Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-16

Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on Price Prediction of Calligraphy and Painting Artworks Based on Machine Learning

Ai Ziyuan†

1. Afiliación. School of International Studies, Beijing Normal University-Hong Kong Baptist University, Beijing, 519000

Submission Info

Communicated by Z. Sabir
Received June 22, 2024
Accepted October 11, 2024
Available online November 22, 2024

Abstract

With the rapid development of the Chinese economy, people's consumption structure has changed, and the investment in art and the tendency towards finance are increasing. The 13th Five Year Plan for the Development of National Cultural Relics clearly proposes to increase the protection of cultural relics and take multiple measures. Artworks have long been an essential investment target in the international capital market and belong to high value-added assets. In order to accurately predict the prices of calligraphy and painting artworks, this study proposes a machine learning based art price prediction method by combining ARCH model and Random Forest Regression (RFR) algorithm. Firstly, this study utilizes the ARCH model to capture the agglomeration effect of volatility in price time series, and reveals the inherent laws and volatility characteristics of price changes. Secondly, the Random Forest Regression (RFR) algorithm is introduced for price prediction. By constructing multiple decision trees and synthesizing their results, the non-linear and high-dimensional features of the data are effectively addressed. Research has shown that the RFR model exhibits high accuracy and stability in predicting art prices with multiple variables and complex relationships.

Keywords: calligraphy and painting art market; GARCH model; prediction of the value of calligraphy and painting artworks; Random Forest model
AMS 2020 codes: 91B24

†Corresponding author.
Email address: 568147275@qq.com

ISSN 2444-8656
http://doi.org/10.2478/amns-2024-3421

1 Introduction

In today's global art market, art as a unique investment asset, its price fluctuations are usually more violent and uncertain than other markets, and this special nature determines that its value is difficult to assess, and its liquidity is also poor, many collectors and investors are unable to rely on the market value of this indicator intuitively and effectively to objectively assess the value of works of art. However, with the year-on-year growth in the volume of art transactions and the rapid development of Internet technology, the transparency of information in the art market has been gradually improved, providing a broader data base for art price prediction. However, art value assessment is still a complex and challenging task, the value of artwork is often subjective, the volume of transactions in the market is relatively small, and even in major cases the art market may be extreme risk [1]. It is not only affected by the reputation of the artist, the quality of the artwork, the scarcity of the work, the historical record of transactions and other intrinsic factors, but also closely related to the market supply and demand, the macroeconomic environment, collection trends, cultural preferences and other external factors [2,3]. Therefore, clarifying the formation mechanism and fluctuation characteristics of artwork prices and constructing an accurate and effective artwork price prediction model are of great significance for investors' decision-making, artwork risk management and market trend analysis.

In recent years, the successful forecasting applications of time series analysis models and machine learning techniques in the economic and financial fields have opened up new paths for solving artwork price forecasting challenges. Time series models such as Autoregressive Moving Average Model (ARMA), ARIMA, and Seasonal SARIMA can effectively capture the time-dependent and cyclical fluctuations of artwork price data and improve the accuracy and robustness of forecasts. Meanwhile, with the enhancement of big data and computer technology, machine learning and deep learning algorithms such as support vector machine (SVM), random forest (RF), artificial neural network (ANN), and long and short-term memory network (LSTM) [4] have shown excellent performance in complex pattern recognition and non-linear relationship modelling [5], introducing a deeper analytical perspective for artwork price prediction.

This paper explores the dynamic prediction and market trend analysis of Chinese calligraphy and painting artwork prices by integrating time series analysis techniques and machine learning algorithms. Based on the extensive collection and collation of historical artwork transaction data, market macro indicators, Chinese painting and calligraphy artwork characteristic indicators and other multi-source heterogeneous data in recent years, a multi-dimensional analysis framework is constructed to deeply analyse the key factors affecting the value of artwork, improve the generalization ability of the prediction model, and provide a scientific basis for the investment strategy of artwork.

2 Literature review

1) Research on the price formation mechanism and price fluctuation characteristics of works of art

Art has a variety of characteristics such as consumption, investment, heterogeneity, non-standardisation, poor liquidity, etc. The traditional economic theory can not reasonably explain the investment behaviour of the art market, and the conventional valuation means are difficult to carry out objective and fair valuation [6]. Firstly, the business structure of the art market will develop in the direction of more diversification, and it is necessary to pay attention to the development of the art market trading platform and the construction of the credit function, and to introduce a complete market competition and elimination system in the operation link [7]. Secondly, art consumers show different elasticity of supply and demand for different artworks, the change of art supply is analysed from the circulation link, and the increase or decrease of supply can reflect the increase or decrease of the number of artworks in circulation in the market and the good or bad prospects of the development of the art market [8]. Again, the fluctuation of artwork prices mainly refer to two aspects: horizontal type comparison and vertical level comparison. The research of price volatility characteristics is more in the stock market and futures market, such as with the help of GARCH model [9], EGARCH model, DCC-GJR-GARCH model and other methods to explore the asymmetry, seasonality, cyclicality of price volatility of stock market and futures market, etc., and the research of price volatility in the field of works of art is still to be improved [10]. Finally, for the pricing problem of artworks, it can be explored from three perspectives: quantitative, qualitative and game analysis, to establish an accurate artwork price index or artist classification index, and to predict the future development trend of the price of calligraphy, painting, ceramics and other artworks [11,12].

