

Harvard Data Science Review • Issue 2.2, Spring 2020

Can Machine Learning Predict the Price of Art at Auction?

Jason Bailey

Published on: Apr 30, 2020

DOI: 10.1162/99608f92.7f90ce96

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Column Editor's note: *It has been said that beauty is in the eye of the beholder. But how well can beauty be valued by the eye of machine learning? Jason Bailey, art analytics expert, explores in this issue's Recreations in Randomness the capabilities of machine learning to provide reliable valuations of artwork that can be used by buyers and auction houses, and how these valuations might even shape the art market.*

Keywords: art analytics, art auction, art valuation, automated appraisal, pricing

In December of 2019 at the Art Basel Miami art fair, Italian artist Maurizio Cattelan, known for his satirical approach to art, stuck a banana to the wall with duct tape, titled the work *Comedian*, and swiftly sold three copies for \$120,000 to \$150,000 each (Pogrebin, 2019). By the standards of the \$67.4B-a-year art market (McAndrew, 2019), this may have been a bargain.

In the last 3 years, the seemingly unpredictable world of blue chip art auctions has caught the attention of the public and the mainstream media. Upon selling for \$1.4M at Sotheby's, Banksy's painting *Girl With a Balloon* proceeded to shred itself in the auction house through a mechanism hidden in its frame. The public naturally assumed the damage from the shredding made the painting worthless. On the contrary, several Banksy experts believe the notoriety from the stunt likely doubled its value and will raise the prices of all Banksy works (Calfas, 2018; Reyburn, 2018).

In 2019, Jeff Koons's stainless steel sculpture *Rabbit*, modeled after a cheap inflatable toy bunny, fetched \$91M, setting a new record for most expensive work by a living artist sold at auction (Wamsley, 2019). And in 2017, the record for most expensive work by any artist was broken when a heavily restored painting thought to be by Leonardo Da Vinci sold for a whopping \$450M. The very same painting had sold for \$10,000 just twelve years earlier in 2005 (Greene, 2017).

These sales regularly make global headlines because the prices seem to defy human logic. But what about machine logic? Are there specific qualities consistent across artworks and artists that can be detected, broken down, and analyzed by human experts or machine learning models to predict the value of art at auction accurately? Or could the prices paid for art simply fall outside the realm of prediction?

Human Appraisal of Art

On the surface, the art market, built on the fickle tastes of individual collectors, would appear to be particularly unpredictable. Impressionism may be in vogue today and out of style tomorrow. Entire markets can seemingly shift overnight based on the whims of a handful of influential collectors. Unlike fungible stocks that trade in high volume, each artwork is unique and different from the next,

and works can go decades between sales, making it difficult to establish any logical pattern or change in value (Mei & Moses, 2002).

In describing the art market, Plattner (1998) characterized it as having two unusual qualities. First, people spend enormous sums of money on objects that are nearly impossible to value. Second, given the limited number of potential buyers, it is difficult for artists to know if there will be any demand for the work they are creating. These challenges have led some economists to caution would-be art investors against fooling themselves into believing they could somehow turn a profit in a market where prices "float around aimlessly" with the randomness of a dice game (Baumol, 1986).

Regardless of what skeptics might think about the feasibility of price prediction, professional art appraisers must develop a presale price estimate for each artwork that goes to auction. Trained in particular genres and geographical markets, appraisers take into consideration many important factors, including the artist, country of origin, age of the artwork, dimensions, rarity, signature, materials, subject, and condition. The appraiser then looks for similar items recently sold at auction and considers the current market demand. Finally, they develop both a low estimate and a high estimate, creating a range of values within which they believe the work could reasonably sell.

Based on the past performance of art appraisers employed at the two largest auction houses, Christie's and Sotheby's, art prices would appear to be somewhat predictable. Actual prices fell between the appraiser's high and low estimate for Christie's 41% of the time and within range for Sotheby's 37% of the time in a study by Bjerg (2018) looking at 195,479 sales from 2016 and 2017.

It would appear that art appraisal is a valuable skill that improves over time. For example, auction houses with more experience in selling work by a given artist not only produced more accurate presale estimates, but they also had a higher likelihood of completing a sale, according to Bruno et al. (2016). One reason for this might be that auction houses with experience selling works by a particular artist have developed personal relationships with the collectors of that artist's work and, therefore, better understand the demand in the market.

