*2023 International Conference for Advancement in Technology (ICONAT)*
*Goa, India. Jan 24-26, 2023*

# An Efficient Machine Learning Approach: Analysis of Supervised Machine Learning Methods to Forecast the Diamond Price

Md Shaik Amzad Basha
*GITAM School of Business.*
*Gandhi Institute of Technology and Management (Deemed to be University)*
*Bengaluru, India*
amjuamjad66@gmail.com

Peerzadah Mohammad Oveis
*GITAM School of Business.*
*Gandhi Institute of Technology and Management (Deemed to be University)*
Bengaluru, India
peerzadahmohammad@gmail.com

C Prabavathi
*Department of Professional Studies*
*Christ (Deemed to be University)*
Bengaluru, India
prabavathidps@gmail.com

Macherla Bhagya Lakshmi
*Department of Professional Studies*
*Christ (Deemed to be University)*
Bengaluru, India
bhagyamacherladps@gmail.com

M Martha Sucharitha
*Department of Professional Studies*
*Christ (Deemed to be University)*
Bengaluru, India
marthasucharitha@gmail.com

2023 International Conference for Advancement in Technology (ICONAT) | 978-1-6654-7517-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICONAT57137.2023.10080618

***Abstract*— Diamond, a found natural process compound of carbon, is one of the hardest and most immensely expensive material known to men, especially more to women. Investments in expensive gems like diamonds are in significant demand. The rate of a diamond, nevertheless, is not as easily calculated as the value of either gold or platinum since so many factors must be taken into account. Because there is such a broad range of diamond dimensions and qualities; as a result, being able to make reliable price predictions is crucial for the diamond industry. Although, making accurate predictions is challenging. In this study, we implemented multiple machine learning techniques employed to the challenge of diamond price forecasting's such as Linear Regression, Random Forest, Decision Tree Random Forest, Cat-Boost Regressor and XGB Regressor. This article's goal is to develop an accurate model for estimating diamond prices based on its characteristics such as weighting factor, cut grade, and dimensions. We compared the sum of estimated values and test values of predicted values with overestimated, underestimated and exact estimations. We applied cross-validation to calculate how much the model deviates from the actual when faced with a difference between the training set and the test set. We predicted values side by side. We performed a comparative analysis of supervised machine learning models with other models to evaluate the model accuracy and performance metrics. The Study's experimental findings show that out of all the supervised machine learning models, Random Forest performs well with R2score and Low RMSE and MAE values and CV Score.**

***Keywords*— Estimated values, Machine learning, Diamonds, Regression Models and Cross validation***

## I. INTRODUCTION

In terms of market value, diamonds are by far the most sought-after gem kind. Diamonds demand a price tag that is tens of times higher than that of any other gemstones. Diamond's optical characteristics, or how it interacts with light, contribute to its widespread acclaim. Diamonds are popular for a variety of reasons outside their obvious toughness and longevity, including the fashion industry, cultural significance, and producers' persistent advertising. When it comes to jewelry, diamonds are at the very top. the world's most expensive items, as it possesses the astounding capability of spreading light everywhere. The worth of a diamond is determined by various factors, including its shape, cut, inclusions (impurities), and weight in carats. Diamonds have various industrial applications due to their superior slicing, buffing, and piercing capabilities. The tremendous worth of gemstones has ensured their continued worldwide trade for millennia. Faceted diamonds are evaluated based on their color, cut, clarity, and carat weight to assess their overall quality. In the 1950s, the Gemological Institute of America created a standardized technique for evaluating diamonds quality known as the "4Cs of Diamond Quality." [1]

As a rule, the weight of materials such as platinum and gold are used to determine their worth, but the pricing of diamond depends on a number of other aspects as well. The carat, the cut, and a host of others are all relevant. Since diamonds are so expensive, even a little shift in these variables would have a major impact on the diamond's final price.

