

Milestone 2 Report

Current State of the project

We have created a prediction model and implemented the non-ML part using streamlit. We keep improving our model (adding more dropout layers, switching to a less complicated model, splitting the dataset into classes), but the model still has the issue of over-fitting. Currently, the non-ML part could be run locally, but it is not completed since the model is not ready. We also created some necessary files such as Procfile, setup.sh, and requirements.txt for hosting our web app on Heroku. Moreover, we have decided to store our completed model through pickle, so later on, we can easily load and use the model.

Further on training our model, we also tried methods other than switching the fully connected layer of a pretrained model. We tried using elements of style transfer, by sending the outputs of several layers from a VGG model to the input of another model, but the results were not satisfying. Finally, we tried splitting our data into classes, and training a classification model instead. Since we have split our data into fifty or more classes, the model gives the illusion of having continuous output values. While this method gives better results than our previous attempts, the overall accuracy of our model is still far from good.

Feature Changes

We have decided to add another functionality to our website which allows the user to choose an artist after uploading the painting and the model will predict a price assuming the painting comes from the certain artist. We know the price of a painting is heavily influenced by the reputation of the artist, but it would be cool to see just how much the price of a particular painting may change when it is assumed to be coming from a different artist. The reason for this change is that even after trying different methods for training the model, the test loss remains pretty high. Also, we observed that some paintings that look similar but come from different artists have different prices (even though we limited the price to be under 20,000) which might have confused the model. One concern for this change is that the model might just predict high prices for more reputable artists and lower prices for others, but we will see how it goes.

When gathering new data, we found that two paintings, coming from the same artist, with similar content and colors visually, and with the only difference being that one is double the size of the other painting, resulted in the price more than doubled. This led us to realize that the price can also be heavily influenced by the painting's dimension. Thus, when gathering new data, we also included the dimension as a feature.

In the end, we will have two models, one of them predicts the price completely visually, and the other takes the artists (and maybe the dimension) into account as well.

Another feature we have decided to add is to let users draw a picture on our website, and let our model appraise their artworks. If they want, they can also choose to present their painting on a

leaderboard on our website. This can make our website more interactive and engaging than a website that only allows image uploads. We found a Streamlit library that provides this functionality so implementation should not be too difficult.

Current Challenges

The main challenge we have is that our model performance is relatively poor, and we believe this might be caused by our dataset. Some similar drawings in the dataset tend to have different prices. Therefore, we realized that prices are not only affected by its aesthetics but also depends on how famous the artist is. As a result, we are going to collect more datas that has both information of the price and the artist.

Tasks Done and Underway by Each Member

Dylan:

- Experimented with training various models with different parameters

Eva:

- Tried improving the model
- Implement the non-ML part with streamlit
- Create necessary classes for hosting on Heroku

Jane:

- Experimented with different parameters of the model
- Gathered new dataset which includes artists and the painting's dimensions

Jared:

- Trained models and experimented with different parameters
- Studied the distribution of prices in our data in order to evenly split our data into classes
- Explored classification models and whether they are better for our task