There is also strong evidence that appraisals act as anchors in influencing the price that collectors are willing to pay for a work at auction. Appraisers themselves are not immune to the effects of anchoring to past estimates when constructing new appraisals for works returning to market (Beggs & Graddy, 2009).

But is appraisal an art form, relying on a combination of human relationships, soft skills, insider industry knowledge, and human bias, or is it science with discrete steps that can be replicated and mastered by a machine learning model?

Machine Learning Appraisal of Art

One might assume that the sale price of artwork at auction would depend heavily on its visual features. After all, that is what we do with art—we look at it. So it makes sense that when trying to predict the value of art using machine learning, most data scientists start with visual analysis. Convolutional neural networks are specifically designed for this type of research and are a popular choice. But surprisingly, researchers have found convolutional neural networks do a poor job in predicting the price of art at auction on their own. Models based on numeric and textual data perform far better than those looking only at images (Aubrey et al., 2019; Ayub et al., 2017). The implication is that what art looks like is not as important as one might think in determining how much it will sell for at auction. At least not to machine learning. However, companies like Art Recognition have been using computer vision and machine learning to authenticate which artist created a given work of art. This authentication can have enormous impact on its value at auction.

How could it be that the visual properties of art are not that important in establishing its value at auction? After all, shouldn't the inherent qualities of an artwork—its craftsmanship, its aesthetics, and its ability to resonate with viewers—be primary drivers of its price? We know this is not the case. Artworks discovered to be forgeries or misattributed can drop in value from hundreds of thousands or even millions of dollars to almost worthless in a heartbeat despite nothing physically changing about the work. This rapid depreciation would imply that art's value at auction derives almost entirely from the artist's reputation and has little to do with the artwork itself. One might then assume that an artist earns their notoriety based mostly on their artistic skills. Yet, new evidence suggests reputation may have much to do with early access to prestigious galleries and museums.

To better understand how access to prestigious institutions impacts artists' reputation and valuation, Fraiberger et al. (2018) mapped out the careers of 496,354 artists, looking at their exhibitions and sales. They placed the artists in the study into two groups based on the prestige of the institutions where they held their first five exhibitions. The results suggest that access to prestigious institutions had an enormous effect on an artist's career trajectory and played a significant role in predicting the sale price of their work.

Artists with early access to high-prestige institutions had twice as many exhibitions as low-prestige artists. High-prestige artists' work was traded 4.7 times more often at auction and at a maximum price that was 5.2 times higher than for low-prestige artists. The team then collected 442,314 prices of artworks displayed in galleries and found the average maximum price for high-prestige artists was \$193,064 versus \$40,476 for low-prestige artists.

Disappointingly, the findings also suggested that artists born into countries without access to high-prestige institutions tended to stay low prestige throughout their careers. One might attribute this to

selection bias, implying the most prestigious institutions selected the most talented artists. However, it is unlikely that one's artistic talent correlates with the country or region in which they are born. Yet artists born in countries with little or no access to high-prestige institutions consistently struggled, most often starting and finishing their careers as low-prestige artists.

Given these findings, it appears important to mine the textual data describing artists and artworks for signals of reputation and prestige in addition to performing a visual analysis of the art. Though not as accurate as human appraisals, a machine learning model performed well when trained on textual and numeric data, including artist, year of creation, materials, and size, by Aubry et al. (2019). Their machine learning model was significantly more accurate in predicting prices of art at auction than a hedonic regression model trained on the same data. This improvement is likely due to machine learning's strength in analyzing not only the training features but also the complex relationships between those features. For example, understanding the dynamics between artist, color, and dimensions, and their impact on pricing can be challenging to capture with hedonic regression. The machine learning model was also resistant to a systematic bias found in human appraisers who failed to sufficiently reduce their estimates in response to negative information about decreasing sale prices in artists' markets.

A second way to potentially automate the capturing of artists' reputation and prestige would be to perform text analysis on their biographies and evaluate their social media presence. Powell et al. (2019) used text analysis to examine how artists sold their work through sites like Saatchiart.com and ArtFinder.com. The team analyzed descriptions of artworks and artists' biographical information using popular machine learning techniques including k-nearest neighbors, support vector machines, and random forest classifiers to see if word count was indicative of sale price. They selected the random forest model as the base for testing of additional features as it performed the best out of the initial group. In the testing, word count from the description of the artwork had an impact, but the best results came from a combination of artwork description word count and artist biography word count. They also tested to see if the presence of social media accounts impacted prices. Interestingly, links to social media accounts had no real impact on their own, but when combined with word count, they showed to have some influence on sale price. One explanation for this is that social media could primarily be a measure of an artist's marketing savvy, which may already be accounted for in the word count of their biography and artwork descriptions.

