University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 18/2026 (512)

Painting Price: A Machine Learning Approach to Art Valuation. Proof of Concept and Market Structure Diagnosis

Kostiantyn Okhrimenko

Warsaw 2026

ISSN 2957-0506

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Painting Price: A Machine Learning Approach to Art Valuation. Proof of Concept and Market Structure Diagnosis

Kostiantyn Okhrimenko[1*]

[1] *University of Warsaw, Faculty of Economic Sciences*
[*] *Corresponding author: k.okhrimenko@student.uw.edu.pl*

Abstract: This paper investigates the feasibility of predicting art prices using machine learning methods applied to a dataset of 20,905 paintings and drawings scraped from the Artsper online marketplace. We test tree-based ensemble models (Decision Tree, Random Forest, XGboost) and deep learning architectures (MLP, CNN, Fusion) on both tabular metadata and hand-crafted image features. Results consistently show poor predictive performance across all model types and feature sets. We argue that this outcome is not a methodological failure but a substantive finding: it constitutes a diagnosis of the market structure of contemporary art. Drawing on hedonic pricing theory (Rosen 1974), the sociology of cultural fields (Bourdieu 1993), and the economics of valuation (Velthuis 2005; Beckert and Rössel 2013), we propose a three-layer model of art price determinants: physical attributes (observable and partially captured by models), visual-aesthetic features (observable but poorly quantifiable), and narrative-reputational capital (largely unobservable in cross-sectional platform data).

Keywords: art market, machine learning, hedonic pricing, art valuation, cultural economics, XGboost, CNN

JEL codes: Z11, C45, C53

1. Introduction

In 2019, Maurizio Cattelan sold one of his most famous and provocative works, Comedian, for 120,000 dollars. In 2024, the work was resold at Sotheby's in New York for 6.2 million dollars. At first glance, this may appear to be a familiar story of success in the contemporary art market, until one considers that Comedian consists of a banana attached with tape (Cascone 2019; Porterfield 2024). A similarly striking case occurred in October 2018, when a Banksy artwork partially shredded itself immediately after being sold at Sotheby's for 1.04 million British pounds. Although the gesture appeared to challenge the logic of the market, it had the opposite effect: the work's value increased, and the shredded piece was resold in 2021 for 18.6 million pounds (Sotheby's 2018; Reuter's 2021).

Examples such as these raise a broader question of what determines the value of an artwork and how that value is formed. The contemporary art market appears highly unpredictable, and this paper seeks to examine which aspects of price formation can be meaningfully modelled with machine learning methods. To this end, we train and evaluate a range of models, from classical tree-based algorithms to deep learning architectures processing raw image data, using a dataset of 20,905 artworks collected from Artsper. By conventional predictive standards, the results are consistently weak. We argue that this limited predictive performance is not merely a technical outcome, but one of the paper's main findings.

Poor model performance reveals a three-layer structure of price determination. Physical attributes of painting, which are size, age, medium, are measurable and models detect their signal. Visual and aesthetic characteristics are observable yet not easily captured in quantitative terms. The dominant layer which is narrative and reputational capital: the artist's biography, institutional affiliations, gallery prestige, critical reception is structurally absent from cross-sectional platform data. As Bourdieu (1993) showed, the art market is not a conventional commodity market but a symbolic field where cultural capital converts into economic value through mechanisms that leave no trace in a price listing.

The paper proceeds as follows. Section 2 develops the theoretical framework, tracing the economics of art valuation from hedonic pricing theory through cultural sociology. Section 3 describes the data collection and the Artsper dataset. Section 4 presents exploratory data analysis. Section 5 details the methodology and model architecture. Section 6 reports results. Section 7 offers a diagnosis of market structure and directions for future research.

2. The existing literature

2.1 Art as anomalous good

The art market cannot be compared directly to markets for ordinary goods. Its products are extremely heterogeneous: an original artwork is unique and can be created only once, while a copy cannot be considered a true substitute. At the same time, an artwork is not a good with a strictly defined utilitarian function, but rather a product with dual value, both cultural and economic. By acquiring an artwork, the consumer obtains an aesthetic experience, a cultural object, and a potential investment instrument (Throsby 1994).

At the same time, the art market is deeply shaped by uncertainty and information asymmetry (Akerlof 1970). Artists, dealers, experts, and buyers operate with unequal access to information, especially regarding an artwork's provenance, attribution, cultural significance, and market potential. Those with superior knowledge may exploit these asymmetries, but they may also reduce them through forms of certification, expertise, and disclosure (Greenwald 2019; Radermecker 2019).

2.2 Art as an investment

The investment literature asks whether art behaves like a financial asset and whether prices incorporate resale expectations. Stein (1977) modifies CAPM by adding service returns, stressing that paintings are both assets and consumption goods. Repeated-sales studies reach broadly similar conclusions: Pesando (1993), Mei and Moses (2002), and Renneboog and Spaenjers (2013) all find that art generally underperforms stocks, while its comparison with bonds is more mixed and accompanied by higher volatility. These studies also debate whether masterpieces outperform the rest of the market, with inconsistent conclusions across samples and methods.

A related question concerns dynamic valuation mechanisms within the art market itself. Beggs and Graddy (2009) document anchoring, showing that past sale prices affect subsequent prices even when they do not affect the probability of resale. Ursprung and Wiermann (2008) analyse the so-called death effect and argue that its direction depends on the interaction between scarcity and expectations about future reputation: prices may rise when supply becomes fixed, but this effect can be offset when buyers were previously expecting the artist's reputation to grow further.

2.3 Hedonic Pricing and Physical Attributes

The canonical economic framework for analyzing goods with multiple quality dimensions is the hedonic pricing model introduced by Rosen (1974). In this framework, the observed price of a good is a function of its bundle of characteristics: price = $f(x_1, x_2, \ldots, x_n)$, where each x_i is a measurable attribute and its implicit price is estimated from market data.

Applied to art, hedonic models have consistently identified physical attributes as significant price determinants. Czujack (1997) found a significant concave relationship between surface area and Picasso auction prices: larger works sold for more, but the marginal effect of additional size declined.

Barnett Newman explicitly theorized the role of scale. Newman's essay "The Sublime is Now" (1948) articulated the claim that large-format painting changes the phenomenological relationship between viewer and work: the canvas does not frame a scene to be contemplated from a distance but envelops the viewer, creating an immediate affective presence. What Newman described in philosophical terms, the market translates into price. An artist who wishes to be noticed at a salon paints large.

Machine learning can be understood as a non-parametric generalization of the hedonic model estimating the implicit price function without imposing a linear functional form. In this sense, tree-based models and neural networks are the natural methodological heirs of hedonic regression in settings where interactions and non-linearities may be important.

2.4 The second layer: visual features and aesthetic pleasure

A separate line of research examines how visual structure contributes to aesthetic response. Berlyne (1971) proposes an inverted-U relationship between stimulation and pleasure: works that are too simple or too complex should be less rewarding. Nadal et al. (2010) refine this idea by showing that visual complexity is multidimensional and linked to such elements as variety, organisation, asymmetry, and colour. Their results suggest that complexity cannot be captured by a single scalar measure.

Reber et al. (2004) approach aesthetic pleasure from the perspective of processing fluency, arguing that viewers prefer images that are easier to process. Symmetry, clarity, and figure-ground contrast can therefore contribute to beauty, even though beauty itself remains subjective. For the present study, this literature motivates the distinction between visual-aesthetic features that are observable but difficult to quantify and other layers of price determination that remain weakly observed or unobserved.

Although aesthetic valuation is inherently subjective, some visual characteristics of an image, including colour and structural composition, can still be represented quantitatively and incorporated into empirical analysis. In the present study, these characteristics are captured either through hand-engineered features or through feature embeddings learned automatically by convolutional layers in deep learning models.

2.4 Third layer: reputation, narrative and context

Physical attributes explain part of the price. The larger part is explained by what Bourdieu (1993) called the field of cultural production — the structured social space in which artists, critics, curators, dealers, and collectors compete for symbolic capital that is convertible into economic capital. In Bourdieu's framework, the market value of an artwork is not discovered but produced through the accumulated endorsements of legitimate cultural institutions.

The practical mechanics of this process were documented in detail by Velthuis (2005) in his ethnography of contemporary art dealers in New York and Amsterdam. Velthuis showed that galleries are not passive intermediaries who sell a pre-existing product but active producers of value: they construct narratives around artists, calibrate pricing to signal positioning within the field, and manage the social relationships through which reputations are built. A work sold through a prestigious gallery is not simply a work with a favorable distribution channel — it is a categorically different cultural object.