Diamonds, like any other commodity, go through a production process that adds value at each stage until they reach the retail shops. The cost of a polished diamond is established after the raw diamonds has been mined and cut. When a diamond is destined for use as an ornament, it goes through a number of procedures to enhance its appearance and change it into a stunning adornment human practice of adorning oneself. The initial cost of a diamond depends on its rarity, size, and the time and effort required to polish, cut, and mine them. The right price depends on several factors, not just one from the gems. The four Cs refer to the color, clarity, carat, cut, and presentation of the gemstone. dimensions like depth, width, table. In particular, color, cut, clarity, and carat

are four of the most important characteristics. The two most important aspects of a product are its weight and its color. The four characteristics of a diamond that most affect its price are synonymous with the "4Cs of Diamond Quality."[2].

There were several experts we investigated who developed and tested a wide range of machine learning models that learn the optimal solution to a given problem build a model to forecast prices. pricing is a continual, changeable goal variable. we determined that supervised regression algorithms should be employed. Section I introduced, the study is divided into three sections: sec. II reviews the literature of related works, while Sec. III provides Research Methodology. The proposed mechanism for determining diamond prices is outlined in sec III machine learning relies on the use of supervised regression methods, and this article discusses both types of methods as well as the dataset used to train them. The findings and evaluation of regression algorithms based on measured technical specifications are discussed in Sec IV. In Sec V, we concluded the research that has to be done to compare the models for accuracy and prediction of diamond price.

## II. Literature review

G. Sharma et. al [3] In their study they explained the comparative analysis of machine learning techniques that used to predict the diamond price and they evaluated best accuracy model based on the findings of their research study, Experimentation and analysis lead us to the conclusion that diamond pricing evaluated using supervised learning approaches including Ada-Boost, Gradient-Boosting, linear regression, decision tree, Elastic-Net, ridge, lasso regression, and the random forest approach. Their findings that accuracy of the Random Forest Regression Algorithm is 97%. Its strong capability to calculate continuously numerical values allows it to provide such a high level of precision. A. C. Pandey et. al [4] findings of their research they have done the comparative analysis between various ensemble models and regressors. State-of-the-art methods like gradient boosting, ada boost, and bagging have been contrasted to the suggested model's performances. All of the approaches' conclusions have been shown as regression model feature choices. The performance of current ensemble models improved by adding the feature selection and preprocessing stages, as shown by a comparison of the results achieved using ensemble regressors and feature choices with regression models. that collection of data used for instruction. Chu S [5] article serves as an example of how to construct a linear regression model. It is demonstrated that by employing a exponential regression model, one sidestep the appearance of a negative intercept, which would be counter-intuitive. Since these regression methods are naturally linear, typical linear regression techniques used to estimate them once the data has been transformed appropriately. In this Chu s [6] article, they explained about the comparative analysis of regression models are linear in nature. The most effective and precise model is found in this work by S. A. Fitriani et al [7]. The k-Nearest Neighbor (kNN) and Least Absolute Shrinkage and Selection Operator (LASSO) models were created to predict diamond prices (LASSO). In order to achieve maximum precision, we carefully choose features by weighing the k value from k-NN against the alpha value from LASSO. Through a comparison of RMSE and R2, they found that the k-NN method outperforms the LASSO approach. k-NN produces the lowest RMSE value (926.07) and the greatest R2 value (0.9066), or 90.66 percent. A variety of machine learning methods, including Liner regression, random forest, and Ensemble, were compared in their W. Alsuraihi [8] work to aid in the prediction of diamond price. The use of polynomial regression, forest regression, gradient descent, and a system of neurons. Following exhaustive model training, accuracy testing, and data analysis, random forest regression was shown to be the most effective model. In spite of the noise, they found that random forest produced the best outcome. This is why we suggest that you create an ensemble model or use a random forest. By merging many models, ensemble learning can increase machine learning outcomes and address issues with (bagging), bias (boosting).