2) Research on the prediction of the market value of works of art

Market value, as a key indicator for assessing the investment potential of artworks, provides a direct and effective valuation benchmark for auction house professionals, art investors, collectors and hobbyists, ensuring the rationality and objectivity of investment decisions [13]. Compared with the domestic, the international market in the field of fine pricing of artworks started earlier and developed more maturely. The main methods are: one is to use the "benchmark works method", that is, by analysing the past landmark works of the same artist's transaction prices, the construction of art price indexes to analyse, in order to infer the market value of the artist's other works, but the method lacks stability [14]. The "benchmark work method" is based on the logic that an artist's representative work, due to its uniqueness, artistic achievement or historical significance, often establishes a price benchmark in the market, and the value of similar works can be used as a comparison. The second is the repeat sales pricing method, based on the idea of repeat sales to establish the art price index system, but only when the art market is hit, the repeat sales pricing method plays an effective [15]. At present, the relevant scholars for the quantitative study of art value prediction method is still mainly the traditional econometric and statistical methods. Statistical methods are mostly based on information-poor, data-poor grey correlation, wavelet neural network approximation of correlation variable weights, data envelopment analysis DEA, fuzzy hierarchical analysis [16], multi-modal price and other methods for artwork price prediction [17-18]. However, these methods are quantitative without text data prediction, and the rise of machine learning provides a new development direction for artwork classification and pricing [19,20,21]. Compared with the traditional discriminant analysis, machine learning not only has higher accuracy, but also has the ability to continuously learn as the database is updated [22,23], there are now many scholars who have introduced Random Forest into the field of price prediction, and the application area of Random Forest is still expanding.

3 Analysis of the state of art transactions

1) Market size of art transactions

From the scale of China's art market transactions from 2010 to 2022 in Figure 1, the total auction turnover fluctuated, but the overall trend of stability. 2022 by the macro-environmental impact, the annual auction turnover of cultural relics and works of art declined to 16.465 billion yuan, the lowest point since 2010, but thanks to the development of online auctions, the turnover was only reduced by 2.99% compared with the previous year. In addition, the settlement rate of 2022 turnover increased by 8.04 percentage points compared with the same period of the previous year. The average price per lot fluctuated less, falling from RMB 150,000 per lot in 2011 to RMB 110,000 per lot in 2021, with the lowest average price reached in 2022, indicating that while the national economy was growing, people were more cautious about investing in art collections.

Further analyse the activity and growth trends within the Chinese art market. For example, Shanghai International Art Trade Week has become an important node for global art transactions, demonstrating the vitality and potential of China's art market. Meanwhile China's art auction market is also growing rapidly, with the domestic cultural relics and art auction market seeing an increase in both the number of auctions and the volume of transactions in the first half of 2023, with the turnover reaching 16.9 billion yuan, a year-on-year increase of 60.95 per cent. With the constant changes in the global economy and the continued development of China's art market, it is expected that China's position in the global art market will become more consolidated in the future, and its influence on the global art market will become more significant.

Figure 1. Art Market Transaction Scale

2) Market size of the calligraphy and painting artwork trade

The Chinese Painting and Calligraphy segment, which has continued to contract over the past few years, rebounded in terms of the number of lots sold and lot values. The Chinese Painting and Calligraphy segment saw a significant increase in the number of lots sold in the lower price ranges, which resulted in a 9% drop in its average price in 2021. Among the various categories of lots sold in 2022, Ancient Painting and Calligraphy sold for $1.997 billion, a drop of 56.51%; Modern Painting and Calligraphy sold for $4.541 billion, a down 53.79%; Contemporary painting and calligraphy turnover of 895 million yuan, down 14.11%.

Further according to the analysis of national cultural relics and works of art auction market data in 2023, the share of painting and calligraphy gradually increased, becoming the main force to stabilise the market scale. 15 auction companies launched a total of 445 auction specials in the spring and autumn seasons, with 76,345 pieces (sets) auctioned, 58,735 sold, and turnover of 18,067 million yuan, with a turnover rate of 76.93%. The turnover of Chinese painting and calligraphy was 9.711 billion yuan, accounting for 53.75% of the total, 5.24 percentage points higher than the previous year, becoming the largest segment. In addition, the turnover of letters and manuscripts increased by 112.77 per cent, making it the largest segment in terms of growth during the year, followed by the Scholar's Room Clearance, which increased by 88.02 per cent. These figures show that calligraphy and painting art transactions achieved a significant increase in 2023, not only in terms of volume, but also in terms of turnover.

