CISC 839 - G-2 Final Report: Exploring NFT Marketplace

Mahmoud Mohamed Abdien
21mmah@queensu.ca

Mohamed Makram Abo Elsood
21mmza@queensu.ca

Osama Ali El Awadia
osama.elawadia@queensu.ca

1 Background and Objective

NFT stands for Non-Fungible Token. It's generally built using blockchain technology, physical money and cryptocurrencies are "fungible", meaning they can be traded or exchanged for one another. They're also equal in value, one dollar is always worth another dollar. Crypto's fungibility makes it a trusted means of conducting transactions on the blockchain. NFTs are different. Each has a digital signature that makes it impossible for NFTs to be exchanged for or equal to one another (hence, non-fungible). [1] When you pay for an NFT, what you get is the right to transfer the token to your digital wallet. The token proves that your copy of a digital file is the original, like owning an original painting. [2] We demonstrate the usefulness of NFTs to tokenize digital goods, tokenization of digital goods is a perfect fit for NFTs as they can guarantee authenticity and uniqueness, prevent fraud, and improve control over secondary market transactions. [3] The objective here is to either predict NFTs' prices or classify NFTs' images.
Our main goals from this dataset is:

- To get a good price estimate for a new artist so they match the price patterns of similar artists through regression model.
- To sense and block sensitive images from young users (i.e., NSFW content).
- Check the integrity of the given dataset.

Firstly, there are some sort of questions that should be asked to ensure we're meeting the target audience's needs. Those questions are divided into two main parts:
Part 1: Questions that can be answered via statistical analysis
Part 2: Questions that can be answered via predictive analysis

Questions related to part 1:
Q1: Is there a statistical relationship between likes and price?
Answering this question can help those who are interested in collecting NFTs with more likes at an appropriate price and also finding if an NFT is worth what someone is willing to pay for.
This question can be answered using a correlation test like Pearson.
Q2: Is there any significant difference between the prices of the top 5 active creators?
Answering this question can help those who are interested in certain NFT creators and the price related to their artwork.
This question can be answered via ANOVA test. As ANOVA test introduces a p-value, we'd also define the null and the alternative hypotheses
- Null hypothesis (H_0): The price range of all creators is relatively similar.
- Alternative hypothesis (H_1): The price range varies considerably from one creator to another.

Questions related to part 2:
Q1: Are the NFT images enough to predict their price?
Answering this question can help the NFT collectors who are interested in collecting the NFTs that worth their price.
Q2: Can we classify the NFT if it is NSFW or not; based on the NFT image itself?
Answering this question can help the NFT collectors who are searching for a particular categories of artwork that are not NSFW.

2 Dataset

Description:

- The dataset size is about 33 GB
- The art in this dataset is collected from various NFT showrooms and it could be any form of artistic NFT such as photos, GIFs, or videos
- The artwork can be NSFW (i.e., contain sexual content or other inappropriate content)
- Most NSFW art pieces have been labeled as NSFW
- Provider: Arjan de Haan

How are the data collected?!

- He got the data from this website, where people generate their art there and wrote a script to generate the provided data

What types of information are provided in the dataset?
- The dataset contains information about:

- The art itself
- The creator

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

- How many likes
- Whether or not it is NSFW
- The year where art was created
- Prices that the creator had offered
- Path to the files (images, GIF, videos)

What's the connection?

- The main data contains a path to separate files (images, GIF, or videos), all of them are all about various NFT artwork
- We will focus on images, and we will try to classify these images if they are NSFW or not

2.1 Data Preprocessing

- The preprocessing steps we have done on the dataset:

- Dropping 'symbol' and 'royalty' columns as they contain fixed values
- Choosing records with 'PHOTO' only as we worked on images only
- Filtering the top 5 active creators to answer the second question
- Filtering the True NSFW images only to compare them with those labels resulted from the NudeNet classifier.
- Removing outliers in the 'year' column (although it was not used later on)
- Applying label encoding to the 'creator' followed by standardization (i.e., StandardScalar) to make the data zero-mean and unit variance.
- Extract a new column called tags as shown in Figure 1, that contains features extracted from images.

	title	name	tags
30	Mística	Místicismo	{b'Person', b'Human face', b'Fashion accessory...
31	CryptoArts	Satan Lover	{b'Person', b'Human face', b'Man', b'Clothing'...
32	Drawings	Ancient MU	<NA>
33	Static Art	1	<NA>
34	Simplicity	Grow	{b'Person', b'Human face', b'Balloon', b'Toy',...
35	CryptoArts	The Snail In the Rain	{b'Watercraft', b'Poster', b'Boat', b'Vehicle'...
36	About Heraklion	Reality & Dream in front of Koules, Heraklion ...	{b'Box', b'Person', b'Human face', b'Toy', b'C...
37	white owl	white owl	{b'Eagle', b'Human face', b'Falcon', b'Human m...
40	CryptoArts	Cuts and Wounds	{b'Window', b'Person', b'Stairs', b'Man', b'Cl...
41	CryptoArts	Marilyn Monroe - The Zombie	{b'Person', b'Human face', b'Fashion accessory...