This value-producing role of galleries points more broadly to reputation as one of the key mechanisms through which prices are formed in the art market. Beckert and Rössel (2013) identify reputation as a central mechanism through which uncertainty is reduced, and value signals are created. This role is reinforced by intermediaries such as galleries and experts, whose presence helps buyers interpret quality under imperfect information. The institutional dimension of valuation is further visible in Oosterlinck and Radermecker (2021), who show that formal transparency regulation in France did not materially raise prices of autographed works after the introduction of the Markus Decree, suggesting that confidence in the art market is not easily engineered through regulation alone.

More granular studies show that reputation is multidimensional. Ertug et al. (2016) distinguish between museum-oriented and gallery-oriented reputation, arguing that different audiences reward different signals: museums respond more strongly to artistic quality, whereas galleries respond more to commercial potential. Braden and Teekens (2019) further separate reputation from status, showing that network position is particularly valuable at earlier career stages, while the artist's own reputation becomes more important later. In theoretical terms, Schönfeld and Reinstaller (2007) model how both artist and gallery reputation affect pricing in the primary market, with artist reputation strengthening buyers' attachment to a given artist, while gallery reputation operates through competitive intermediation.

Narrative and contextual information also play a substantial role in shaping evaluation. Jucker et al. (2014) show that positive narratives lead individuals to assess similar works more favourably. Swami (2013) finds that contextual information can raise appraisal, particularly for abstract and surrealist works, although Russell (2003) reports a more qualified pattern in which additional meaning does not necessarily increase pleasure and may even reduce it when it introduces disturbing context. Together, these studies suggest that valuation depends not only on the image but also on interpretive framing.

Time, effort, and meaning also affect valuation. Moore and West (2012) show that representational works tend to be preferred to abstract ones even when visual clarity is reduced, indicating the importance of extractable meaning. Kruger et al. (2004) demonstrate that perceived production effort increases both appreciation and stated monetary valuation. These findings align with the broader argument that art prices are shaped by narratives attached to works and by the beliefs viewers hold about artistic labour.

2.5 Empirical literature: econometrics and machine learning

Empirical work on art prices has followed two broad directions. The econometric tradition typically aims to identify interpretable determinants of value, often in carefully delimited market segments. Pownall and Graddy (2016), for example, test whether colour intensity and lightness explain auction prices for Andy Warhol prints, combining low-level image descriptors with hedonic regression. Candela et al. (2002) develop a much broader explanatory framework in which object-specific variables, auction characteristics, artist performance indicators, and macroeconomic controls are modelled separately across artistic schools and attribution levels. In this stream, the emphasis is on parameter interpretation and on isolating economically meaningful correlates of price.

The machine-learning literature shifts the focus from explanation to predictive performance. Strezoski and Worring (2017) show that deep learning is highly effective for image-based tasks such as classifying author, material, type, or date, but the translation of this success to price prediction is far less straightforward. Bailey (2020), Ayub et al. (2017), and Nho and Park (2019) all report that image-only models perform poorly relative to models based on metadata, indicating that price cannot be inferred from visual appearance alone.

Subsequent studies therefore move toward multimodal and sequence-based architectures. Aubry et al. (2023) show that combining metadata with visual embeddings improves prediction, although expert pre-sale estimates remain superior overall. De et al. (2025) further enrich fusion models with historical auction-price sequences and report substantial reductions in error compared with image-only CNNs. Liu (2022) models artwork prices as time-series data and reports improved prediction from a combined bidirectional and one-way LSTM architecture, while Li and Liu (2021) show that textual descriptions can be highly informative and may outperform image-only approaches.

Recent evidence also suggests that the usefulness of visual information depends on the informational environment. Mei et al. (2025) find that visual embeddings are particularly valuable when artworks lack sale history and therefore lack anchoring information. By contrast, Carugno et al. (2025) show that deep learning on tabular data does not automatically dominate simpler models: in their setting, CNNs applied to non-spatial tabular features are outperformed even by hedonic OLS. Taken together, the empirical literature suggests a clear distinction between econometric studies aimed at interpretable determinants and machine-learning studies aimed at prediction. It also indicates that predictive success depends on which layer of valuation is being observed: physical attributes are easiest to model, visual-aesthetic features require either handcrafted descriptors or learned embeddings, and narrative-reputational capital remains only weakly captured in most datasets.

3. Data Collection and Dataset Description

3.1 Data Collection Process

The basis of every effective predictive model is a consistent and reliable training dataset. Freely available art datasets present serious problems: outdated information, non-reliable or anonymous sources, and a lack of variables that may be important for analysis. Taking these factors into consideration, the author collected the data independently.

All information was gathered from Artsper, an online marketplace where artists and galleries sell original artworks. The categories selected were drawings and paintings.

At the initial stage, URLs to individual artwork pages were collected using the Parsehub application. When the individual URLs were collected, the Selenium library was used to extract information from HTML content via XPath and CSS selectors, with time delays added between requests to minimize server load. After the collection process, data was cleaned using pandas and re libraries, removing empty observations, artifacts, and converting variables to appropriate data types.

3.2 Dataset Description

After cleaning, the dataset contains 20,905 observations. Table 1 describes the variables; Table 2 presents basic descriptive statistics of the numeric variables.

Table 1. Description of the variables.

Variable name	Description	Hypothetical effect on price
price_numeric	price of an artwork measured in US dollars.	The independent variable.
gender_guessed	Gender of the artist, obtained from the names with gender_guesser library.	It might be hypothesised that male artists might charge more for their artworks.
Height	Height of the artwork measured in inches.	It might be suggested that artists charge more for bigger works, as such works need more materials and time to produce.
Width	Width of the artwork measured in inches.	See the description of "Height" variable.
age	The age of the work, measured in years.	Older artworks might be more appreciated than the newer ones.
years_selling	Number of years an artist or a gallery sell works through the website.	Higher number of years might serve as evidence of seller's reliability and trustworthiness. As a result, reliable seller might charge higher prices.

Variable name	Description	Hypothetical effect on price
Uniqueness	Binary variable identifying whether an artwork is unique or a copy.	Usually, unique works cost more than mass products.
is_signed	Binary variable telling whether a work is signed or not.	It might be suggested that signed work may cost more compared to not signed one, as signature serves as guarantee of uniqueness.
painting	Categorical variable showing type of the painting material.	Hypothetically, paintings made with oil cots more.
gallery	Binary variable identifying whether an artwork is sold by a gallery or an individual artist.	Galleries tend to charge more than individual artists due to higher costs and level of risk.
location	Categorical variable standing location of an artwork. The most popular countries in the sample were put in separate categories, whereas all other countries were encoded as “Other”.	Some countries might be more famous in terms of contemporary art, so artworks originated in these countries might cost more.

Table 2. Descriptive statistics of numeric variables

Variable	Mean	Minimum	Maximum	Std. dev.	Median
price_numeric	3176.32	61	750000	12247.7	1600
Height (in)	29.32	0.4	153	14.28	27.6
Width (in)	29.24	0.4	248	16.34	25.6
age (years)	6.74	0	125	14.33	3
years_selling	5.35	1	12	3.15	5

The dependent variable shows high variance: mean price is 3,176 dollars but the median is $1,600 and the maximum is $750,000. The sample is heterogeneous along all dimensions: from miniature works to quite huge canvases, from works created in 2025 to pieces over a century old.

4. Exploratory Data Analysis

Figure 1 presents the distribution of prices on logarithmic scale. The log transformation reveals an approximately normal distribution with a long right tail, confirming that a small number of highly priced works substantially inflate the mean.

Figure 1. Distribution of artwork prices on a logarithmic scale.

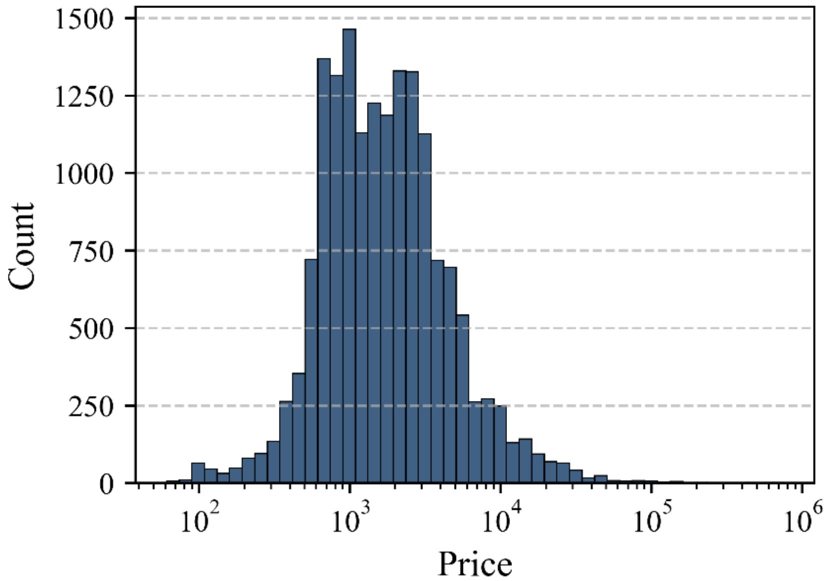


Figure 2 shows that the data set predominantly consists of contemporary works, with the majority created within the last decade. This reflects the composition of the Artsper platform, which focuses on living artists.