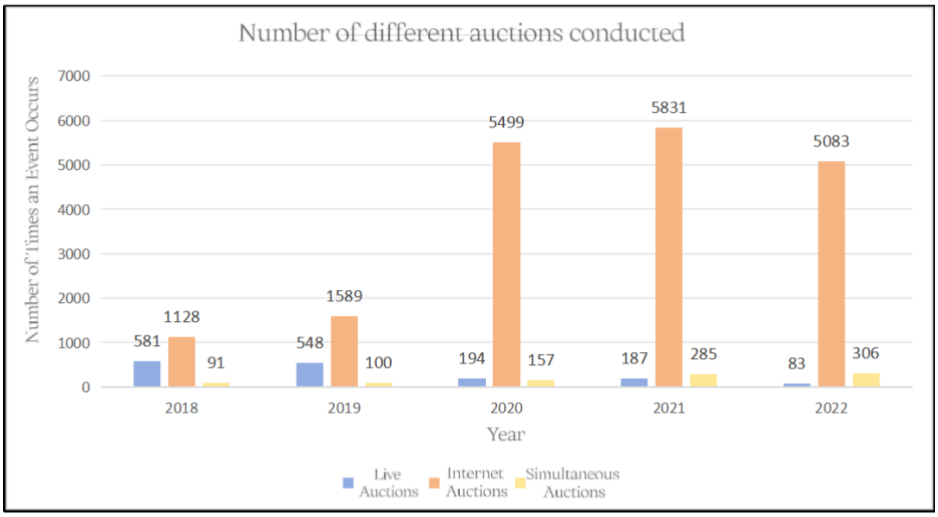


Figure 2. Scale of calligraphy and painting artwork transactions

3) Laws and trends in art transactions

In 2022, technological innovations are driving the art industry to embark on a new journey, and the art trading market is experiencing digitalisation and specialisation trends, showing a steady rise in development. Firstly, the development of China's art market tends to be steady. With the continuous development of Internet technology, art transactions rely more and more on Internet platforms and mobile devices. Digitisation not only makes art appraisal and evaluation more accurate and transparent, but also promotes the overall development of the art trading market. In addition, specialisation is also an important trend in the art trading market. Market participants are deepening their knowledge and understanding of artworks, which helps to improve the overall quality and level of the market. Secondly, art provides beauty with quality and substance. Thirdly, uncertainty in the wealth market has increased, and the importance of investment portfolios has been highlighted. Fourth, people are paying more attention to inheritance in art purchase and collection. When purchasing and managing artworks, there is a shift from satisfying immediate needs to considering long-term value. Fifth, the power of youth in the art world is gradually increasing. The power of the younger generation is steadily rising, and the influence of young artists, budding collectors and emerging art organisations is growing significantly, as they are becoming the core driving force behind the development of China's art scene. Sixthly, the financialisation of the art market is becoming increasingly apparent. The financial industry's interest in the art market has increased, and the main method of art transactions is dominated by auctions. The future development of the art market will no longer be confined to the traditional form of auctions, but will also include more financial products and services, which will bring new development opportunities for the art market.

4 Design of empirical analyses

1) Characterisation of price volatility of works of art

The rapid development of China's art market has led to a surge in demand for art. While the level of demand continues to grow, the level of supply is extremely limited, resulting in significant price movements. With the continuous development of the global economy and the continuous expansion of the art market, the impact of art price fluctuations is also increasing. In this paper, ARCH class model is chosen to econometrically analyse the price volatility of China's painting and calligraphy artwork segment, to explore whether the price volatility of artwork is clustered and asymmetric.

(1) Variable selection and data sources

Art price data is calculated based on the National Painting 400 Index 2000-2023 published by Artron Market Testing Centre (AMMA). Focusing on art auction activities in spring and autumn, Artron Art Network conducts in-depth observation and statistical analysis of its data, aiming to accurately track price fluctuations in the art market. In order to enhance the timeliness and accuracy of the reflection of price dynamics, a data conversion strategy is adopted, whereby the original semi-annual artwork price index is refined into quarterly data by means of the frequency conversion technique used in time-series analysis, thus enhancing the index's sensitivity and practicality.

The first order difference and ADF unit root test are performed on the price returns of paintings and calligraphy. The first-order difference is obtained by calculating the difference between each data point in the time series and its predecessor, aiming at eliminating non-stationarity in the data.The ADF test (Augmented Dickey-Fuller Test) is a unit-root test used to determine whether a time series has a unit root, i.e., to judge whether the time series is smooth or not.The original The assumption is that there is a unit root, i.e., the time series is non-stationary, and the calculation process is based on an autoregressive model. If the results of the ADF test reject the original hypothesis, i.e., at a given level of significance, the time series can be considered to be smooth. The results in Table 1 show that the price returns on paintings and calligraphy pass the significance test at 1%, 5% and 10% significance levels, indicating that the series of price returns on paintings and calligraphy artwork is smooth.

Table 1. ADF test of price returns of paintings and calligraphy

art index	ADF test value	1 per cent threshold	5 per cent threshold	10 per cent threshold	P-value
calligraphy	-3.743	-3.675	-2.969	-2.617	0.000

(2) Autocorrelation test, ARIMA model selection

1. ARIMA model principle.

Understanding the data series generated by the evolution of the research objective over time as a stochastic process, the aim is to approximate the characteristics and trends of this series using an appropriate mathematical model. Once the model has been accurately established, future values can be predicted based on the past trend and current state of the time series.

ARIMA (1, 1) model equation:

$$y_t = \alpha + \rho y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t \tag{1}$$

Basic procedure for ARIMA model forecasting:

A. Identify the smoothness of the series based on the scatterplot, autocorrelation function and partial autocorrelation function plots of the time series with the ADF unit root test for its variance, trend and its seasonal pattern of change.

B. Smoothing of non-smooth series.

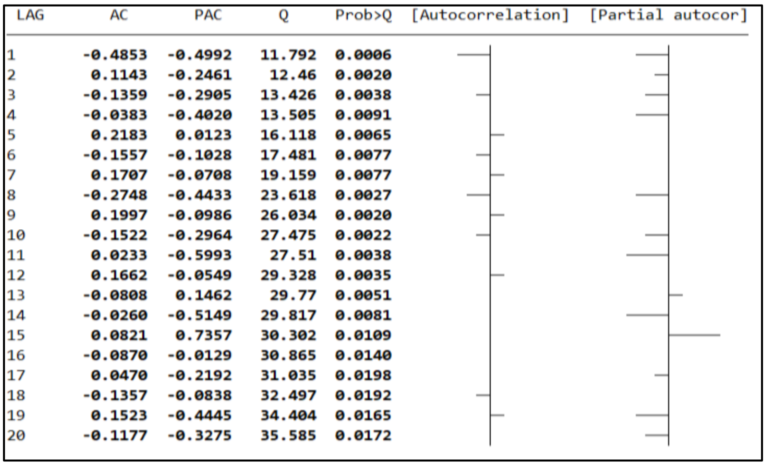
C.Build the corresponding model according to the identification rules of the time series model.

D. Perform parameter estimation and test for statistical significance.

E. Hypothesis testing is performed to diagnose whether the residual series is white noise or not, and predictive analyses are performed using the model that has passed the test.

