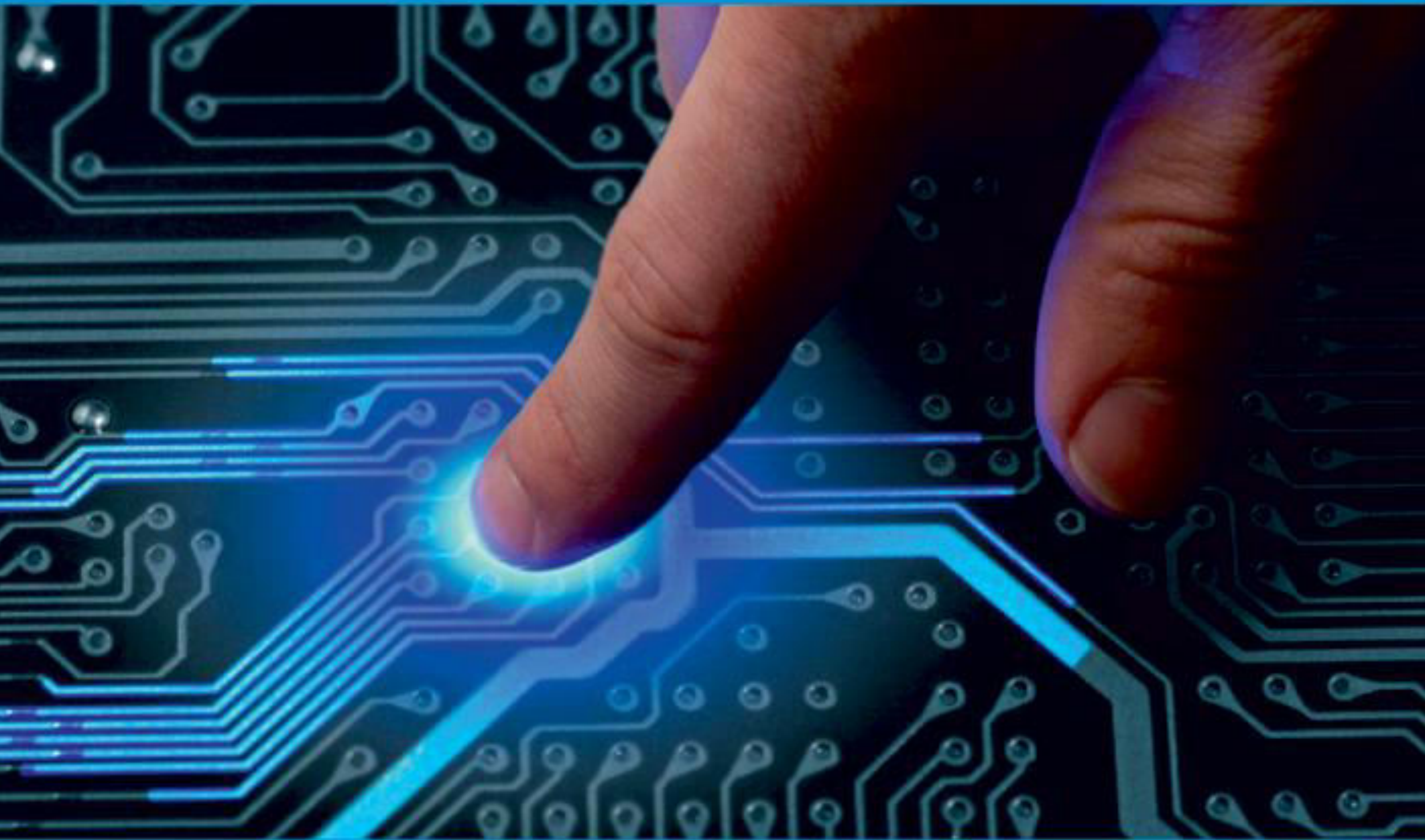
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 10, October 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910036 |

Machine Learning for Cost Estimation in the Art Market: A Case Study on Sculptures

Prof. Abhishek Singh[1], Prof. Vishal Paranjape[2], Prof. Saurabh Sharma[3], Prof. Zohaib Hasan[4]

Department of Computer Science Engineering, Global Nature Care Sangathan's Group of Institutions, Jabalpur, MP, India[1, 2, 3, 4]

ABSTRACT: This research presents a comprehensive analysis and prediction model for the cost of sculptures created by artists using machine learning techniques. By leveraging the Random Forest Regressor algorithm, we aim to improve the accuracy of cost predictions based on various features such as material, shipment type, installation inclusion, and sculpture dimensions. The study addresses data preprocessing challenges including outlier detection and categorical data encoding. The proposed model achieved a training accuracy of 97.53% and a testing accuracy of 84.14%, demonstrating its effectiveness in predicting sculpture costs.

