The Price of Art

Viktoriya Andonova
SFDAT22, General Assembly

Table of Contents

Problem overview

$66B industry but it's still old-school [1]

Emerging online salesmen such as Artsy, Amazon Art, and Saatchi Art [2]

The price of art seems to be very subjective – determined by a select few who decide what is good or not

Is there a formula for determining the value of art based on its features?

1. http://www.bloomberg.com/news/articles/2014-03-12/global-art-market-surged-to-66-billion-in-2013-report, 2014
2. http://fortune.com/2015/01/22/artsy-galleries/,

What do you think?

Image A

Image B

What do you think?

Image A

Man on A Horse (Full Wide Shot) A. Schallenberg, 2007

Image B

#13 from the series We Communicate Only Through Our Shared Dismissal of the Pre-Linguistic, 1995

Click on the image to zoom in

IN ROOM

Specifications | About the Artist | About the Gallery

ANSEL ADAMS

Fresh Snow, Yosemite, CA, 1947

Silver Gelatin
Edition of 1

Size : 10.75 x 12.75 in.

In stock.

Price: $14,250.00 + $60.00 shipping or Make an Offer

Estimated Delivery Date: June 9 - 14 when you choose Standard at checkout.

Make an Offer

Add to Cart

After you check out, we may contact you to confirm purchase of this item. Learn more

Add to List

Ships from and sold by Perspective Fine Art.

Need help? Contact the gallery.

Share

About This Artwork

Titled "Winter Forest" in an unknown hand together with studio stamp on mount verso. Printed ca 1965-70 An exceptionally beautiful print with an important provenance. A rare example of this image printed by Adams prior to the 1974 limited edition Portfolio VI. Between 1947 and the production of Portfolio VI, Adams made fewer than eight prints from this negative which included two screens and a mural sized print. It is estimated that only one or two prints of this size and quality were made.

Data Cleaning and Exploratory Analysis

Max price/sqin was $1,714.29

Created 'area', 'price_sqin', 'price_class' features

```
amazon_art.price_class.value_counts()
```

```
3.0    1777
2.0    1213
1.0     810
4.0     400
0.0     244
5.0      72
6.0      10
Name: price_class, dtype: int64
```

Average sentiment and clusters

```
model_df.loc[model_df.price_sqin < .50, 'price_class'] = 0
model_df.loc[(model_df.price_sqin >= .50) & (model_df.price_sqin < 1), 'price_class'] = 1
model_df.loc[(model_df.price_sqin >= 1) & (model_df.price_sqin < 2), 'price_class'] = 2
model_df.loc[(model_df.price_sqin >= 2) & (model_df.price_sqin < 10), 'price_class'] = 3
model_df.loc[(model_df.price_sqin >= 10) & (model_df.price_sqin < 50), 'price_class'] = 4
model_df.loc[(model_df.price_sqin >= 50) & (model_df.price_sqin < 200), 'price_class'] = 5
model_df.loc[(model_df.price_sqin >= 200), 'price_class'] = 6
```

Hypothesis

While style and content matter significantly in determining the price of a photograph, the photographer's popularity ultimately drives the amount paid for a piece of art.

Model Building: artist vs price_class

Logistic Regression

```
In [67]: # define X and y
         X = artist_dummies
         y = model_df.price_class

         artist_logreg(X, y)
```

```
Accuracy:  0.782685512367
     |   0    1    2    3    4    5    6 |
     |   .    .    .    .    .    .    . |
     |   0    0    0    0    0    0    0 |
----+-------------------------------+
0.0 | <32> 22    .    6    .    .    . |
1.0 |   3<132> 28   27    .    .    . |
2.0 |   .  34<249> 34    .    .    . |
3.0 |   .    .  29<414> 13    .    . |
4.0 |   .    .    .  40 <56>   .    . |
5.0 |   .    .    .   5    4  <3>   . |
6.0 |   .    .    .    .    1    .  <.>|
----+-------------------------------+
(row = reference; col = test)

Null Accuracy Rate:  0.392620415378
```