Evaluation of the three linear algorithms Using a combination of regression, neural network, and M5P, they investigated estimating the value of diamonds price Pena et. al [9]. Within the scope of this study, they placed a procedure for reducing the number of dimensions and the issue of similarity between elements of the dataset. In a number of cases, this in general, the accuracy drops when there are associated characteristics in the dataset. the model itself A good illustration of this seen in the collection of data, where the Width are extremely important characteristics. connected to one another As a result, the model's accuracy can be improved by excluding linked characteristics. Using the diamonds dataset, the M5P model's accuracy is with a score of 98.7 percent on the training data, we were able to for dimensionality reduction, the M5P method is quite precise. reached a new high of 99.03%. Based on the research presented in their study, conclusion that the M5P algorithm is the most appropriate. Using a custom-built algorithm, we are able to anticipate future dataset by means of reducing its dimensions.

By analyzing the results of many regression methods [10]. using the dataset, we were able to determine which ones would be most useful for predicting diamond prices in the future. We analyze cross validation score as a result, we have examined various techniques to identify the most appropriate algorithm for creating and forecasting the value of diamonds.

## III. Proposed Methodology

The whole execution of diamond price forecasting is broken down into several sub-steps, such as preprocessing the dataset, normalization of the data, removing duplicate values, missing numbers, label encoding as categorical and numerical and train the supervised machine learning approaches to determine the optimal methods for the price forecasting based on a number of metrics, including R2 score, root mean square error (RMSE), and mean absolute error (MAE). The technique we provided and the procedures we used to put the models into action are depicted in Fig. 1. As shown in the workflow sequence of Figure 1, the data is normalized and cleaned before it is utilized for data to train. The dataset is divided into two parts: the train data, and uutilized to create the models, and the test data, which is utilized to assess the quality of the algorithms and calibrate the important performance characteristics. After extensive training and testing, the

parameters for all of the algorithms' process variables are determined. The regression models are then used to predict diamond prices using these statistics.

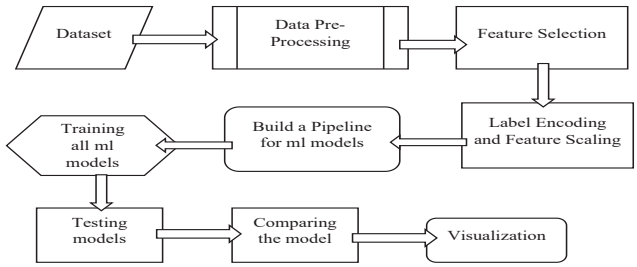


Fig. 1. Workflow of Regression Models

## A. Dataset

Kaggle is a platform for sharing and discussing machine learning and data sciences techniques. For the purpose of training ML models, it provides easy access to hundreds of datasets. In addition to sharing their models, users could develop them and tests with other tools. Specifically, we have trained our supervised machine learning models using Diamond file collection [11].

## B. Data Description

There are a total of 53,940 distinct entries in the diamond collection dataset. Attributes of the given dataset depicted in Table 1. The best-guess price of diamonds predicted with the use of the data's qualities. The 4Cs of the diamond refer to these four characteristics and are commonly used terminology among diamond and jewelry experts.

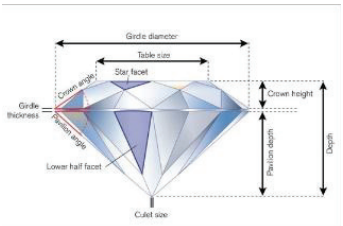


Fig. 2. Diamond Dimensions

TABLE I. DATA SET FEATURES

| ***Features*** | ***Range*** |
|---|---|
| Cut | (Fair, Good, Very Good, Premium, Ideal) |
| Color | J (worst) - D (best) |
| Clarity | (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best)) |
| Carat | 0.2 - 5.01 ct |
| Depth | (0 - 31.8) mm |
| Table | (43 - 95) mm |
| Price | ($326--$18,823) |
| X(Length) | (0 - 10.74) mm |
| Y(Width) | (0 - 58.9) mm |
| Z (Depth) | (0 - 31.8) mm |

From Table 1 shows from All the diamonds in the data collection the vary in carat weight from 0.21 to 5.01 kg. There are five possible levels of excellence for the cut: excellent, premium, decent, and fair. Diamonds could be any color, from the poorest (J) to the greatest (D), depending on the scale. There are eight distinct values for the clarity trait, ranging from the lowest possible clarity (I1) to the highest possible clarity (IF). Values for depth, table, price, and x, y, and z can range from integers to float points. Table, Width, and Depth of a Diamond are Measured as Depicted in Fig.2. Since the cost of a diamond is based largely on these four factors—carat weight, diamond cut quality, color, and clarity—they deserve special attention.