Figure 2. Distribution of artwork age variable.

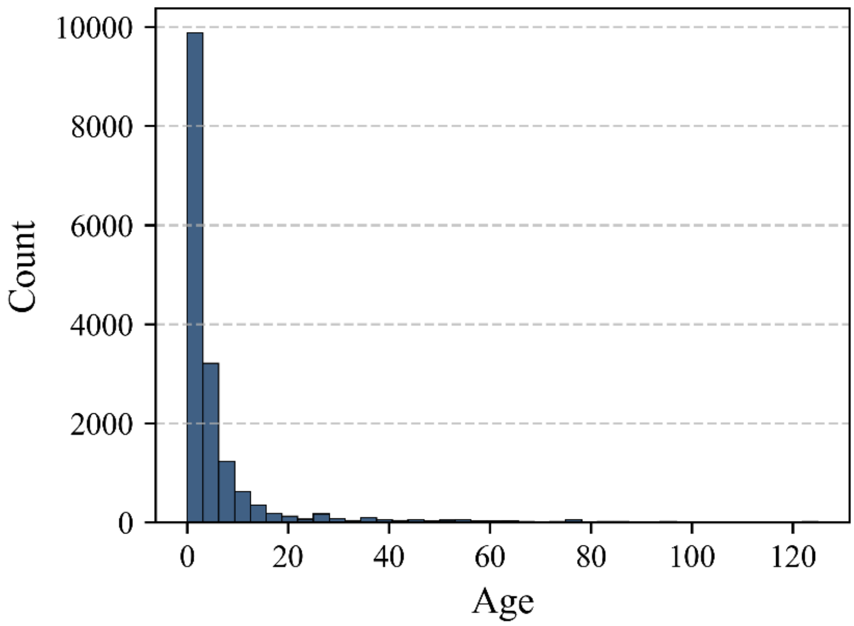


Figure 3 shows the distribution of years each seller has been active on Artsper. The bounded range (maximum 12 years, corresponding to the platform's age) limits the usefulness of this variable as a reputation proxy.

Figure 3. Distribution of years of selling variable.

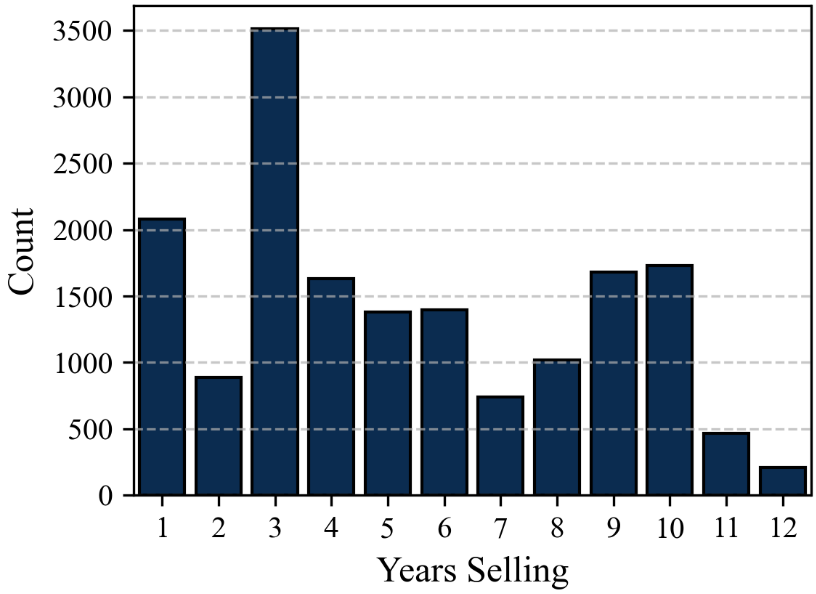


Figure 4 presents the distributions of height and width. Both variables are approximately normally distributed with a moderate right skew and few extreme outliers, consistent with the market norm of paintings produced for domestic and gallery interiors.

Figure 4. Distributions of height and width variables.

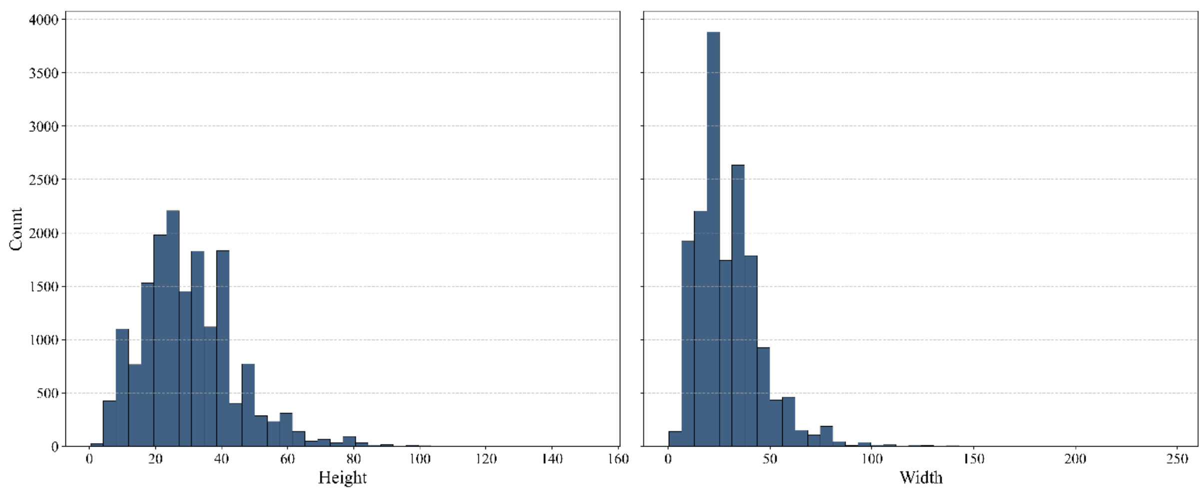


Figure 5 shows distributions of categorical variables. The majority of works are unique, signed, sold by galleries rather than individual artists, and originate from France (the location of Artsper's primary market). Oil and acrylic are the dominant materials.

Figure 5. Distributions of categorical variables.