2. Modelling analysis

From Fig. 3, it can be seen that there is autocorrelation in the series of painting and calligraphy price index, so it is necessary to establish ARIMA model in the mean equation to eliminate autocorrelation. According to the autocorrelation diagram - partial autocorrelation analysis diagram can be seen that the autocorrelation function and partial correlation function of the calligraphy and painting price index series are both manifested as trailing, the preliminary judgement can be established for the calligraphy and painting price return series ARIMA (p,q) model. By comparing the AIC values after several estimation tests, it is found that the ARIMA (1,3) model fits best for the series of painting and calligraphy price index. Further to its residuals test, it can be learnt that the residuals are not normally distributed, the p-value are greater than 0.05, accepting the original hypothesis that the series is a white noise series.



LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial autocor]
1	-0.4853	-0.4992	11.792	0.0006		
2	0.1143	-0.2461	12.46	0.0020		
3	-0.1359	-0.2905	13.426	0.0038		
4	-0.0383	-0.4020	13.505	0.0091		
5	0.2183	0.0123	16.118	0.0065		
6	-0.1557	-0.1028	17.481	0.0077		
7	0.1707	-0.0708	19.159	0.0077		
8	-0.2748	-0.4433	23.618	0.0027		
9	0.1997	-0.0986	26.034	0.0020		
10	-0.1522	-0.2964	27.475	0.0022		
11	0.0233	-0.5993	27.51	0.0038		
12	0.1662	-0.0549	29.328	0.0035		
13	-0.0808	0.1462	29.77	0.0051		
14	-0.0260	-0.5149	29.817	0.0081		
15	0.0821	0.7357	30.302	0.0109		
16	-0.0870	-0.0129	30.865	0.0140		
17	0.0470	-0.2192	31.035	0.0198		
18	-0.1357	-0.0838	32.497	0.0192		
19	0.1523	-0.4445	34.404	0.0165		
20	-0.1177	-0.3275	35.585	0.0172		

Figure 3. Autocorrelation test results of price returns of paintings and calligraphy

(3) ARCH model fitting for ARIMA process

1. Principle of ARCH model

The ARCH model is the autoregressive conditional heteroskedasticity model. It integrates all currently available information to characterise changes in variance through an autoregressive mechanism. In this modelling framework, for a given time series, the information set at each time point varies, resulting in the conditional variance at that point in time. Thus, the ARCH model effectively captures and expresses the characteristics of the dynamic evolution of conditional variance over time in a time series.

ARCH (m) model:

$$y_t = x_t \beta + \varepsilon_t \tag{2}$$

$$\sigma_t^2 = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \gamma_2 \varepsilon_{t-2}^2 + \cdots + \gamma_m \varepsilon_{t-m}^2 \tag{3}$$

where ε^2 is the residual squared and volatility and γ_i is the coefficients of the ARCH model.

GARCH (m,k) model:

$$y_t = x_t\beta + \varepsilon_t \tag{4}$$

$$\sigma_t^2 = \gamma_0 + \gamma_1\varepsilon_{t-1}^2 + \gamma_2\varepsilon_{t-2}^2 + \cdots + \gamma_m\varepsilon_{t-m}^2 + \delta_1\sigma_{t-1}^2 + \cdots + \delta_k\sigma_{t-k}^2 \tag{5}$$

where γ_i is the coefficient of the ARCH model and δ_i is the coefficient of the GARCH model.

2. Modelling analysis

In order to further accurately model the changes in volatility of the serial variables of the price index of paintings and calligraphy and to grasp the risk more precisely, after the constant regression analysis with first-order differences was implemented on the returns of the prices of paintings and calligraphy, the ARCH-LM test was further applied to examine whether there is conditional heteroskedasticity in the residuals of the regression model. The test results show that the probability value (PROB>CHI2) is 0.0265, which is less than the significance level of 0.05, so the initial hypothesis is rejected. This means that the residual series of the price returns of paintings and calligraphy do show an ARCH effect, i.e., the residual fluctuations show a non-random, aggregated character. The result of this ARCH-LM test strongly proves that the movement of the price of paintings and calligraphy tends to concentrate in a specific period of time, which manifests the phenomenon of clustering of fluctuations, further emphasising the non-uniform distribution characteristics of market fluctuations.

From the above, it can be seen that there is autocorrelation and volatility in the residual series of paintings and calligraphy, and the GARCH (p,q) model should be considered for fitting. Usually, the GARCH (1,1) model is able to describe a large amount of time series data well. Meanwhile, according to the AIC and SC criteria, suitable distributional assumptions are selected, and the estimation results are shown in Table 2.

Table 2. GARCH model estimation results

	calligraphy
C	0.007839
RESID(-1)^2	0.6819
GARCH(-1)	0.4872

Starting from the basic principle of GARCH model, we can understand that the coefficient of the ARCH term reflects the degree of influence of new information or external shocks on the current volatility, while the coefficient of the GARCH term embodies the continuity effect of the previous forecast volatility itself on the current volatility forecast. Specifically to the estimation results of the GARCH (1,1) model used, the ARCH effect coefficient of the painting and calligraphy market is 0.6819, which indicates that the newly occurring external shocks have a significant and immediate impetus to the price volatility of the paintings and calligraphy; at the same time, the GARCH effect coefficient is 0.4872, which implies that the expected value of the volatility in the previous period also strongly influences the volatility status in the current period. It is worth noting that the sum of the two coefficients exceeds 1 (i.e., 0.6819 + 0.4872 > 1), which implies that the volatility of painting and calligraphy prices is not only strongly driven by the shocks, but also that this volatility has the characteristics of self-amplification and continuous diffusion, which makes the market uncertainty increase.

In addition, both coefficients exhibit statistical significance at the 5% significance level, further confirming the decisive role of sudden external factors and the legacy of prior volatility in shaping the pattern of price volatility of paintings and calligraphy. In short, market prices of paintings and calligraphy are not only sensitive to immediate external shocks, but also profoundly influenced by the memory of historical volatility, which together exacerbate the dynamic complexity of price fluctuations.

