could also include more photos of "noise" that features a larger array of random objects to enhance the model's understanding of what is classified as noise, so in other words, what is neither a music box nor a lighter. In this context, I consider an improvement of the model's performance to be based on its level of accuracy in identifying the exact versions of the music box and of the Clipper lighter that I photographed.

Another approach to potentially improve the model would be to train it to identify different versions of music boxes (e.g. wooden ones, ones with more gears) and classify them all as "music box". Likewise, the model could be trained to identify different types of lighters (e.g. varying shapes, colours) and classify them all under the umbrella of "lighter". This would be done by providing the model with photographs of different versions of the objects, giving them all the same label, and re-training the model. In this context, I consider the improvement of the model to be based on its ability to identify and group multiple versions of an object together to classify them under one umbrella term, to give the model a more "general understanding" of what these objects are. This could give the model a broader use.

A SCENARIO USING OBJECT DETECTION

An app could be designed that uses object detection to help users pick out outfits from their wardrobes. The app would pull images from online fashion-related resources like Vogue magazines and runaway shows, along with descriptions of the clothing if available, to train the model to detect current trends such as colours of the seasons, popular jean cuts and pattern matching [Figure 1 in storyboard]. The user would take photos of their own clothing articles and upload them to the app [Figures 2 and 3]. They can then prompt the app to generate an outfit for them based on the photographs they uploaded [Figure 4]. This generated outfit [Figure 5] would be based on the images collected from the fashion resources. The outfit then gets added to the app's database and is displayed on the main page for other users to see, along with a description of why the machine learning model put together this particular outfit (e.g. the colour blue and bootcut jeans are trending this season). Users can then rate these outfits in terms of accuracy to help train the model [Figure 6]. For instance, if the description says "this outfit was chosen because boot cut jeans and blue knit camisoles are trending this season" but features a green camisole in the photograph, users may report this as having low accuracy to inform the model that it misinterpreted the data [Figure 6]. This app could be an interesting research-creation project to explore how a machine learning model can be designed and trained to "understand" and "interpret" more abstract concepts, like fashion.

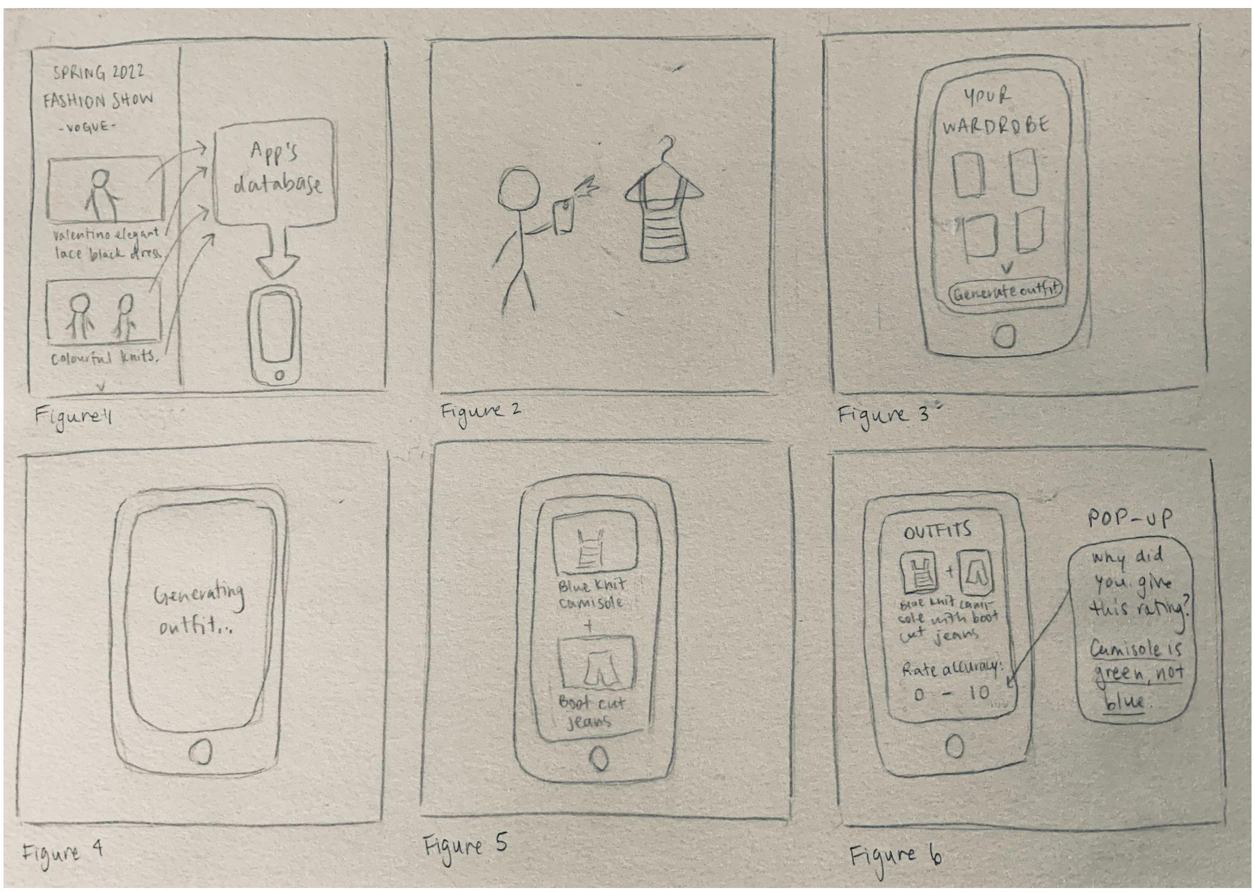

Figure 10. A storyboard sketch demonstrating the object detection app