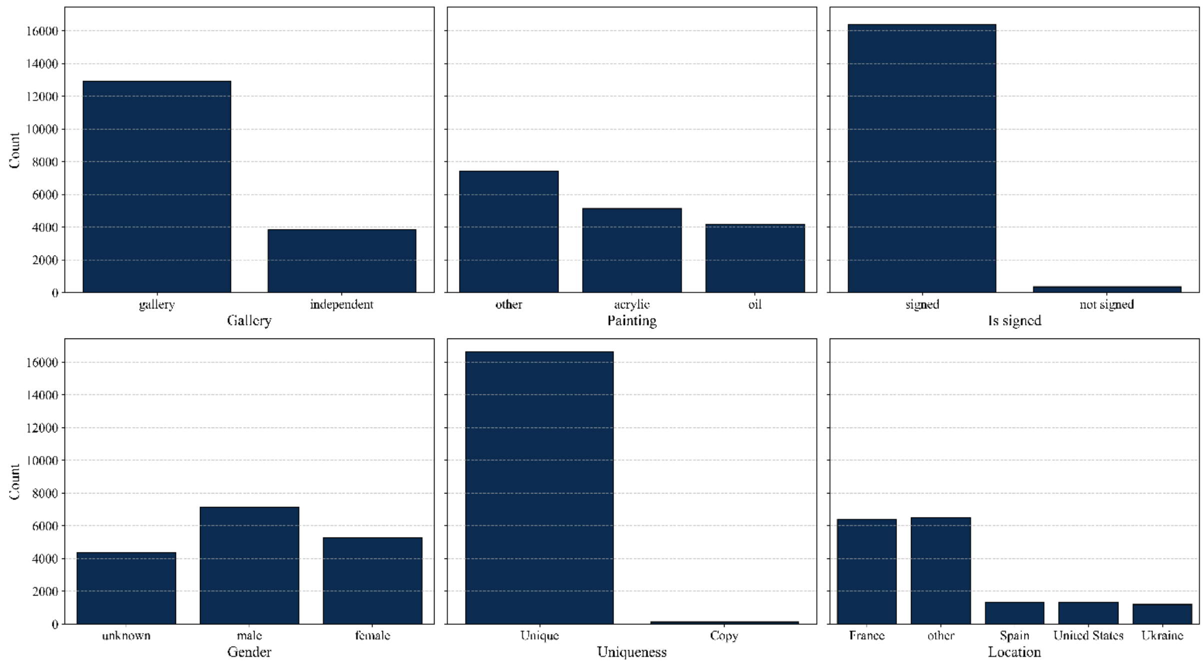
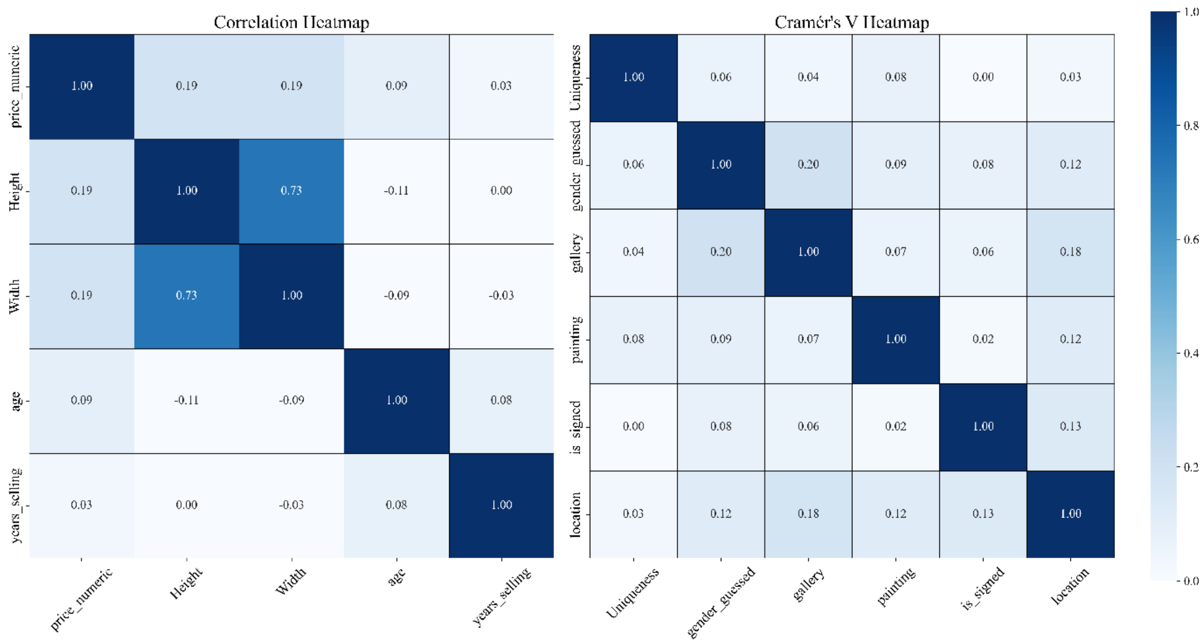


Figure 6 presents both the Pearson correlation matrix for numeric variables and the Cramér's V heatmap for categorical variables. No variable shows a strong linear relationship with price. The correlation between height and width is expected. The absence of strong linear correlations does not rule out non-linear relationships, which tree-based and deep learning models are designed to capture.

Figure 6. Correlation matrix and Cramér's V heatmap



.

In addition to standard EDA, we also utilized K-prototypes algorithm to explore the cluster structure of the images. All the observations from the train data were clustered using all numeric and categorical features (K-prototypes algorithm allows for it). Then, we identified the observations having the smallest distance to clusters centroids. The representative images are presented on Figure 7.

Figure 7. Nearest-to-centroids images.

5. Methodology

Art price prediction using machine learning is a quite difficult problem. Models trained on image data alone show poor performance (Ayub et al. 2017; Bailey 2020). Models combining images and metadata fail to outperform art experts (Aubry et al. 2023). Richer data sources — historical prices, text descriptions, artist career data — can improve performance but were not available in our setting (Liu 2022; Li and Liu 2021). Given these constraints, we adopt a multi-architecture approach designed to establish a baseline and identify where the predictive signal, however weak, originates.

Due to the need for a separate validation set for neural networks, the training sets for classical ML and deep learning models are not identical in size, which precludes direct metric comparison across families. We treat each family as a separate experiment and focus on cross-model patterns rather than absolute rankings.

5.1 Classical Tree-Based Models

We apply Decision Tree, Random Forest, and XGboost regressors to tabular metadata. Tree-based methods are chosen because they impose no linearity assumption (unlike OLS or ridge regression) and are not affected by multicollinearity (unlike distance-based methods such as SVM or KNN). Uddin and Lu (2024) confirm the consistent superiority of tree-based approaches for tabular data across benchmark comparisons.

Hyperparameters for each model were tuned using five-fold cross-validation with a preprocessing pipeline applied inside each fold to prevent data leakage. For the Decision Tree, we tuned maximum depth, minimum samples per split and leaf, features per split, and cost-complexity pruning parameter alpha. For Random Forest, the same tree parameters were tuned alongside the number of estimators. For XGboost, we additionally tuned learning rate, row and column subsampling fractions, minimum child weight, and gamma.

5.2 Hand-Crafted Image Features

We also apply tree-based models to features extracted from artwork images. Raw pixel values are uninformative at scale; we instead engineer features capturing color palette and spatial structure. After resizing all images to 128×128 pixels, we extract: Histograms of Oriented Gradients (HOG), capturing edge directions and structural patterns; Local Binary Patterns (LBP), encoding local micro-texture; RGB histograms, capturing color distribution in the standard color space; and HSV histograms, describing perceptual properties such as hue, saturation, and brightness.

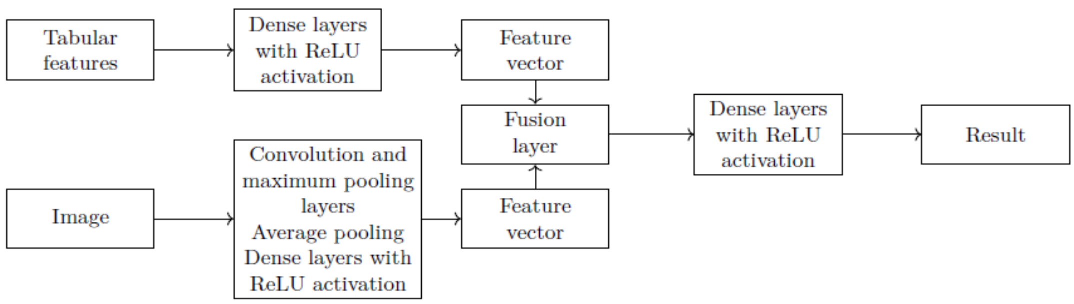
The resulting feature vector contains 8,302 dimensions per image. To prevent overfitting and reduce training time, we apply mutual information regression for feature selection — a method capable of detecting non-linear dependencies and more appropriate than linear correlation for this purpose (Frénay et al. 2013).

5.3 Deep Learning Architectures

We train three deep learning architectures: a Multi-Layer Perceptron (MLP) on tabular features, a Convolutional Neural Network (CNN) on images, and a Fusion model combining both.

The MLP consists of four hidden layers with 128, 64, 32, and 16 neurons respectively, with dropout after the third and fourth layers, and ReLU activations throughout. The CNN comprises three convolutional layers (32, 64, and 128 kernels of size 3×3) each followed by max pooling, an average pooling layer, and three dense layers of 128, 32, and 8 neurons. Fusion architecture concatenates the 8-dimensional output vectors of the MLP and CNN and passes them through three additional dense layers of 128, 64, and 8 neurons with dropout after the first two. All models use the Adam optimizer (Kingma and Ba 2015), which offers superior stability and efficiency compared to alternatives. The optimal number of training epochs is determined using a validation set comprising 25% of the training data

Figure 8. Schematic diagram of the Fusion model architecture.



For evaluation, we report normalized RMSE and normalized MAE on both training and test sets. Normalized RMSE is sensitive to outliers; normalized MAE provides a complementary view. A normalized RMSE of approximately 10% is generally considered indicative of acceptable prediction accuracy (Khoshvaght et al. 2025), though this threshold is domain-dependent.