Expanding the Art Market Through Automated Valuation

Manual appraisal of art is slow, expensive, and limited by the number of human experts available. While not as accurate, machine learning can dramatically increase the volume, speed, and frequency of appraisals. Instead of only estimating prices for art headed to auction, all artworks, on and off the

market, could regularly undergo automated valuation using machine learning. This approach is already in use in the real estate market with companies like Zillow, Redfin, and Trulia.

Public pricing estimates for all artwork and not just art headed to auction could help grow the art market. Economists generally agree that increased transparency leads to increased liquidity in markets (Lang & Maffett, 2011; Pagano & Röell, 1996). While some market participants may not want greater transparency because they benefit from an inefficient system, there is evidence that the lack of publicly available pricing is the number one deterrent to collectors buying more art (Artsy, 2019).

Various automated appraisal systems have been around for close to a decade, but have enjoyed limited success. Currently, at least one major auction house, Sotheby's, is seeking to improve the use of machine learning for appraisal automation to help drive liquidity (A. Qamar & A. Shum, personal communication, October 24, 2019). Ahmed Qamar, VP director of machine learning, and Andrew Shum, VP director of product, lead Sotheby's machine learning team. According to Qamar and Shum, their team has two goals: making selling easier and driving business impact. Shum shared that Sotheby's sells over 50,000 lots annually, and each lot is currently getting a human-generated price estimate. Since the number of lots sold at Sotheby's is presently limited by how fast human experts can produce estimates, automating even a small piece of the manual process can drive up efficiency and have a significant impact on the business. To accomplish this, they are working on augmenting human appraisers' abilities to identify comparable works with machine vision and machine learning.

Shum explained that Sotheby's goal is not to replace human appraisers with machine learning, but instead to make them "super predictors." They hope to free appraisers from menial and repetitive tasks so they can focus more on the higher end of the market where the volume is lower and the accuracy of estimates becomes more critical due to higher dollar-value items.

For middle-market appraisals, Qamar wants to remove manual processes altogether and fully automate not only price estimation, but the entire end-to-end sales process. Qamar would also like to automatically generate up-to-date appraisals for every work that Sotheby's has ever sold. He hopes this might highlight works that offer surprisingly high returns, which Sotheby's can then take to the owners, encouraging them to sell at a profit. If the owners agree to sell, Sotheby's can also use their recommendation engine created by Shum and Qamar to automatically advertise those works to Sotheby's massive client database based on past buying patterns. The strategy is essentially to automate the construction of bidding wars to achieve the highest possible sale price for the seller.

As Shum explained it, in examining which features drive up the price of artworks at auction, the number of bidders was second only to the artist who created the work—and according to Shum, Sotheby's recommendation engine would enable them to influence the number of bidders heavily.

With these machine learning–driven capabilities, not only could Sotheby's estimate prices—they could influence them, achieving optimal returns for their clients at auction.

Conclusion

Many researchers have expressed beliefs that prices for art are perhaps unpredictable. Yet, there is a practical need to create presale estimates for all work sold at auction. Humans are currently more accurate than machines in crafting these estimates. However, machine learning models can potentially scale to appraise all art, not just the work going to auction at any given time. Automated pricing of all artworks on and off the market could drive up liquidity by providing additional information to buyers. To this end, Sotheby's has been exploring machine learning to price all past lots and to develop an automated end-to-end sales process.

Although economists might have advised against it, purchasing the duct tape banana may have been a good investment. The research suggests that extrinsic features like artists' reputation and the number of bidders are more important than features unique to the artwork. Cattelan's stunt has no doubt increased his global reputation, which should drive up the value of the work. And if the current owners wait just a few more years before selling, Sotheby's automated sales process might even trigger a bananas bidding war to assure the best possible return.