2) Research on the prediction of the value of works of art based on the RFR algorithm

With the advent of the digital era, more and more auction houses are integrating traditional auction formats into online platforms, providing convenience and diversity of choices for global participants and bringing broader opportunities. The trend of auction format development is shown in Figure 4, which shows that the online auction format has been increasing year by year since 2019, with the number far exceeding that of live and synchronised auction formats.

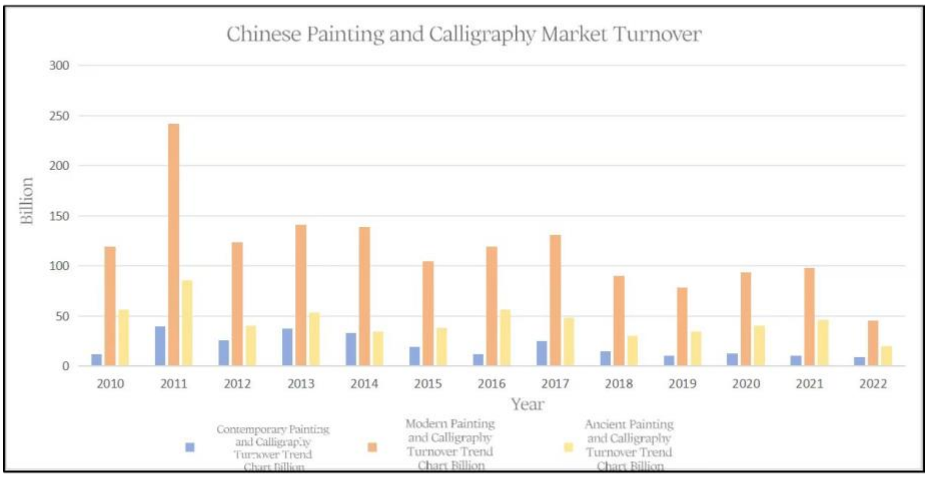


Figure 4. Comparison of the number of auctions in different formats

In this paper, we use Python to write web crawler code to crawl the data of two historical art auctions of "Chinese painting and calligraphy" in December 2022 from the authoritative Artron Art webpage in China, one is an online auction and the other is an on-site auction. The input features include the name of the painting and calligraphy artwork, classification, material, form, pattern, size, transaction price, etc., which are used as the data basis for the later modelling training. The study of the importance of features affecting the value of painting and calligraphy artworks and the prediction effect is carried out from both qualitative and quantitative perspectives.

(1) Random Forest Regression Algorithm Principles and Steps

1. Algorithmic principles

Random Forest Regression (RFR) is one of the Bagging-based algorithms in machine learning, where each decision tree is trained on different randomly selected subsets of samples, and finally the average of the outputs of all the regression trees is aggregated as the final prediction, which is insensitive to the noise in the training set, and is more conducive to obtaining a robust model.

The formula is shown below:

$$\bar{y}(x) = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{6}$$

Where, $\bar{y}(x)$ is the final prediction result of the model, x is the sample input feature, y_i is the output value of a single regression decision tree, and N is the number of regression decision trees.

In order to verify the feasibility of the model, the coefficient of determination R^2 was used to reflect the model goodness of fit. The calculated expression is:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - y_i^{'})^2}{\sum_{i=1}^{N}(y_i - \bar{y}(x))^2} \tag{7}$$

The closer R^2 is to 1, the better the fitted regression equation. The opposite indicates a poorer predictive ability of the model.

2. Algorithmic steps

A. Data Preparation: Firstly, prepare the dataset for training and testing the model. The dataset should contain features and corresponding target variables. The features are the attributes or properties used to predict the target variable, while the target variable is the value to be predicted by regression, where the training set is used to train the model and the testing set is used to evaluate the performance of the model.

B. Constructing Random Forest: use the RandomForestRegressor class in the Scikit-learn library to construct a random forest regression model. Set parameters to control the behaviour of the random forest, such as the number of decision trees, the way of feature selection, the way of decision tree growth, etc., and adjust the parameters according to the actual problems and needs.

C. Training the model: a random forest regression model is trained using a training set. The model will construct multiple decision trees based on the samples in the training set and the values of the target variables, and perform feature selection and segmentation on each tree.

D. Prediction results: the samples in the test set are predicted using a trained random forest regression model. The model will average or weight the predictions of each decision tree to get the final regression prediction.

E. Model evaluation: The performance of the model is assessed by comparison with the true target variables. Various regression performance metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination (R-squared) are used to assess the accuracy and generalisation ability of the model.

F. Model tuning: According to the results of the model evaluation, the random forest regression model is tuned to adjust the parameters of the random forest, such as increasing or decreasing the number of decision trees, adjusting the way of feature selection, and adjusting the way of decision tree growth.

G.Model application: use the trained random forest regression model to make actual predictions. New input samples are fed into the model to get the corresponding regression predictions.

(2) Data pre-processing

The raw data collected was sourced from Artron Art, and this data was subjected to a data preprocessing process before it could be used for model training. The process involves removing duplicate records, filling in or eliminating missing values, identifying and dealing with outliers, and mining the information through text analysis to extract features that are valuable to the model. This series of operations is designed to sanitise the dataset and ensure that the information fed into the model is both accurate and relevant.