5.4 Outlier Sensitivity Analysis

Given the highly skewed price distribution, we additionally train and evaluate all models on cleaned datasets from which outliers have been removed using Tukey's method. This eliminates 2,223 observations (10.6%), leaving 18,682. The outlier removal cannot be used as valid strategy because we omitted observations from both training and test sets; however, it allows to model the "perfect scenario" in which extreme heterogeneity is absent — providing an upper bound on achievable performance with the available features.

6. Results

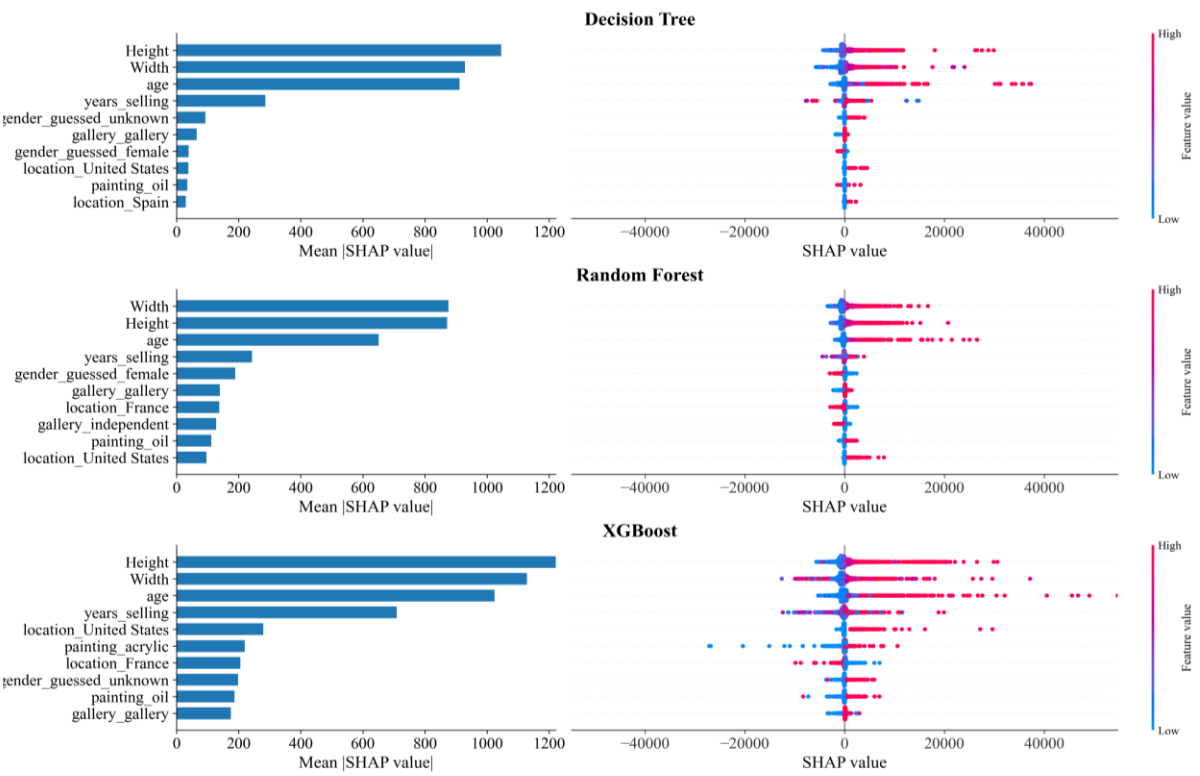
6.1 Classical Models: Tabular Data, Full Dataset

Table 3. Metrics for tree-based models on tabular data, full dataset.

Metric	Decision Tree	Random Forest	XGBoost
Train MSE	125 045 640.61	77 394 656.19	2 941 828.95
Test MSE	153 307 419.33	130 156 372.66	111 404 234.35
Train RMSE	11 182.38	8 797.42	1 715.18
Test RMSE	12 381.74	11 408.61	10 554.82
Train Normalized RMSE	3.52	2.77	0.54
Test Normalized RMSE	3.71	3.42	3.16
Train MAE	2 153.90	1 637.27	678.80
Test MAE	2 333.37	1 984.93	2 044.18
Train Normalized MAE	0.68	0.52	0.21
Test Normalized MAE	0.70	0.59	0.61

XGboost achieves the best absolute test performance but also shows the highest degree of overfitting, with a large gap between training and test metrics. No model achieves normalized RMSE approaching the 10% threshold. Normalized MAE, less sensitive to extreme values, also indicates poor performance. The SHAP analysis (Figure 9) shows that height, width, age, and years of selling are the most important predictors across all three models.

Figure 9. SHAP values for tree-based models trained on tabular features, full dataset.



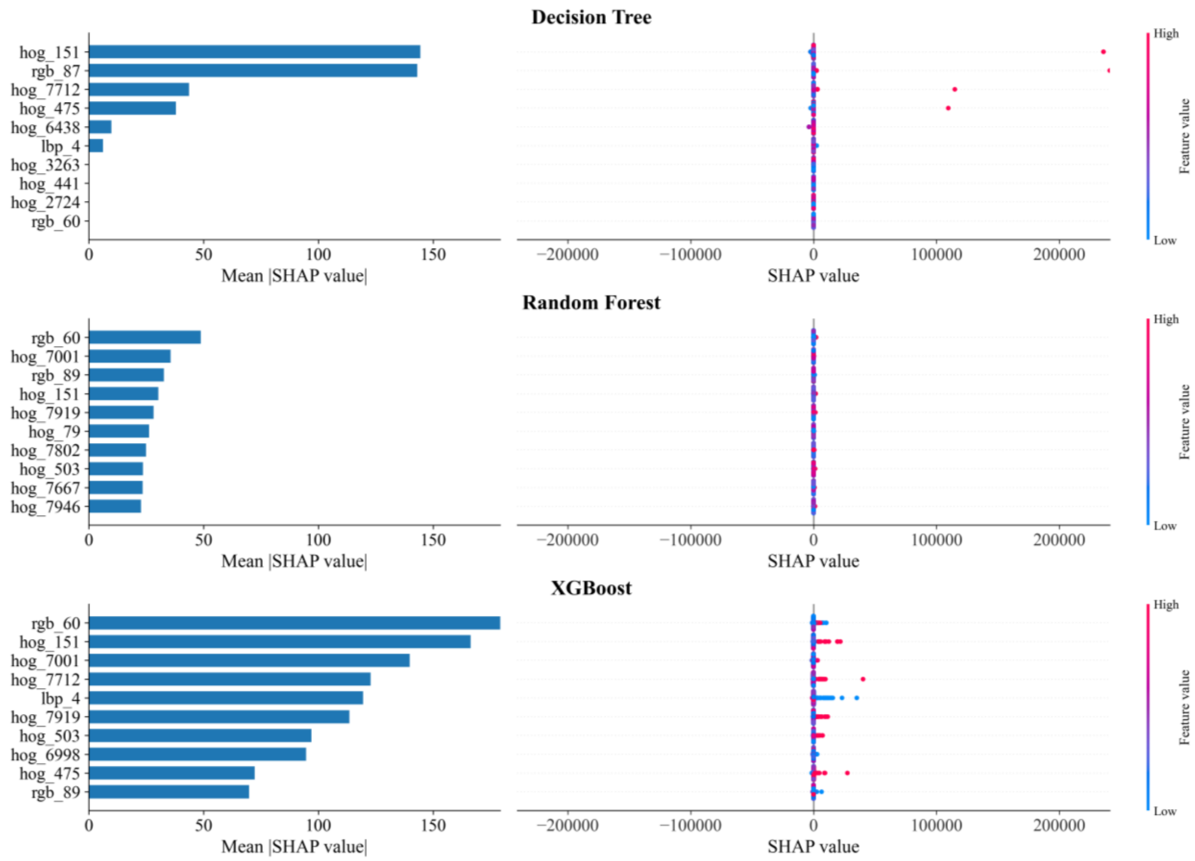
6.2 Classical Models: Image Features, Full Dataset

Table 4. Metrics for tree-based models on hand-crafted image features, full dataset.