References

- Artsy. (2019). *The online art collector report 2019*.
http://files.artsy.net/documents/artsy_2019_onlineartcollectorreport.pdf
- Aubry, M., Kräussl, R., Manso, G., & Spaenjers, C. (2019). *Machines and masterpieces: Predicting prices in the art auction market*. SSRN. <https://doi.org/10.2139/ssrn.3347175>
- Ayub, R., Orban, C., & Mukund, V. (2017). *Art appraisal using convolutional neural networks*.
- Baumol, W. J. (1985). Unnatural value: Or art investment as floating crap game. *Journal of Arts Management and Law*, 15(3), 47-60. <https://doi.org/10.1080/07335113.1985.9942162>
- Beggs, A., & Graddy, K. (2009). Anchoring effects: Evidence from art auctions. *American Economic Review*, 99(3), 1027-1039. <https://doi.org/10.1257/aer.99.3.1027>
- Bjerg, M. (2018). *Sotheby's & Christie's hammer price vs estimate for 2016 and 2017*. Mearto.
<https://mearto.com/accuracy-of-sothebys-and-christies-estimations-revealed>

Bruno, B., Garcia-Appendini, E., & Nocera, G. (2016). Experience and brokerage in asset markets: Evidence from art auctions. *Financial Management*, 47(4), 833–864. <https://doi.org/10.1111/fima.12207>

Calfas, J. (08 Oct. 2018). Banksy shredded a piece of art that sold for \$1.4 million. Now it's worth double, according to an art expert. *Money*. <https://money.com/banksy-girl-with-a-balloon-self-destruct-double/>

Fraiberger, S. P., Sinatra, R., Resch, M., Riedl, C., & Barabási, A. L. (2018). Quantifying reputation and success in art. *Science*, 362(6416), 825–829. <https://doi.org/10.1126/science.aau7224>

Greene, K. (2017). Da Vinci's "Salvator Mundi" painting took a winding 60-year path from under \$200 to a record-breaking \$450 million. CNBC. <https://www.cnbc.com/2017/11/17/da-vincis-salvador-mundi-went-from-under-200-to-a-record-breaking-450-million.html>

Lang, M., & Maffett, M. (2011). Transparency and liquidity uncertainty in crisis periods. *Journal of accounting and economics*, 52(2-3), 101–125. <https://doi.org/10.1016/j.jacceco.2011.07.001>

McAndrew, C. (2019). *The art market. An Art Basel and UBS Report*. <https://www.artbasel.com/about/initiatives/the-art-market>

Mei, J., & Moses, M. (2002). Art as an investment and the underperformance of masterpieces. *American Economic Review*, 92(5), 1656–1668. <https://doi.org/10.1257/000282802762024719>

Pagano, M., & Röell, A. (1996). Transparency and liquidity: A comparison of auction and dealer markets with informed trading. *The Journal of Finance*, 51(2), 579–611. <https://doi.org/10.1111/j.1540-6261.1996.tb02695.x>

Plattner, S. (1998). A most ingenious paradox: The market for contemporary fine art. *American Anthropologist*, 100(2), 482–493. <https://doi.org/10.1525/aa.1998.100.2.482>

Pogrebin, R. (2019, December 8). Banana splits: Spoiled by its own success, the \$120,000 fruit is gone *The New York Times*. <https://www.nytimes.com/2019/12/08/arts/design/banana-removed-art-basel.html>

Powell, L., Gelich, A., & Ras, Z. W. (2019). Developing artwork pricing models for online art sales using text analytics. In T. Mihálydeák, F. Min, G. Wang, M. BanerjeeIvo, I. Düntsch, Z. Suraj, & D. Ciucci (Eds.), *International Joint Conference on Rough Sets* (pp. 480–494). Springer. https://doi.org/10.1007/978-3-030-22815-6_37

Reyburn, S. (2018, October 7). How Banksy's prank might boost his prices: "It's a part of art history." *The New York Times*. <https://www.nytimes.com/2018/10/07/arts/design/banksy-artwork->

[painting.html?smtyp=cur&smid=tw-nytimes](#)

Wamsley, L. (2019, May 16). Jeff Koons' "Rabbit" fetches \$91 million, auction record for work by a living artist. NPR. <https://www.npr.org/2019/05/16/723888420/jeff-koons-rabbit-fetches-91-million-auction-record-for-work-by-living-artist>

This article is © 2020 by Jason Bailey. The article is licensed under a Creative Commons Attribution (CC BY 4.0) International license (<https://creativecommons.org/licenses/by/4.0/legalcode>), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the authors identified above.