In order to deeply analyse the link between the intrinsic value and price of painting and calligraphy artworks, the scope of feature selection is defined according to the feature classification framework proposed by scholar Zhang Xiaohong (2019)[18] . Based on the guidance of his research results, feature indicators that can fully reflect the unique attributes and market performance of calligraphy and painting artworks are selected to construct a data model that comprehensively reflects the characteristics of calligraphy and painting artworks, and to ensure that the model is able to learn and predict price trends based on the unique attributes of calligraphy and painting artworks. The classification characteristics of calligraphy and painting are divided into two indicators: painting and calligraphy; the material characteristics of calligraphy and painting are divided into four indicators: colour on paper, ink on paper, oil on canvas, and colour on silk; the shape characteristics of calligraphy and painting are divided into seven indicators: lens, mirror, vertical scroll, framed, fan, hand scroll, horizontal scroll, etc.; and the characteristics of calligraphy and painting motifs are divided into three indicators: landscape motifs, human figure motifs, and text motifs; based on this classification standard, the data are analysed using text analysis. Based on this classification standard, text analysis is used to extract the indexes of classification, material, form and pattern of calligraphy and painting artworks from the expert description content in the data set.

The above pre-processed qualitative and quantitative data were randomly divided into 232 sample data in a ratio of 7:3, of which 162 samples (70%) were used to train the model and the other 70 samples (30%) were used for model testing.

(3) Exploratory data analysis

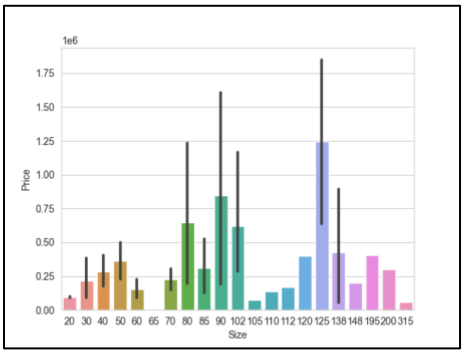


Figure 5. Statistics of size distribution characteristics

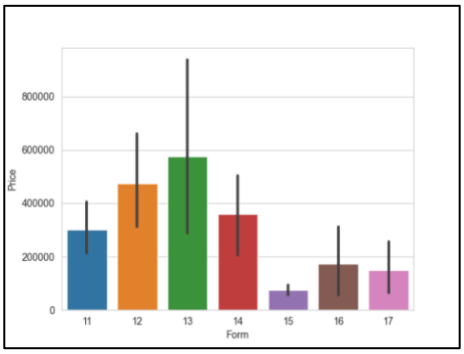


Figure 6. Statistics of form distribution characteristics

Data processing, size data to take the average of the length and width and to measure, classification, material, shape, pattern data for independent coding processing, the use of N binary trees to code N feature indicators, each feature of the indicators have its independent binary tree. Observing the size and shape distribution characteristics of the most label classification, it can be seen from the statistical chart of size distribution characteristics in Fig. 5 that the pricing of painting and calligraphy artworks with a size range of 125cm is higher, and the pricing of the increasing size is lower, so the painting and calligraphy artworks within the range of 80-125 sizes are more valuable; it can be seen from the statistical chart of the distribution characteristics of the shape distribution in Fig. 6 that the number of artworks with the shape of the vertical scroll is the highest in the price range of 400,000 yuan to 600,000 yuan, which is the highest percentage of the number of artworks within the price range of 400,000 yuan to 600,000 yuan. From Figure 6, it can be seen that the number of artworks with vertical axis shape is the highest in the high price range of RMB400,000 to RMB600,000, and the number of artworks with fan shape is in the low price range of RMB200,000, which indicates that there is a certain fluctuation in the range of shape distribution characteristics of artworks in the dataset, and that shape can be used as a significant input feature for the training of model.

(4) RFR model training

The extracted characteristics of the painting and calligraphy artworks are used as input variables for the RFR model prediction, aiming at modelling and predicting the market value of the artworks. The key to verifying the effectiveness of this model is to compare its predictions with the real transaction prices recorded by Artron. R^2 (coefficient of determination) was chosen as the evaluation criterion to quantify the accuracy of the model's predictions. Considering the large number of prediction instances, the top 40 items of the prediction results were selected for data visualisation in order to show the analysis effect visually (see Figure 7). From the visualisation results, it is observed that there is a significant positive correlation and a high degree of consistency between the predicted prices and the actual transaction prices in the RFR model output, indicating that the model predictions in most cases closely follow the actual situation. Although there are a few cases of large prediction errors, on the whole, the model demonstrates a reliable prediction ability, further confirming its effectiveness and accuracy in assessing the value of painting and calligraphy artworks.

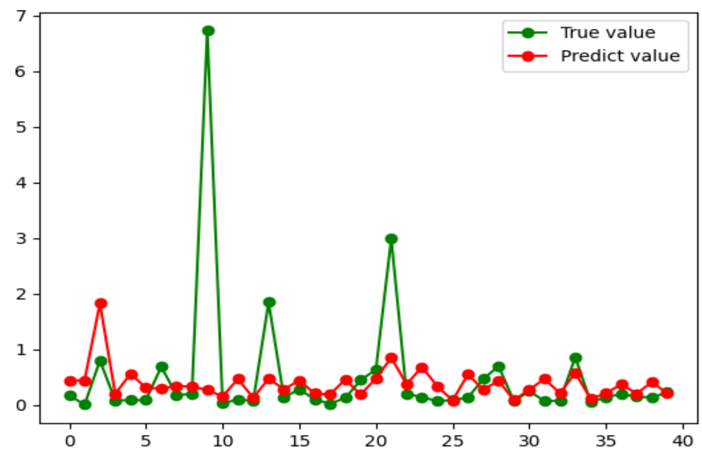


Figure 7. Comparison of actual values and results of RFR predicted values

sklearn provides very convenient metrics function to evaluate the model, Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2 Score) are used to measure the predictive ability of the regression model, which can be assessed by calculating the difference between the true value and the predicted value. Among them, R^2 measures the percentage of changes in the explanatory variables of the model, and the closer the value is to 1 indicates that the model fits better.The results of the RFR model assessment are shown in Table 3, and it is sufficient to focus only on the R^2 Score among these four values, and the score of the RFR model is 0.76, which is close to 1, indicating that the RFR model is suitable for the research on the value of painting and calligraphy artworks.