Metric	Decision Tree	Random Forest	XGBoost
Train MSE	91 396 682.05	140 511 567.11	88 225 491.03
Test MSE	286 214 762.22	165 191 439.62	170 588 163.40
Train RMSE	9 560.16	11 853.76	9 392.84
Test RMSE	16 917.88	12 852.68	13 060.94
Train Normalized RMSE	3.01	3.73	2.96
Test Normalized RMSE	5.07	3.85	3.91
Train MAE	2 624.25	2 764.38	2 646.00
Test MAE	3 131.78	2 933.02	3 015.53
Train Normalized MAE	0.83	0.87	0.83
Test Normalized MAE	0.94	0.88	0.90

Models trained on image features perform worse than those trained on tabular metadata. Random Forest and XGboost show similar performance; the Decision Tree is worst and most overfit. SHAP values (Figure 10) show no single feature group dominating, consistent with the interpretation that visual features carry minimal systematic price signal.

Figure 10. SHAP values for image feature models, full dataset.



6.3 Deep Learning Models: Full Dataset

Table 5. Metrics for deep learning models, full dataset.

Metric	MLP	CNN	Fusion
Train MSE	116 269 095.44	126 540 994.05	116 476 668.87
Validation MSE	220 129 170.84	230 219 628.20	220 133 758.44
Test MSE	155 914 514.65	168 439 331.85	155 571 421.24
Train RMSE	10 782.81	11 249.04	10 792.44
Validation RMSE	14 836.75	15 172.99	14 836.91
Test RMSE	12 486.57	12 978.42	12 472.83
Train RMSE ratio	3.47	3.62	3.47
Validation RMSE ratio	4.38	4.48	4.38
Test RMSE ratio	3.74	3.89	3.74
Train MAE	1 648.09	2 218.26	1 660.94
Validation MAE	1 986.67	2 528.19	2 001.62
Test MAE	1 884.66	2 451.96	1 898.78
Train MAE ratio	0.53	0.71	0.53
Validation MAE ratio	0.59	0.75	0.59
Test MAE ratio	0.56	0.73	0.57

The MLP trained on tabular features achieves the best performance among deep learning models and is comparable to but does not substantially outperform XGboost. The CNN and Fusion model do not improve on the tabular baseline. This is consistent with findings by Ye et al. (2024) that deep learning does not systematically outperform tree-based methods on small and medium-sized tabular datasets.

6.4 Outlier-Cleaned Dataset

After applying Tukey's method to remove outliers from the training and test sets, performance improves substantially across all model families.

Table 6. Metrics for tree-based models on tabular data, cleaned dataset.

Metric	Decision Tree	Random Forest	XGBoost
Train MSE	638 426.69	146 557.04	257 348.57
Test MSE	1 090 242.31	811 017.21	806 425.01
Train RMSE	799.02	382.83	507.30
Test RMSE	1 044.15	900.56	898.01
Train Normalized RMSE	0.43	0.21	0.28
Test Normalized RMSE	0.56	0.48	0.48
Train MAE	533.17	244.78	324.42
Test MAE	703.04	582.40	589.63
Train Normalized MAE	0.29	0.13	0.18
Test Normalized MAE	0.38	0.31	0.32

Table 7. Metrics for tree-based models on image features, cleaned dataset.

Metric	Decision Tree	Random Forest	XGBoost
Train MSE	1 743 741.17	1 007 472.23	580 854.51
Test MSE	1 858 843.75	1 790 466.11	1 777 899.43
Train RMSE	1 320.51	1 003.73	762.14
Test RMSE	1 363.39	1 338.08	1 333.38
Train Normalized RMSE	0.72	0.54	0.41
Test Normalized RMSE	0.73	0.72	0.72
Train MAE	1 049.33	785.69	602.23
Test MAE	1 076.95	1 052.41	1 044.85
Train Normalized MAE	0.57	0.43	0.33
Test Normalized MAE	0.58	0.57	0.56

Table 8. Metrics for deep learning models, cleaned dataset.

Metric	MLP	CNN	Fusion
Train MSE	1 118 407.57	1 971 048.38	1 241 190.37
Validation MSE	1 160 847.39	2 021 747.62	1 266 733.60
Test MSE	1 178 375.00	2 074 088.85	1 288 982.48
Train RMSE	1 057.55	1 403.94	1 114.09
Validation RMSE	1 077.43	1 421.88	1 125.49
Test RMSE	1 085.53	1 440.17	1 135.33
Train RMSE ratio	0.58	0.76	0.61
Validation RMSE ratio	0.58	0.77	0.61
Test RMSE ratio	0.58	0.77	0.61
Train MAE	687.58	1 017.52	732.75
Validation MAE	710.95	1 031.50	749.21
Test MAE	710.23	1 040.03	750.14
Train MAE ratio	0.37	0.55	0.40
Validation MAE ratio	0.38	0.56	0.40
Test MAE ratio	0.38	0.56	0.40

The improvement on the cleaned dataset is informative: it confirms that models can learn from the available features when extreme outliers do not dominate the loss function. However, the practical value of outlier removal for art price prediction is limited, since the most expensive works are the outliers. The cleaned data results represent a best case rather than a realistic scenario.

SHAP values on the cleaned dataset (Figures 11 and 12) show the same pattern as the full dataset: physical dimensions and age dominate in tabular models; image features show diffuse importance.

Figure 11. SHAP values for tabular models, cleaned dataset.

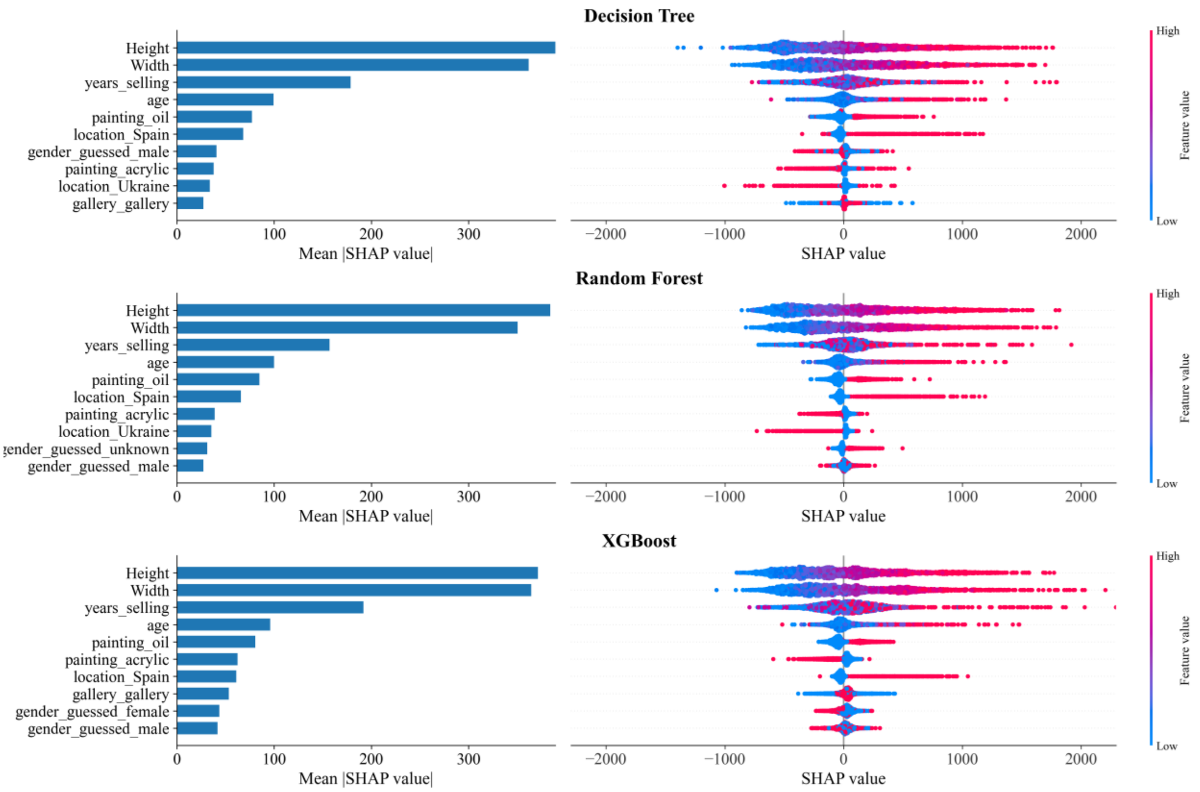
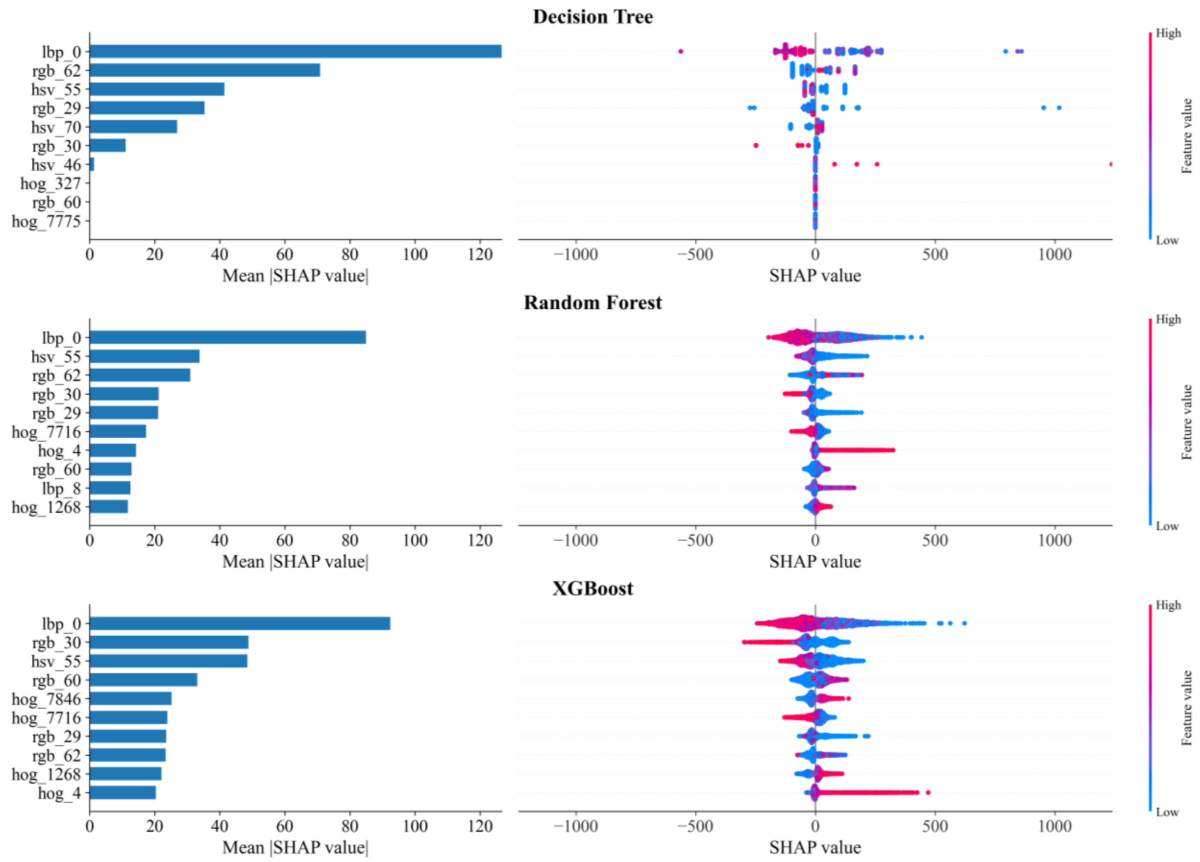