Figure 1. The generated feature 'tags'

2.2 Basic Statistics of the Dataset

- The basic statistics of the dataset:

- Using Pandas-Profiling tool in different project's phases to provide us with the most needed statistics from the dataset.
- For example, the report provided by Pandas-Profiling was the guide which let us drop 'symbol' and 'royalty' columns due to their fixed values, see Figure 2.
- Conducting Pearson Correlation Test to can answer the first question.
- Performing ANOVA test to can answer the second question.

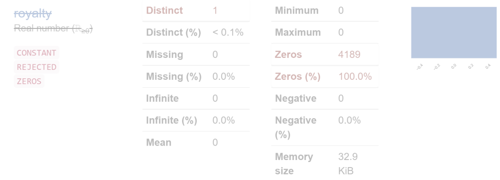


Figure 2. The 'royalty' feature contains a fixed value 'tags'

3 Answers to the research questions

3.1 Answer to question #1: Is there a statistical relationship between likes and price?

We have answered this question using Pearson correlation test. This conducted hypothesis test assumes the null hypothesis is that there is a statistical relationship between likes and price while the alternative hypothesis is there is not. The Pearson's correlation resulted coefficient was 0.025 which implies low degree of correlation which means we would reject the null hypothesis and accept the alternative on that states there is no relationship between likes and price. This indicates that there is no rule to follow; for those who are interested in collecting NFTs with more likes at an appropriate price and also finding if an NFT is worth what someone is willing to pay for.

3.2 Answer to question #2: Is there any significant difference between the prices of the top 5 active creators?

This question can be answered via ANOVA test.
- Null hypothesis H_0: The price range of the top five active creators is relatively similar.
- Alternative hypothesis H_1: The price range varies considerably from one creator to another.
We have used the top five active creators to perform ANOVA Test as shown in Figure 1. After performing One-Way ANOVA Test (as we only have one independent variable) the p-value was 0.006 which gives a strong evidence to reject H_0. Instead, we have performed the Test using the top 10 active creators but the p-value was still significantly low (0.005), see Figure 2.

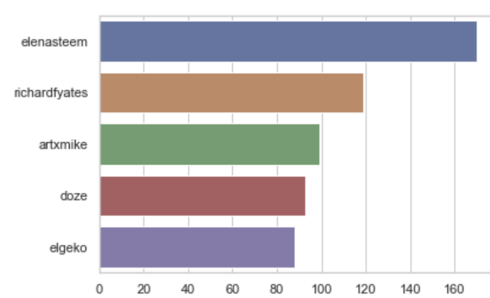


Figure 3. Top 5 active creators

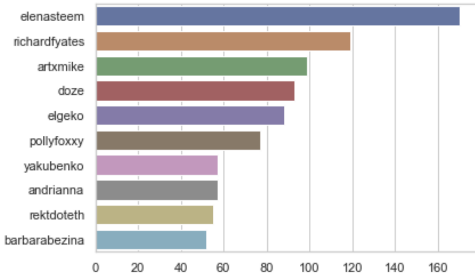


Figure 4. Top 10 active creators

3.3 Answer to question #3: Are the NFT images enough to predict their price?

Regarding price, we have built two models, one with (images text) and another one with image only. First model is implemented using word embedding then attention layers then Bidirectional-GRU which is a Recurrent Neural Network that allow model to use the previous and next time steps to make prediction about current state, embedding layer turns positive integers into dense of fixed size, attention layers to attend to different positions of the input sequence to compute a representation of that sequence. For images: a convolution layer convolved with input images to produce tensor output then max pooling layer to downsample the input along its spatial dimensions, then Flatten to reshape input to one vector. output layer with linear activation function. Then compile that with Adam optimizer and mean square error as a loss function on validation set which was 20.0526. The second model was based on the same layers as images of first model(Convolution, MaxPooling, Flatten) and loss function on validation set was 34.3052. The model started to learn as loss got much higher but because prices are considered not reasonable. Our models were able to detect some patterns as you can see that images and text help the model to identify patterns better than image only.