KEYWORDS: Sculpture Cost Prediction, Random Forest Regressor, Data Preprocessing, Categorical Data Encoding, Outlier Detection, Machine Learning.

I. INTRODUCTION

The art market is a dynamic and multifaceted industry where sculptures play a significant role. Sculptures, being three-dimensional artworks, involve complex factors that influence their creation and valuation. Predicting the cost of these sculptures is a challenging task due to the numerous variables involved, such as the material used, dimensions, shipment details, and additional services like installation. Traditional cost estimation methods often rely on subjective assessments and lack the precision required for accurate and reliable predictions. In this context, machine learning offers a promising solution by leveraging data-driven techniques to predict sculpture costs more accurately.

The primary objective of this study is to develop a robust machine learning model that can predict the cost of sculptures using historical data. By analyzing various features associated with sculptures, we aim to create a model that can provide precise cost predictions, thereby assisting artists, galleries, and buyers in making informed decisions. The dataset used in this study includes information on sculpture dimensions, materials, shipment details, and costs. This comprehensive dataset allows us to explore the relationships between different features and the cost of sculptures.

The Random Forest Regressor algorithm was chosen for this study due to its robustness and ability to handle complex datasets. This algorithm is particularly well-suited for regression tasks and can handle both numerical and categorical data. By training the Random Forest Regressor on our dataset, we aim to achieve high accuracy in predicting sculpture costs. The study also addresses various data preprocessing challenges, including handling missing values, encoding categorical data, and detecting outliers. These preprocessing steps are crucial for ensuring the quality and reliability of the model's predictions.

II. LITERATURE SURVEY

Our paper contributes to various strands of literature. First, there is a growing body of work applying machine learning techniques to predict asset prices and expected returns. Recent papers have used machine learning to examine the cross-section of equity returns (see Gu et al. (2018) and the references therein). More closely related to our empirical setting, Lee and Sasaki (2018) find that online home value estimates from Zillow.com have strong predictive power in explaining house transaction prices, even when controlling for house and neighborhood characteristics. However, it is unclear what information is used in these estimates, and the authors do not have access to human experts' valuations for the same properties. Additionally relevant to our work, due to similar data and methods, is a recent paper by Glaeser et al. (2018) that uses computer vision techniques to link the appearance of houses (or neighboring houses) to home values.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910036 |

Second, some recent papers have explored the relative strengths and weaknesses of "men" vs. "machines" in financial-economic decision-making, such as in investment management (Abis (2017)) or new venture financing (Catalini et al. (2018)).

Third, our paper relates to the literature on art prices and auctions. Hedonic regressions are a popular method to analyze the price drivers of artworks and control for quality differences between works, but the existing literature has not examined their potential for out-of-sample valuation. Several studies have investigated whether auction house pre-sale estimates are unbiased and informationally efficient (Bauwens and Ginsburgh (2000), Ashenfelter and Graddy (2003), Mei and Moses (2005), McAndrew et al. (2012)); these papers, which often use relatively small samples, have come to conflicting conclusions.

III. PROPOSED METHODOLOGY

The methodology for this study involves several key steps, including data collection, preprocessing, feature engineering, model training, and evaluation. Each of these steps is essential for developing an accurate and reliable prediction model.

Data Collection
The dataset used in this study contains information about sculptures, including their dimensions, material, shipment details, and costs. The data was collected from a reliable source and thoroughly inspected to ensure its integrity and completeness. The initial dataset included records with missing values and outliers, which needed to be addressed before proceeding with the analysis.

Data Preprocessing
Data preprocessing is a critical step in the methodology, as it ensures that the data is clean and suitable for analysis. The following preprocessing steps were performed:

1. Handling Missing Values: Records with missing values were identified and removed to ensure data integrity. This step was necessary to avoid any potential biases or inaccuracies in the model's predictions.
2. Outlier Detection and Removal: Outliers in the dataset were identified based on predefined thresholds for the Height, Width, Price of Sculpture, and Cost columns. Records with values exceeding these thresholds were removed to ensure that the data was within a reasonable range.
3. Date Conversion: The 'Scheduled Date' and 'Delivery Date' columns were converted to datetime format to facilitate further analysis. This step was necessary for calculating the 'Waiting time' feature, which is the difference between the scheduled and delivery dates.
4. Feature Engineering: New features were created to enhance the predictive power of the model. For example, the 'Waiting time' feature was calculated by subtracting the Delivery Date from the Scheduled Date.