Figure 12. SHAP values for image feature models, cleaned dataset.



7. Discussion: A Diagnosis of Market Structure

7.1 Three Layers of Price Determination

The pattern of results across all model families is consistent and interpretable. We propose a three-layer model of art price determinants that explain both what models detect and what they miss.

The first layer is physical: dimensions, medium, age, and the seller's tenure on the platform. These attributes are measurable, present in the data, and carry a genuine price signal. SHAP analysis confirms that height, width, age, and years of selling dominate model predictions across all tree-based architectures. This is consistent with the hedonic pricing literature and with the aesthetics of scale developed by Newman: physical presence has economic value that markets reliably price.

The second layer is visual aesthetic: color palette, compositional structure, stylistic distinctiveness. These attributes are in principle visible in the image data, but our results confirm what prior literature has found that they do not yield reliable price signals neither from hand-crafted features nor from CNN-created embeddings. Whether this reflects the overall unimportance of visual attributes for price, or simply the inadequacy of our feature engineering, cannot be fully determined from the current analysis. More powerful image representations (pre-trained convolutional features, CLIP embeddings) may perform differently.

The third and dominant layer is narrative-reputational: artist biography and career trajectory, gallery prestige, exhibition history, critical reception, auction provenance. This layer is structurally absent from Artsper listing data. Velthuis (2005) documented how this layer is constructed through the daily work of dealers and institutions; Beckert and Rössel (2013) showed it is the foundation on which prices in singular-goods markets are stabilized. Our models cannot learn what is not in the data. The poor performance is not a failure of algorithm design but a consequence of data structure.

7.2 The Irreducibility of Narrative

Banksy's self-shredding canvas is a useful thought experiment for this conclusion. At the moment of the auction hammer, a standard model would have predicted a price based on dimensions, medium, signature status, and gallery affiliation. These features were unchanged when the shredder activated. What changed was the narrative: the work became the artwork that was destroyed, a performance piece, a comment on the market itself. The price tripled over three years. No feature engineering applied to the original listing data would have predicted this.

This logic can be applied not only to extreme cases. In fact, the same painting sold by an unknown artist in their first year on Artsper carries a different narrative weight than the same physical object sold by an artist with a major retrospective, museum acquisitions, and a prominent gallery representation. The difference is entirely in the third layer. Our proxy, which is years of selling, captures only a shadow of this.

7.3 Limitations and Directions for Future Research

The findings of this study suggest several concrete directions for future research that could substantially improve predictive performance.

First, text embeddings from artwork descriptions and artist biographies would provide access to part of the narrative layer. Large language models trained on art criticism and auction catalogues could generate representations of narrative capital that are currently absent from our feature set. Li and Liu (2021) and Mei et al. (2025) demonstrate the value of this approach.

Second, artist career metadata which is education institution, number of solo exhibitions, presence in public collections, auction history would operationalize the reputational layer more directly than any proxy available in platform listing data.

Third, panel data on price histories would allow models to learn the dynamics of reputation accumulation rather than treating price as a cross-sectional outcome. An artist's price trajectory over time encodes information about their reputational trajectory that no single observation can provide.

Fourth, more powerful image representations, particularly features extracted from pre-trained vision transformers or CLIP-style models may capture visual aspects that our hand-crafted features or the ones extracted with custom CNNs miss. The question of whether visual attributes independently predict price, controlling for artist identity, remains open.

Fifth, the sample composition of Artsper predominantly contemporary works, heavily French, with prices ranging from 61 to 750,000 dollars limits generalizability. A study combining multiple platforms or incorporating auction data from Sotheby's and Christie's would allow better coverage of the market's upper end, where the narrative layer is most determinative.

8. Conclusion

This paper aimed at predicting art prices using machine learning and found that it cannot be reliably performed that way. Poor model performance should be treated rather as a finding than a methodological error. The fact that various models trained on both physical and visual features failed to effectively predict the prices gives insight into the real structure of the art market.

Art price is not only a property of the canvas. It is the outcome of a social process in which physical attributes constitute only the first, most measurable layer. The dominant layers — visual distinctiveness, narrative capital, reputational positioning — are either poorly quantifiable or structurally absent from cross-sectional marketplace data. Bourdieu described the art market as a field were symbolic capital converts into economic one through mechanisms largely invisible to standard economic analysis. Our models confirm this empirically: they can read the painting but not the story the painting tells.

The practical implication is clear. A useful art price prediction model requires richer data: text, biography, exhibition history, auction provenance. The algorithmic challenge is secondary to the data challenge. Future work that assembles this richer data perhaps combining web scraping, artist database linkage, and NLP on auction catalogue descriptions would be better positioned to test the limits of machine prediction in this market.

What the current study contributes is a rigorous baseline, a clear articulation of where prediction fails and why, and a theoretical framework the three-layer model of physical, visual, and narrative determinants that can guide future data collection and model design. Our results suggest that predicting the price of art requires not only understanding the physical and visual features of an object but also the narrative and story behind it.

References

Akerlof, G.A., 1970. The market for “lemons”: Quality uncertainty and the market mechanism. The Quarterly Journal of Economics, 84(3), 488–500. https://doi.org/10.2307/1879431

Aubry, M., Kräussl, R., Manso, G., Spaenjers, C., 2023. Biased auctioneers. The Journal of Finance, 78(2), 795–833. https://doi.org/10.1111/jofi.13203.

Ayub, R., Orban, C., Mukund, V., 2017. Art appraisal using convolutional neural networks [Unpublished project report]. Stanford University. https://cs229.stanford.edu/proj2017/final-reports/5229686.pdf

Bailey, J., 2020. Can Machine Learning Predict the Price of Art at Auction? Harvard Data Science Review, 2(2). https://doi.org/10.1162/99608f92.7f90ce96.