Table 3. Assessment of model effects

PREDICTIVE MODELLING	MAE VALUE	MSE VALUE	RMSE VALUE	R^2 SCORE
RFR MODEL	16726.1891	4536308.3645	2122.8067	0.7562

(5) Analysis of the importance of influencing factors

In the prediction model of painting and calligraphy artwork prices, the degree of contribution of each parameter to the value of the artwork is analysed in depth by calculating the average influence of each feature in all decision trees in the random forest. This approach effectively highlights the relative importance of different attributes in the prediction process. After implementing this analysis with the help of Python programming, a visual presentation of the importance of features was obtained. According to Figure 8, the size of the painting and calligraphy work is identified as the primary factor affecting its market value, highlighting the key role that size plays in the pricing of artworks, with an importance value of 0.3418, followed by the material and form of the painting and calligraphy with a relatively high and similar level of importance of 0.2844 and 0.2267, respectively, compared to the low importance of the classification of the painting and calligraphy, which accounts for only 0.0137.

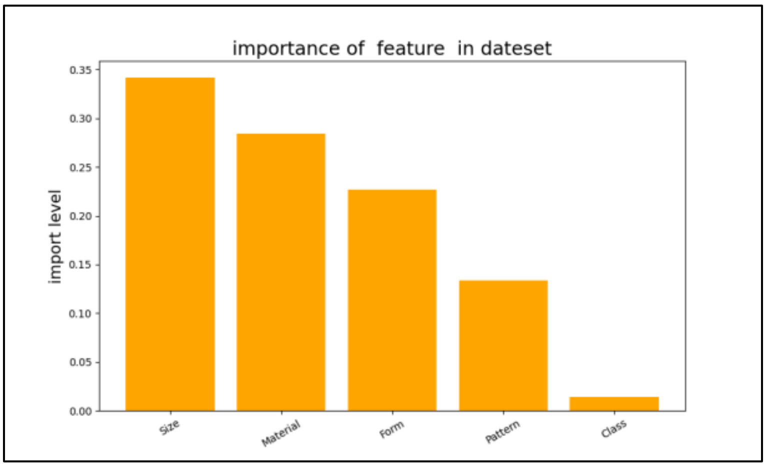


Figure 8. Histogram of the importance of the characteristics of paintings and calligraphy

5 Conclusions and policy recommendations

1) Conclusion

By integrating machine learning algorithms and time series analysis techniques, we have pioneered the exploration of dynamic prediction of Chinese painting and calligraphy artwork prices and market value trend analysis. Extensively collecting and collating the transaction data of Chinese painting and calligraphy artworks in recent years through crawler technology and literature analysis, a multi-dimensional analysis framework is constructed with the aim of revealing the complex driving factors behind artwork values and accurately predicting future price trends. Key variables affecting artwork prices are first identified, including but not limited to the style type of the work, market circulation and macroeconomic environment. Advanced feature engineering techniques are used to improve data quality. Secondly, time series models ARIMA and GARCH models are used to effectively capture the time-dependent and cyclical fluctuations of artwork prices and improve the accuracy of forecasting. Once again, integrated learning methods are introduced to cross-validate and comprehensively assess the prices of artworks with the help of Random Forest Regression (RFR) algorithm. The stable performance of the model predictions under different market conditions is ensured. The results show that the hybrid prediction framework combining time series and machine learning models significantly improves the accuracy of price prediction, especially in identifying market turning points and abnormal fluctuations, which opens up a new way of thinking for quantitative analysis of the art market.

2) Policy recommendations

(1) Enhancing standardised data sharing

An art information database is a database dedicated to the collection, preservation and management of information related to works of art. Nowadays, art information databases mainly have the problems of limited data volume, insufficient data standardisation and insufficient openness, which will have an impact on the comparability and usability of the data. Therefore, the government and industry associations need to promote the establishment of a unified database of art information and improve the art value evaluation system under the premise of fully respecting China's national conditions. By standardising data formats and promoting the open sharing of data, researchers and market participants can advance their studies on the assessment of the value and price of works of art.

(2) Development of a scientific pricing system for works of art

The art pricing system is a guiding framework for the art market, which helps buyers and sellers in the art market to determine the actual value of artworks. Based on time series modelling and machine learning, the system should combine multi-dimensional factors such as artistic value, historical significance and market demand to provide a reference for art pricing, reduce market bubbles and protect investors from irrational fluctuations. Finally, a sound accountability mechanism for art value assessment and pricing should be put in place to increase the penalties for failures in the assessment process and ensure the effectiveness of the art pricing system.

(3) Improve the art market trading system

The art trading system refers to a set of legal, cultural and economic rules governing the buying and selling of artworks. The lack of information disclosure in the art market has led to information asymmetry, making it difficult to judge whether transaction prices are fair, and the existence of problems such as false advertising and forgery. At the same time, the lack of an effective regulatory mechanism in the art market has led to the existence of some illegal behaviours in the market, and the profit-seeking mentality has led to the overpricing of some works of art, which in the end may jeopardize the entire art market. Therefore, the art market trading system needs to be further improved. Strengthen the connection between artists and collectors and the art market, promote the innovation and demand expansion of artworks, and provide a solid foundation of elemental resources for the art market.

(4) Enhance investor education and risk alerts

Art investment has a high professional threshold and uncertainty. Regulators and financial institutions have strengthened art knowledge and risk education for investors, especially regarding the volatility of art prices and the identification of authenticity, so as to help investors make more rational decisions.

(5) Formulation of laws and regulations adapted to new technologies

The government should encourage and support the application of technological innovation in the art market, using artificial intelligence and big data analysis to enhance the efficiency of art appraisal, or broadening the channels for art display and trading through virtual reality technology. With regard to the emerging artwork price prediction technology based on time series models and machine learning, relevant laws and regulations should be revised and improved in a timely manner to clarify the authority to use data, protect intellectual property rights, and define the boundaries of responsibility, so as to ensure that the application of the technology does not infringe on personal privacy, while protecting the value of the original artworks from being infringed upon.