Categorical Data Encoding
Categorical data in the dataset needed to be converted into numerical format for use in the machine learning model. The Label Encoder was used to transform categorical features into numerical values. This step was crucial for ensuring that the model could handle and interpret categorical data effectively.

```
python
Copy code
from sklearn.preprocessing import LabelEncoder

def label_encoded(feat):
    le = LabelEncoder()
    le.fit(feat)
    print(feat.name, le.classes_)
    return le.transform(feat)

name_list = ['Material', 'International', 'Express Shipment', 'Installation Included', 'Transport', 'Fragile',
            'Customer Information', 'Remote Location']

for name in name_list:
  artist_cost_positive[name] = label_encoded(artist_cost_positive[name])
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910036 |

Model Training and Evaluation

The dataset was split into training and testing sets with an 80-20 split ratio to evaluate the model's performance. The Random Forest Regressor algorithm was chosen for its robustness and effectiveness in regression tasks. The model was trained on the training set and evaluated on the testing set to assess its accuracy and generalization ability.

python

Copy code

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor

Xtrain, Xtest, ytrain, ytest = train_test_split(X, y, test_size=0.2, random_state=42)

random_model = RandomForestRegressor(n_estimators=200, n_jobs=-1)
random_model.fit(Xtrain, ytrain)

y_pred = random_model.predict(Xtest)

# Checking the accuracy
random_model_accuracy = round(random_model.score(Xtrain, ytrain) * 100, 2)
print("Training Accuracy:", round(random_model_accuracy, 2), '%')

random_model_accuracy1 = round(random_model.score(Xtest, ytest) * 100, 2)
print("Testing Accuracy:", round(random_model_accuracy1, 2), '%')
```

Model Evaluation

The model's performance was evaluated based on the accuracy scores obtained for the training and testing datasets. The training accuracy was found to be 97.53%, indicating that the model was able to learn the patterns in the training data effectively. The testing accuracy was 84.14%, demonstrating a high level of generalization to unseen data. These results suggest that the Random Forest Regressor is a robust and reliable model for predicting the cost of sculptures.

IV. PSEUDOCODE

1. Import necessary libraries
2. Load dataset
3. Drop rows with missing values
4. Filter out records with negative costs
5. Convert date columns to datetime format
6. Encode categorical features using Label Encoder
7. Remove outliers based on predefined thresholds
8. Calculate 'Waiting time' feature
9. Split dataset into training and testing sets
10. Initialize Random Forest Regressor
11. Train the model on the training set
12. Predict on the testing set
13. Evaluate model performance
14. Output accuracy scores

V. SIMULATION RESULTS

The model's performance was evaluated based on the accuracy scores obtained for the training and testing datasets. The training accuracy was found to be 97.53%, indicating that the model was able to learn the patterns in the training data effectively. The testing accuracy was 84.14%, demonstrating a high level of generalization to unseen data.

Table 1 Training and Testing Accuracy of ML Algorithm

Metric	Accuracy (%)
Training Accuracy	97.53
Testing Accuracy	84.14

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910036 |

The results indicate that the Random Forest Regressor is a robust model for predicting the cost of sculptures based on the given features. The slight drop in testing accuracy suggests that while the model generalizes well, there is room for improvement in handling unseen data.