## C. Data Pre-Processing

The dataset is preprocessed and optimized so that it used effectively during training, as illustrated in the flowchart of Fig. 1. There are two distinct subsets of the dataset: the training data, which is used to develop the models, and the testing data, which is used to evaluate the developed models and determine the values of the relevant performance parameters. All of the models' performance parameter values are acquired after training and testing. These numbers are utilized to determine the optimum regression models, which are then used for diamond price forecasting. The supervised regression models are fed these parameters, and an USD-based forecast is produced. We employed graphical representations to examine the distributions of the dataset's three categorical variables: cut, clarity, and color. Across all three categories, we find sufficient data to enable our algorithm to understand Fig 3,4,5. Because there would be less reference points for each category if the lines were skewed, the machine learning system would have difficulties in learning.

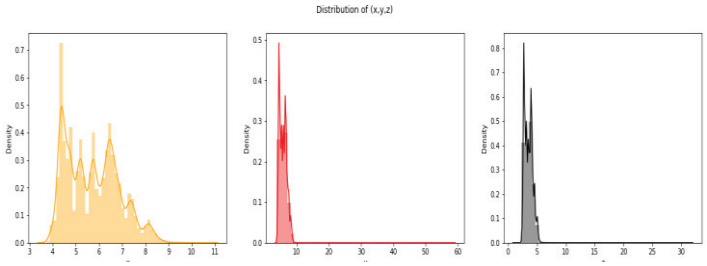


Fig. 3. Distribution of x,y,z

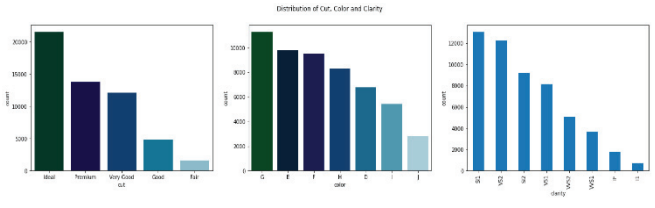


Fig. 4. Distribution of Cut, Clarity and color

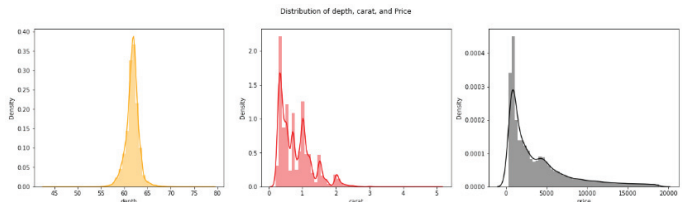


Fig. 5. Distribution of Depth, Carat and Price

We use plots to learn about the continuously factors' distribution, and a slightly skewed bell curve is what we want to see most of the time. The variables carat, depth, and table are skewed; however, the parameters x, y, and z are skewed, therefore we need to eliminate extreme values to get them nearer to the bell - shaped curve.

### *D. Label Encoding*

Various pre-processing techniques, such as Labeling Encode, are applied to the real dataset prior to training the various ml models using it, in order to optimize the dataset for training. To improve training efficiency for ml algorithms, we conduct Labeling Encode on the dataset to encode its categorical features as numeric values.