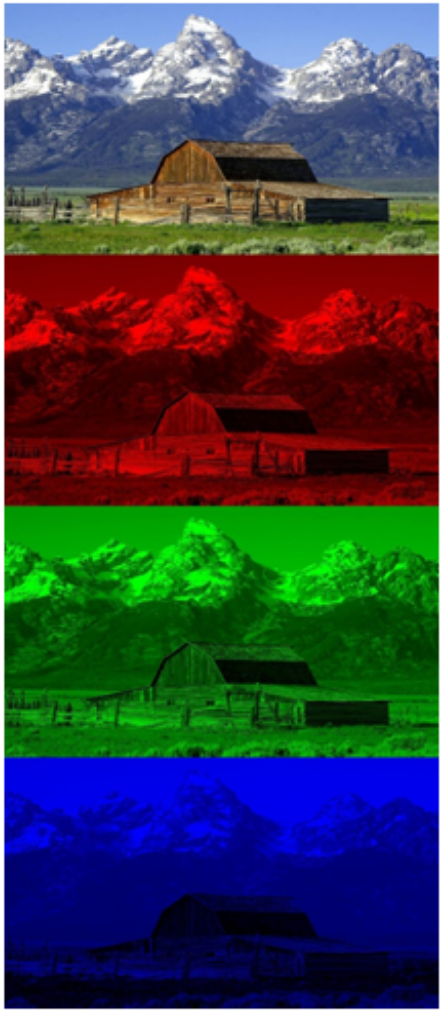
Naïve Bayes

```
In [69]: # define X and y
         X = artist_dummies
         y = model_df.price_class

         artist_nb(X, y)
```

```
Accuracy:  0.78445229682
     |   0    1    2    3    4    5    6 |
     |   .    .    .    .    .    .    . |
     |   0    0    0    0    0    0    0 |
----+-------------------------------+
0.0 | <27> 27    .    6    .    .    . |
1.0 |   .<146> 28   16    .    .    . |
2.0 |   .  35<243> 39    .    .    . |
3.0 |   .    .  22<424> 10    .    . |
4.0 |   .    .    .  51 <45>   .    . |
5.0 |   .    .    .   5    4  <3>   . |
6.0 |   .    .    .    .    1    .  <.>|
----+-------------------------------+
(row = reference; col = test)

Null Accuracy Rate:  0.392620415378
```

Computer Vision

- Images = 3D numpy array
- Extract average value for each channel
- 6 new features in the DataFrame