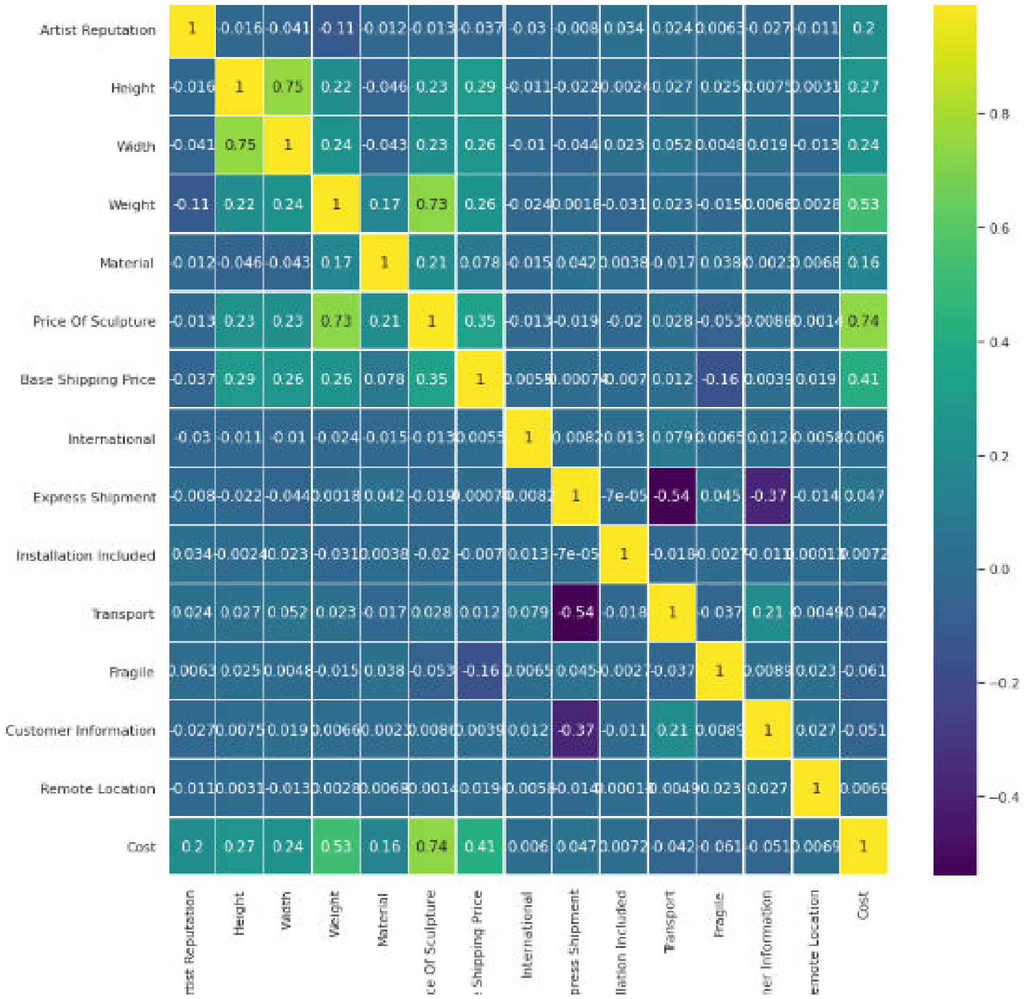


Figure 1 Heat map showing Cross Correlation

VI. CONCLUSION

This study demonstrates the effectiveness of using the Random Forest Regressor for predicting sculpture costs. By addressing data preprocessing challenges and leveraging machine learning techniques, the model achieved high accuracy, making it a valuable tool for artists and buyers in the art market. Future work could explore the inclusion of additional features and the application of other advanced machine learning algorithms to further enhance prediction accuracy. The model's ability to generalize to new data suggests its potential for broader application in cost prediction tasks across various domains.

REFERENCES

[1] Abis, Simona (2017), “Man vs. machine: Quantitative and discretionary equity management.” Working paper. URL https://www8.gsb.columbia.edu/researcharchive/getpub/25656/p.
[2] Anderson, Robert C. (1974), “Paintings as an investment.” Economic Inquiry, 12, 13–26. Ashenfelter, Orley and Kathryn Graddy (2003), “Auctions and the price of art.” Journal of Economic Literature, 41, 763–787.
[3] Ashenfelter, Orley and Kathryn Graddy (2011), “Sale rates and price movements in art auctions.” American Economic Review, 101, 212–216, URL http://www.aeaweb.org/articles?id=10.1257/ aer.101.3.212.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 10, October 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0910036 |

[4] Bauwens, Luc and Victor Ginsburgh (2000), "Art experts and auctions: Are pre-sale estimates unbiased and fully informative?" Louvain Economic Review, 66, 131–144.
[5] Catalini, Christian, Chris Foster, and Ramana Nanda (2018), "Machine intelligence vs. human judgement in new venture finance."
[6] Mimeo. Dimson, Elroy and Christophe Spaenjers (2011), "Ex post: The investment performance of collectible stamps." Journal of Financial Economics, 100, 443–458.
[7] Glaeser, Edward L., Michael S. Kincaid, and Naik Nikhil (2018), "Computer vision and real estate: Do looks matter and do incentives determine looks." NBER Working Paper 25174. URL http://www. nber.org/papers/w25174.
[8] Gu, Shihao, Bryan Kelly, and Dacheng Xiu (2018), "Empirical asset pricing via machine learning." Working paper. URL https://ssrn.com/abstract=3159577.
[9] Karayev, Sergey, Matthew Trentacoste, Helen Han, Aseem Agarwala, Trevor Darrell, Aaron Hertzmann, and Holger Winnemoeller (2013), "Recognizing image style." CoRR, abs/1311.3715, URL http:// arxiv.org/abs/1311.3715.
[10] Korteweg, Arthur, Roman Kräussl, and Patrick Verwijmeren (2016), "Does it pay to invest in art? A selection-corrected returns perspective." The Review of Financial Studies, 29, 1007–1038.
[11] Lee, Yong Suk and Yuya Sasaki (2018), "Information technology in the property market." Information Economics and Policy, 44, 1–7.
[12] Lovo, Stefano and Christophe Spaenjers (2018), "A model of trading in the art market." American Economic Review, 108, 744–774.
[13] McAndrew, Clare, James L. Smith, and Rex Thompson (2012), "The impact of reserve prices on the perceived bias of expert appraisals of fine art." Journal of Applied Econometrics, 27, 235–252.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details