3.4 Answer to question #4: Can we classify the NFT if it is NSFW or not; based on the NFT image itself?

Classifying NSFW images is a challenging task to do manual, it is highly inappropriate for someone to classify it one by one, but it also challenging for Machine Learning models too because the definition of NSFW may vary from someone to another, and there could be a hidden context or hand signals that make the photo NSFW, a Deep Learning model won't recognise the difference between full and partial nudity, it will only consider the features of exposed parts, which unfortunately is common on both labels, to solve this problem we needed a model that consider the whole picture and to not give high weights to small details. For the data, as it is highly imbalanced data, we have tried two different approaches, the traditional one and nudity detection neural network NudeNet. Table 1 shows the accuracy of the traditional approach for the classification model that followed the official guide from TensorFlow. We have used ImageDataGenerator to apply augmentation and preprocessing (shearing, rotation, zoom and scaling). Table 1 shows the accuracy of first model. Figure 5 shows the accuracy-loss curve.

Table 1. Traditional approach to classify NSFW image

	Accuracy
Train set	95.81%
Validation set	97.56%

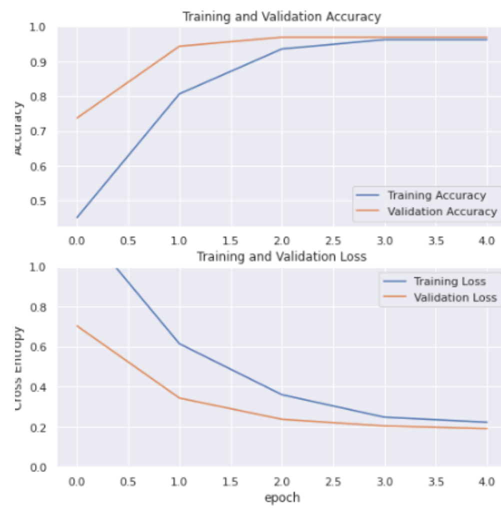
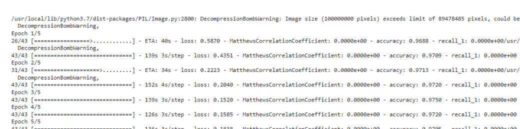


Figure 5. The Loss-Accuracy curve

Figure 6 shows the accuracy, recall and MCC.



```
/usr/local/lib/python3.7/dist-packages/PIL/Image.py:2800: DecompressionBombWarning: Image size (100000000 pixels) exceeds limit of 89478485 pixels, could be
  DecompressionBombWarning,
Epoch 1/5
26/43 [=================>...........] - ETA: 40s - loss: 0.5870 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9688 - recall_1: 0.0000e+00/usr/l
  DecompressionBombWarning,
43/43 [==============================] - 139s 3s/step - loss: 0.4351 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9709 - recall_1: 0.0000e+00 -
Epoch 2/5
31/43 [====================>........] - ETA: 34s - loss: 0.2223 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9713 - recall_1: 0.0000e+00/usr/l
  DecompressionBombWarning,
43/43 [==============================] - 152s 4s/step - loss: 0.2040 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9720 - recall_1: 0.0000e+00 -
Epoch 3/5
43/43 [==============================] - 139s 3s/step - loss: 0.1520 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9750 - recall_1: 0.0000e+00 -
Epoch 4/5
43/43 [==============================] - 126s 3s/step - loss: 0.1585 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9720 - recall_1: 0.0000e+00 -
Epoch 5/5
43/43 [==============================] - 136s 3s/step - loss: 0.1838 - MatthewsCorrelationCoefficient: 0.0000e+00 - accuracy: 0.9705 - recall_1: 0.0000e+00 -
```

Figure 6. The evaluation metrics of the model

We can see that the model is saturated and the generalization gap is closed, but the f1, recall remains 0 This is probably because the model is only classifying SFW, and the accuracy of NSFW is 0 so it is divided by 0. We trained the model multiple times and tried different splitting methods, still the results remained the same, there is 3 potential reasons why we got this results.

- The corrupted images were too many
- We defined the metrics wrong

- The NSFW images doesn't containing enough information to be distinguished from SFW images
- The miss-labeling that we found in the SFW/NSFW caused the model to be confused and not learn the difference
- We tried to solve this problem by using pre-trained model NudeNet, but the results didn't also match the labels.
- From this result we conclude that we cannot classify NSFW from images only.