R=red

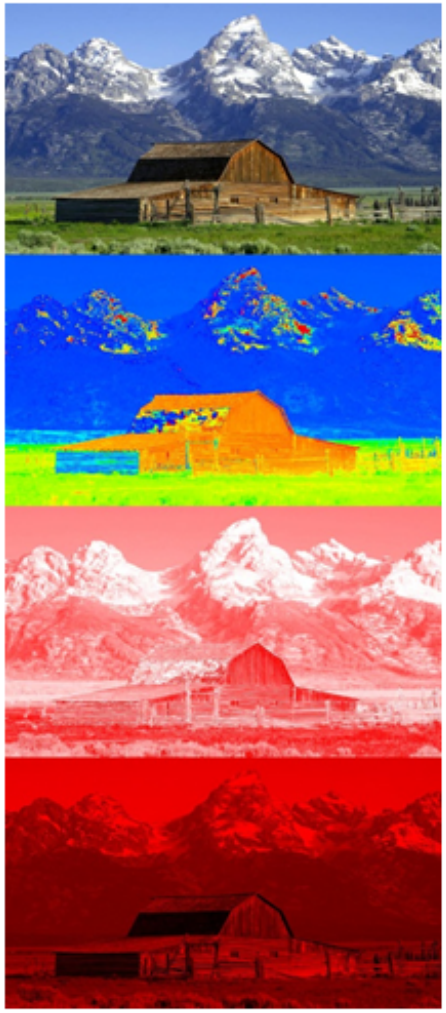
G=green

B=blue

Hue=H

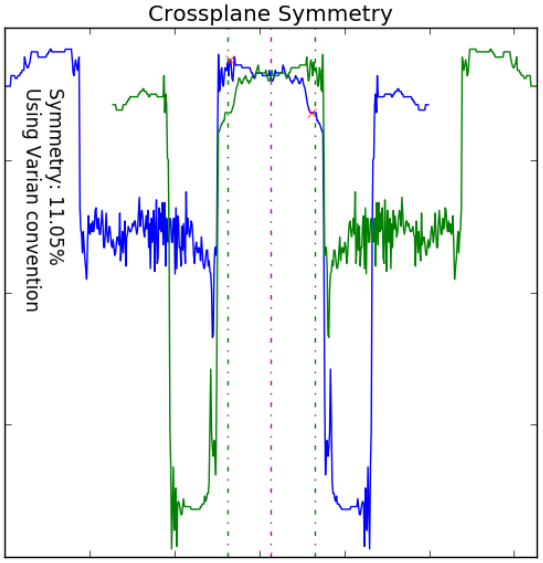
Saturation=S

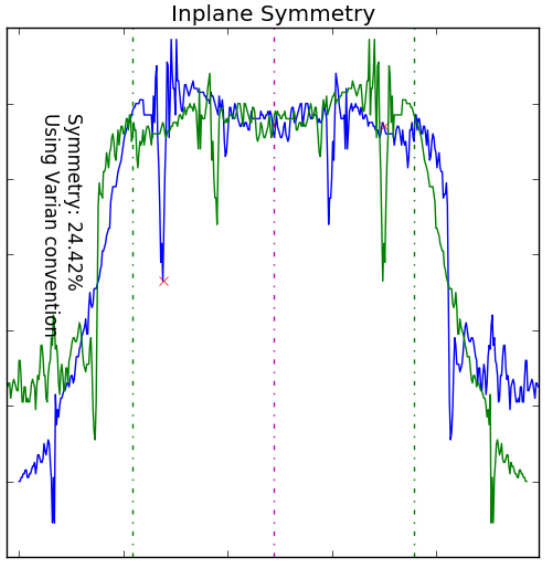
Value=V

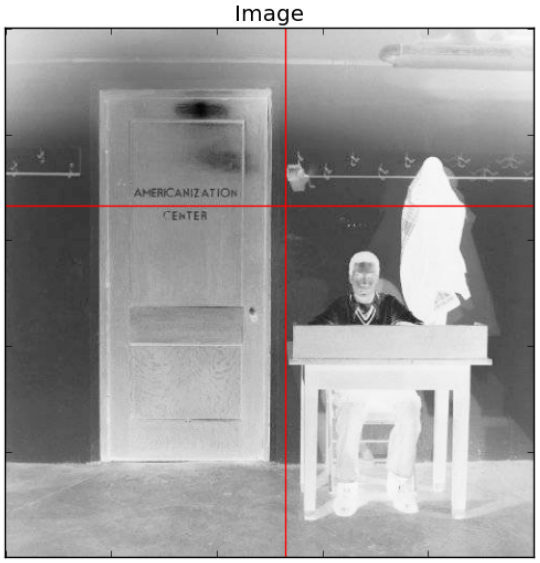


Feature Extraction: Image Processing

- Symmetry extraction
  - pylinac
- Dimension reduction
  - Canny edge detection





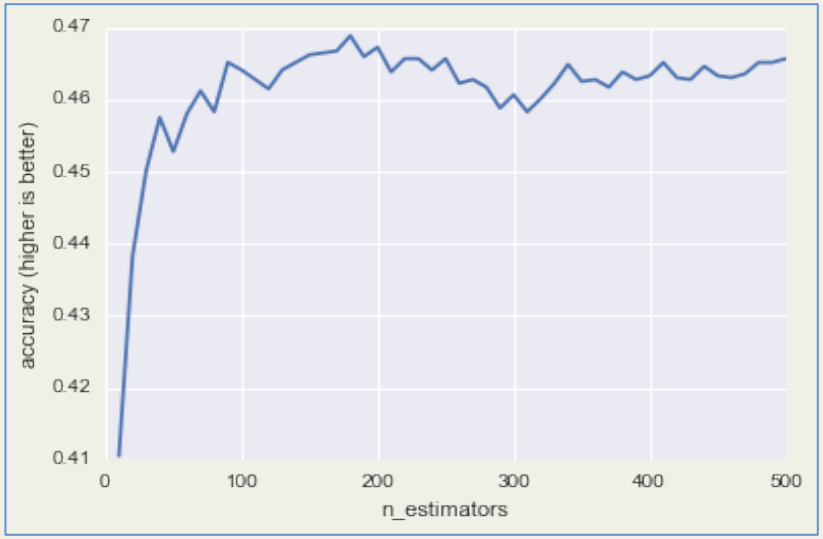


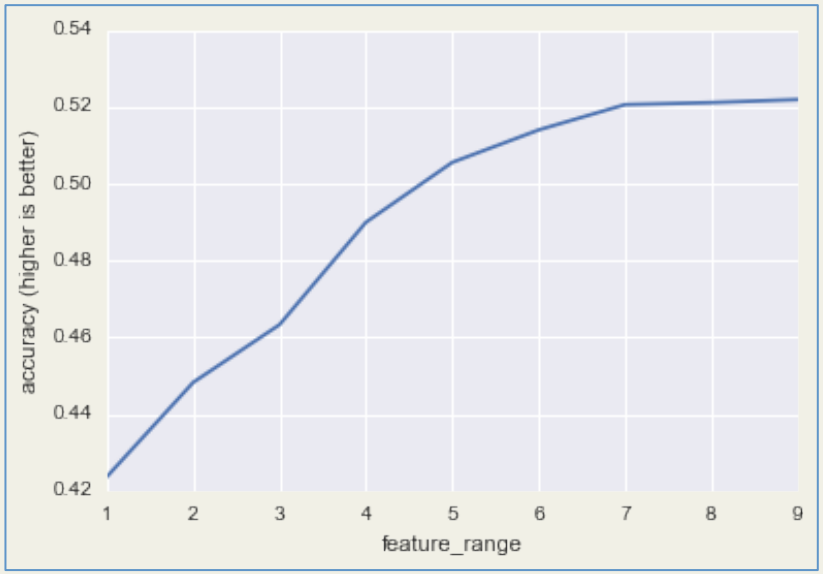
Canny Edge Detection

Original Image

Edge Image

Model Building: img_features vs price_class



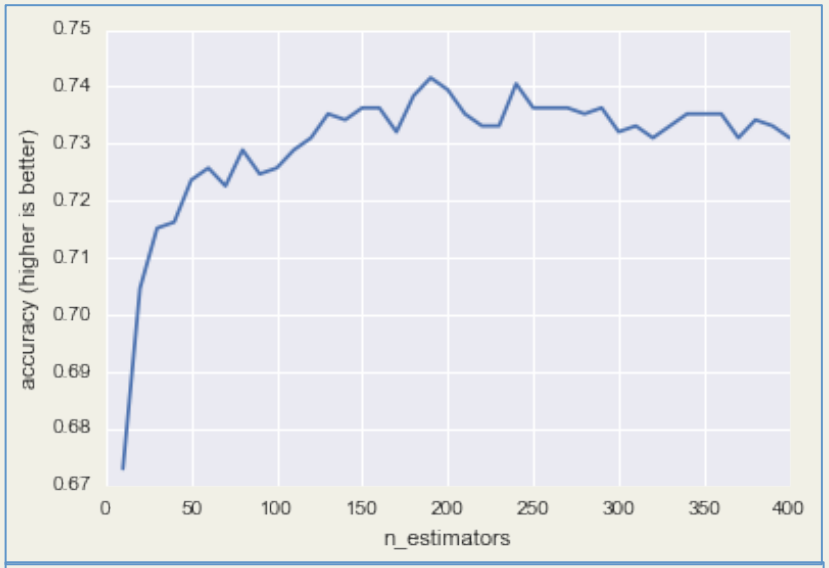


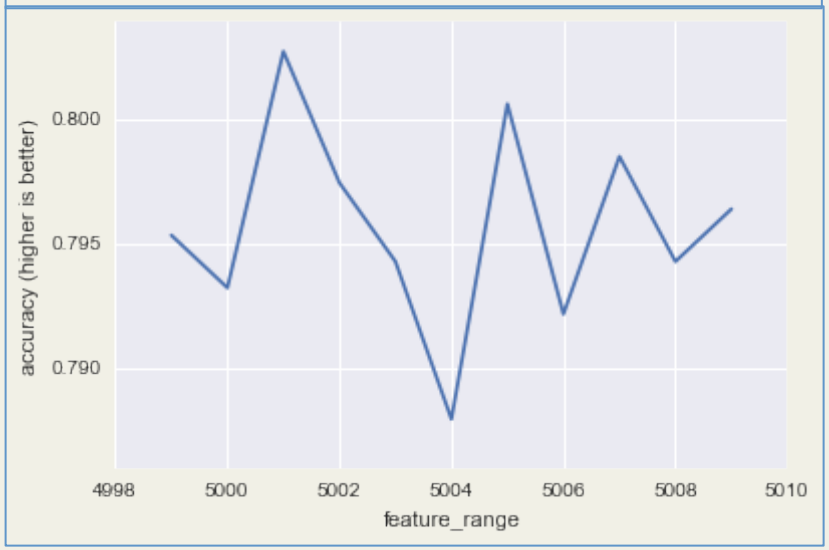
Feature	Importance
area [1]	.283392
crossplane_sym	.107661
inplane_sym	.106354
avg_hue	.105140
avg_saturation	.103728
avg_b_value	.076908
avg_g_value	.075597
avg_value	.070818
avg_r_value	.070402

Accuracy Score: .522664128397

Out of Bag Score: .610817941953

Model Building: img_features + description vs price_class





Feature	Importance
area	.203182
avg_hue	.039236
avg_saturation	.037951
crossplane_sym	.036318
inplane_sym	.031865
avg_b_value	.024325
avg_value	.023587
avg_g_value	.022860
avg_r_value	.021606
titled	.018570
recto	.021606
limited	.005061
american	.004892
edition	.004643
dated	.004431

Accuracy Score: .802742616034

Out of Bag Score: .765306122449

Conclusion

Content matters!

Next Steps

Continue to refine dataset and scrape more data

Define more functions to better visualize the results

Incorporate Google's content analysis API

Look for an artist popularity metric

Create a recommendation engine for artists

- Where to sell
- How much to sell for