Beckert, J., Rössel, J., 2013. The price of art: Uncertainty and reputation in the art field. European Societies, 15(2), 178–195. https://doi.org/10.1080/14616696.2013.767923

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.

Berlyne, D.E., 1971. Aesthetics and psychobiology. Appleton-Century-Crofts. https://doi.org/10.2307/429334.

Bourdieu, P., 1993. The field of cultural production: Essays on art and literature. Columbia University Press.

Braden, L.E.A., Teekens, T., 2019. Reputation, status networks, and the art market. Arts, 8(3), Article 81. https://doi.org/10.3390/arts8030081

Candela, G., Figini, P., Scorcu, A.E., 2002, June 13–15. Hedonic prices in the art market: A reassessment [Paper presentation]. ACEI Conference, Rotterdam, The Netherlands.

Carugno, S., Fedderke, J.W., 2025. Machine learning algorithms and fine art pricing: Comparative performance to regression results for the South African market, 2009–22. Expert Systems with Applications, 270, Article 126468. https://doi.org/10.1016/j.eswa.2025.126468

Cascone, S., 2019, December 4. Maurizio Cattelan’s duct-taped banana artwork sold at Art Basel Miami Beach’s VIP preview for $120,000. Artsy. https://www.artsy.net/article/artsy-editorial-maurizio-cattelans-duct-taped-banana-work-sold-art-basel-miami-beachs-vip-preview

Czujack, C., 1997. Picasso paintings at auction, 1963–1994. Journal of Cultural Economics, 21, 229–247. https://doi.org/10.1023/A:1007453919073

De, S., Sargam, S., Shrivastava, P., Mahesh, J., Rao, P.H., Raj, N., Lahane, P.S., 2025. ART MARKET PREDICTIONS THROUGH DEEP LEARNING. ShodhKosh: Journal of Visual and Performing Arts, 6(4s), 255–265. https://doi.org/10.29121/shodhkosh.v6.i4s.2025.6847

Ertug, G., Yogev, T., Lee, Y.G., Hedström, P., 2016. The art of representation: How audience-specific reputations affect success in the contemporary art field. Academy of Management Journal, 59(1), 113–134. https://doi.org/10.5465/amj.2013.0621.

Frénay, B., Doquire, G., Verleysen, M., 2013. Is mutual information adequate for feature selection in regression? Neural Networks, 48, 1–7. https://doi.org/10.1016/j.neunet.2013.07.003

Greenwald, D.S., 2019. Used cars and canvases: Information economics, art history, and the art market. American Art, 33(3), 5–9. https://doi.org/10.1086/707468

Jucker, J.-L., Barrett, J.L., Wlodarski, R., 2014. "I just don't get it": Perceived artists' intentions affect art evaluations. Empirical Studies of the Arts, 32(2), 149–182. https://doi.org/10.2190/EM.32.2.c.

Khoshvaght, H., Permala, R.R., Razmjou, A., Khiadani, M., 2025. A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction. Journal of Environmental Chemical Engineering, 13(6), Article 119675. https://doi.org/10.1016/j.jece.2025.119675

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, San Diego, CA, United States. https://arxiv.org/abs/1412.6980

Kruger, J., Wirtz, D., Van Boven, L., Altermatt, T.W., 2004. The effort heuristic. Journal of Experimental Social Psychology, 40, 91–98. https://doi.org/10.1016/S0022-1031(03)00065-9

Li, B., Liu, T., 2021. An analysis of multi-modal deep learning for art price appraisal, in: 2021 IEEE International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking (ISPA/BDCloud/SocialCom/SustainCom) (pp. 1509–1513). IEEE. https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00203

Liu, C., 2022. Prediction and analysis of artwork price based on deep neural network. Scientific Programming, 2022, Article ID 7133910. https://doi.org/10.1155/2022/7133910.

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

Mei, J., Moses, M., Wälty, J., Yang, Y., 2025. Deep learning for art market valuation. doi: https://doi.org/10.48550/arXiv.2512.23078

Moore, K., West, A., 2012. Global Perception, Meaning and Aesthetic Appreciation, EMPIRICAL STUDIES OF THE ARTS. 30(1) 23-38. https://doi.org/10.2190/EM.30.1.d

Nadal, M., Munar, E., Marty, G. and Cela-Conde, C.J., 2010. Visual complexity and beauty appreciation: Explaining the divergence of results. Empirical Studies of the Arts, 28(2). 173–191. doi: 10.2190/EM.28.2.d

Newman, B., 1948. The sublime is now. The Tiger's Eye, 1(6), 51–53.

Nho, H., Park, H., 2019. The art of predicting art auction price [Course project report]. Stanford University. https://cs230.stanford.edu/projects_fall_2019/reports/26261328.pdf.

Oosterlinck, K., & Radermecker, A.-S. V. (2021). Regulation or reputation? Evidence from the art market (Working Papers CEB No. 21-006). Université Libre de Bruxelles.

Pesando, J.E., 1993. Art as an investment: The market for modern prints. American Economic Review, 83(5), 1075–1089

Porterfield, C., 2024, November 21. Maurizio Cattelan's banana sells for $6.2m at Sotheby's. The Art Newspaper. https://www.theartnewspaper.com/2024/11/21/maurizio-cattelan-banana-sothebys-6-million-auction-comedian

Pownall, R.A.J., Graddy, K., 2016. Pricing color intensity and lightness in contemporary art auctions. Research in Economics, 70(3), 412–420. https://doi.org/10.1016/j.rie.2016.06.007

Radermecker, A.-S. V.E., 2019. Artworks without names: An insight into the market for anonymous paintings. Journal of Cultural Economics, 43(3), 443–483. https://doi.org/10.1007/s10824-019-09344-5

Reber, R., Schwarz, N., Winkielman, P., 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? Personality and Social Psychology Review, 8(4), 364–382. https://doi.org/10.1207/s15327957pspr0804_3

Renneboog, L., Spaenjers, C., 2013. Buying beauty: On prices and returns in the art market. Management Science, 59(1), 36–53. https://doi.org/10.1287/mnsc.1120.1580

Reuters, 2021, October 14. Half-shredded Banksy fetches $25.4 mln on return to Sotheby's. https://www.reuters.com/lifestyle/5-8-million-half-shredded-banksy-try-sothebys-2021-10-14/

Rosen, S., 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy, 82(1), 34–55. https://doi.org/10.1086/260169

Russell, P.A., 2003. Effort after meaning and the hedonic value of paintings. British Journal of Psychology, 94(1), 99–110. https://doi.org/10.1348/000712603762842138.

Schönfeld, S., Reinstaller, A., 2007. The effects of gallery and artist reputation on prices in the primary market for art: a note. Journal of Cultural Economics, 31(2), 143–153. https://doi.org/10.1007/s10824-007-9031-1

Sotheby's, 2018, October 5. Sotheby's gets Banksy'ed at contemporary art auction in London. https://www.sothebys.com/en/articles/sothebys-gets-banksyed-at-contemporary-art-auction-in-london

Stein, J.P., 1977. The monetary appreciation of paintings. Journal of Political Economy, 85(5), 1021–1036. https://www.jstor.org/stable/1830343

Strezoski, G., Worring, M., 2017. OmniArt: Multi-task Deep Learning for Artistic Data Analysis. https://doi.org/10.48550/arXiv.1708.00684

Swami, V., 2013. Context matters: Investigating the impact of contextual information on aesthetic appreciation of paintings by Max Ernst and Pablo Picasso. Psychology of Aesthetics, Creativity, and the Arts, 7(3), 285–295. https://doi.org/10.1037/a0030965

Throsby, D., 1994. The production and consumption of the arts: A view of cultural economics. Journal of Economic Literature, 32(1), 1–29.

Uddin, S., Lu, H., 2024. Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data. PLOS ONE, 19(4), Article e0301541. https://doi.org/10.1371/journal.pone.0301541

Ursprung, H.W., Wiermann, C., 2008. Reputation, price, and death: An empirical analysis of art price formation (CESifo Working Paper No. 2237). CESifo.

Velthuis, O., 2005. Talking prices: Symbolic meanings of prices on the market for contemporary art. Princeton University Press.

Ye, H.-J., Liu, S.-Y., Cai, H.-R., Zhou, Q.-L., & Zhan, D.-C., 2024. A closer look at deep learning methods on tabular datasets. arXiv. https://doi.org/10.48550/arXiv.2407.00956

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl
ISSN 2957-0506