However, we have applied the NudeNet to the whole images we have in the dataset then extract the 'unsafe' percentage from the resulted dictionary (which shows the safe-unsafe percentage) and compare it with a ground truth 0.5 (i.e., if the unsafe percentage is > 0.5, it is considered as unsafe, otherwise it is safe). The results were surprising-as shown in Table 2-as approximately 5% of the images were classified as NSFW; while the 'nsfw' feature itself states that nearly 3% is only the percentage of NSFW images in the dataset. This means that there are images that are classified as SFW but they are actually NSFW, and vice versa. Then, we have used the same approach with the NSFW images only (i.e., the images where 'nsfw' = True). Table 2 shows the results that support our investigation that images are not classified correctly whether they are SFW or not. More than 75% of NSFW images are classified as SFW by NudeNet classifier.

Table 2. The whole images using NudeNet vs 'nsfw' feature

Label	Whole images - NudeNet	'nsfw' feature
True	5.246%	3.342%
False	94.754%	96.658%

Table 3. NudeNet on the NSFW images only

Label	NSFW images - NudeNet
True	21.186%
False	78.814%

4 Limitations

- This dataset is not yet complete, the prices are not of sold art, but a presented prices (i.e., not the real prices).
- We also noticed a lot of missed labeled data in the 'nsfw' column which confuses the model during training, for example seeing some part of the body exposed in one NSFW image and seeing it again as SFW, this type of inconsistency resulted a lower recall.
- NFTs are only 4 years old, both artists and customers don't have enough experience to value the prices of an NFT.
- During training we found that some of the images have too large size for the model to handle and some are corrupted they don't contain JFIF header which TensorFlow requires.

5 Take-away messages

- The main challenge in this dataset, that it seems a bit unfinished, in order to be able to predict prices accurately, we need to have an actual prices of sold art not presented art, and we also need to have an accurate price of the crypto-currency itself.
- For the NSFW prediction, even after using a pre-trained model for detecting NSFW images, the model found that some of the SFW pictures are NSFW.
- We haven't manually examined the NSFW data ourselves but NudeNet could only recognise small portion of the data, because it's probably not suitable in context, and since the NSFW data was only 118 images, this was clearly not enough.

6 Replication package

- Our first deliverable is the notebook "DAG2exploringnft", the packages required to be installed can be found on Colab.
- Kaggle doesn't allow us to modify the directory, the reason why we need to change directory is to split the data into (train, validation, test) for the classification task.
- The Drive we used after splitting the NSFW and SFW data can be found here.
- The notebook named "Pandas-Profiling + NudeNet Classifier" contains the Pandas-Profiling report in addition to NudeNet Classifier described before.
- The notebook named "price$_{p}rediction_{u}sing_{i}mages_{K}aggle$”.

7 Distribution of Workload

so far we have made some collaborative work in data analysis and data exploration, we distributed the upcoming work based on what we found each one comfortable with:

- Osama El Awadia: Data Cleaning & Data Analysis.
  Osama has done further data exploration, found important insights/patterns that helped us to achieve our goal, and applied NudeNet classifier to compare its results with those resulted from the traditional approach. He will also work closely with Abdien in the feature engineering task to proof our hypothesis and regression analysis problem.
- Mahmoud Abdien: Feature engineering & creating new features.
  Abdien has worked on extracting new features and prepossessing the inputs for our BERT and CNN models, suggested most suitable models to be implemented in our pipeline. He also created the object detection

algorithm and extract a new feature contains tags that may help in the NSFW classification task.

- Mohamed Makram: Regression model and statistical analysis like Pearson Correlation Test and ANOVA Test.
  Makram has implemented the predictive pipelines that we find suitable for the task, fine tune the model and hyperparameter search.

The three of us have worked closely with each other especially because our problems are co-dependant and finish-to-start tasks.

Acknowledgments

This project would not have been possible without the rich content Dr. Yuan provide us with. We must also thank Dr. Mona for her unstoppable support and encouragement. Last but not the least, we would like to recognize the invaluable assistance of the TAs Sadegh, Shahenda, and Ashraqat.

References

[1] John Schmidt Robyn Conti. *What Is An NFT? How Do NFTs Work?* URL: https://www.forbes.com/advisor/in/investing/what-is-an-nft-how-do-nfts-work/. (Updated: Mar 25, 2022).

[2] Dragan Boscovic. *How nonfungible tokens work and where they get their value.* URL: https://theconversation.com/how-nonfungible-tokens-work-and-where-they-get-their-value-a-cryptocurrency-expert-explains-nfts-157489. (Published: March 31, 2021).

[3] Ferdinand Regner, Nils Urbach, and André Schweizer. "NFTs in practice–non-fungible tokens as core component of a blockchain-based event ticketing application". In: (2019).