References

[1] Coslor E. (2016). Transparency in an opaque market: Evaluative frictions between "thick" valuation and "thin" price data in the art market. *Accounting, Organizations and Society*, 50, 13-26. https://doi.org/10.1016/j.aos.2016.03.001

[2] Turan F. K., & Tosun Z. (2023). Sustainable development of art industry and a statistical analysis of the factors that influence the gallery prices of contemporary artworks. *Sustainable Development*, 31(3), 1790-1804. https://doi.org/10.1002/sd.2484

[3] Horvitz J. E. (2009). An overview of the art market. *Collectible Investments for the High Net Worth Investor*, 85-117. https://doi.org/10.1016/B978-0-12-374522-4.00005-6

[4] Huang F., Xiong H., & Chen S., et al. (2023). Slope stability prediction based on a long short-term memory neural network: Comparisons with convolutional neural networks, support vector machines and random forest models. *International Journal of Coal Science & Technology*, 10(1), 18. https://doi.org/10.1007/s40789-023-00579-4

[5] Abiodun O. I., Jantan A., & Omolara A. E., et al. (2019). Comprehensive review of artificial neural network applications to pattern recognition. *IEEE Access*, 7, 158820-158846. https://doi.org/10.1109/ACCESS.2019.2945545

[6] Mandel B. R. (2009). Art as an investment and conspicuous consumption good. *American Economic Review*, 99(4), 1653-1663. https://doi.org/10.1257/aer.99.4.1653

[7] Sidorova E. (2019). The cyber turn of the contemporary art market. *Arts*, 8(3), 84. https://doi.org/10.3390/arts8030084

[8] Durlauf S., & Blume L. E. (2016). The new Palgrave dictionary of economics. *Springer*. https://doi.org/10.5860/choice.46-0666

[9] Kakhki M. D., Farsi M. M., & Fakari B., et al. (2019). Volatility Transmission of Barley World Price to the Domestic Market of Iran and the Role of Iran Mercantile Exchange: An Application of BEKK Model. *New Medit*, 18(3), NA-NA. https://doi.org/10.30682/nm1903h

[10] Bocart F. Y. R. P., & Hafner C. M. (2015). Volatility of price indices for heterogeneous goods with applications to the fine art market. *Journal of Applied Econometrics*, 30(2), 291-312. https://doi.org/10.1002/jae.2355

[11] 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

[12] Lee B., Fraser I., & Fillis I. (2022). To sell or not to sell? Pricing strategies of newly-graduated artists. *Journal of Business Research*, 145, 595-604. https://doi.org/10.1016/j.jbusres.2022.03.035

[13] Wijnberg, N. M., & Gemser, G. (2000). Adding value to innovation: Impressionism and the transformation of the selection system in visual arts. *Organization Science*, 11(3), 323-329. https://doi.org/10.1287/orsc.11.3.323.12499

[14] Botha F., Snowball J., & Scott B. (2016). Art investment in South Africa: Portfolio diversification and art market efficiency. *South African Journal of Economic and Management Sciences*, 19(3), 358-368. https://doi.org/10.17159/2222-3436/2016/v19n3a4

[15] Galbraith J. W., & Hodgson D. J. (2018). Econometric fine art valuation by combining hedonic and repeat-sales information. *Econometrics*, 6(3), 32. https://doi.org/10.3390/econometrics6030032

[16] Zhao J., Wang Y., & Zheng H., et al. (2021). An Evaluation/Price Prediction Model for Art Investment. In *Proceedings of the 2021 5th International Conference on E-Business and Internet*, 68-75. https://doi.org/10.1145/3497701.3497713

[17] Indrawan J. O., Filia B. J., & Parmonangan I. H. (2023). Multimodal Approach for Painting Price Prediction. In *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, 1-5. https://doi.org/10.1109/ICORIS60118.2023.10352208

[18] Zehtab-Salmasi A., Feizi-Derakhshi A. R., & Nikzad-Khasmakhi N., et al. (2023). Multimodal price prediction. *Annals of Data Science*, 10(3), 619-635. https://doi.org/10.1007/s40745-021-00326-z

[19] Guo J., & Tanaka T. (2020). Dynamic transmissions and volatility spillovers between global price and US producer price in agricultural markets. *Journal of Risk and Financial Management*, 13(4), 83. https://doi.org/10.3390/jrfm13040083

[20] Jang D., & Park M. (2019). Price determinant factors of artworks and prediction model based on machine learning. *Journal of Korean Society for Quality Management*, 47(4), 687-700. https://doi.org/10.7469/JKSQM.2019.47.4.687

[21] Liu C. (2022). Prediction and analysis of artwork price based on deep neural network. *Scientific Programming*, 2022(1), 7133910. https://doi.org/10.1155/2022/7133910

[22] Wang D. (2022). Research on the art value and application of art creation based on the emotion analysis of art. *Wireless Communications and Mobile Computing*, 2022(1), 2435361. https://doi.org/10.1155/2022/2435361

[23] Spee B. T. M., Mikuni J., & Leder H., et al. (2023). Machine learning revealed symbolism, emotionality, and imaginativeness as primary predictors of creativity evaluations of western art paintings. *Scientific Reports*, 13(1), 12966. https://doi.org/10.1038/s41598-023-39865-1

About the Author

Ai Ziyuan , male, Han nationality,2001/5, Research interests: news communication, calligraphy and painting art prices and communication integration, Beijing Normal University, Hong Kong Baptist University, United Nations International School of Journalism and Communication Studies in graduate students, postal code 519000 E-mail: 568147275@qq.com.