TABLE II. LABEL ENCODING

| *Carat* | *Cut* | *Color* | *Clarity* | *Depth* | *Price* | *Table* | *x* | *y* | *z* |
|---|---|---|---|---|---|---|---|---|---|
| 0.23 | 2 | 1 | 3 | 61.5 | 55.0 | 326 | 3.95 | 3.98 | 2.43 |
| 0.21 | 3 | 1 | 2 | 59.8 | 61.0 | 326 | 3.89 | 3.84 | 2.31 |
| 0.23 | 1 | 1 | 4 | 56.9 | 65.0 | 327 | 4.05 | 4.07 | 2.31 |
| 0.29 | 3 | 5 | 5 | 62.4 | 58.0 | 334 | 4.20 | 4.23 | 2.63 |
| 0.29 | 3 | 5 | 5 | 62.4 | 58.0 | 334 | 4.20 | 4.23 | 2.63 |

Using Labeling Encode, we transform the dataset's properties from their original alphanumeric form into meaningful numerical identifiers in Table 1. These elements are labeled from 0 to n, where n is the number of categories. Categorical variables in the diamond dataset are encoded using Label Encoding.

These include cut, color, and clarity. In addition, 70% of the records are kept for training and 30% are kept for testing, thus dividing the dataset in half. Finally, we maintain a random state of '42' to ensure that we are drawing from the same pool of potential samples throughout our investigation.

### *E. Correlation of Features*

The concept of correlation is useful for understanding the connections between different variables. These factors could be elements of the input data that have been employed in previous attempts to predict the outcome variable. With the use of correlation, a statistical method, we would learn about the connections between two variables and how they develop over time. Here, we applied a heatmap to visually show the correlation matrix and locate highly correlated components.

Fig. 6. Correlation Matrix

From the Fig 6 We find a strong relationship between the characteristics x, y, z, as well as carat and the dependent parameter cost. We can easily consider these characteristics while training our models. The characteristics 'depth,' 'cut,' and 'table' have minimal correlation and might be removed. Though the number of characteristics in the collection is small, we have opted to retain it. After carefully examination we analyzed the outliners before and after removal of outliners from the fig 7 and 8.

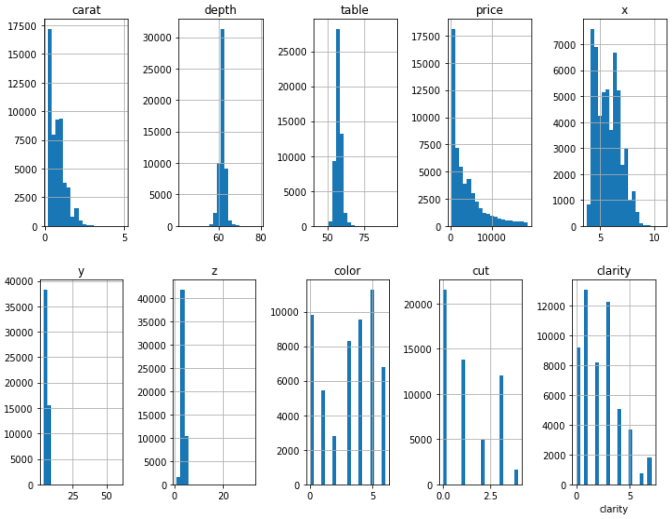


Fig. 7. Outliners

An outlier is a data point in a dataset whose value is extremely out of line with the rest of the data. Data inconsistency or scientific or human mistake during data collection are two possible causes. Statistical testing and model training could both suffer from the presence of outliers. Some of the various methods available for finding outliers include the Z-score, the IQR (Interquartile Range) approach, and DBSCAN clustering. After examining our data using boxplots, we used the IQR technique to eradicate any outliers. These points serve as a brief overview of the method: The first step is to determine the IQR for each column. To solve problem 2, find the constant that corresponds to 1.5 * (IQR). Third, multiply this constant by the third quartile, and then discard any information that is higher than this threshold. Specifically, take the first quartile and subtract the constant, and then get rid of any and all data points that are below the new lower limit. Figure 4 shows the boxplots of the features, which we use to identify outliers in the dataset and exclude using the

interquartile range (IQR) method before continuing to train our model. Figure 7 displays the boxplots that resulted when the data was cleared.

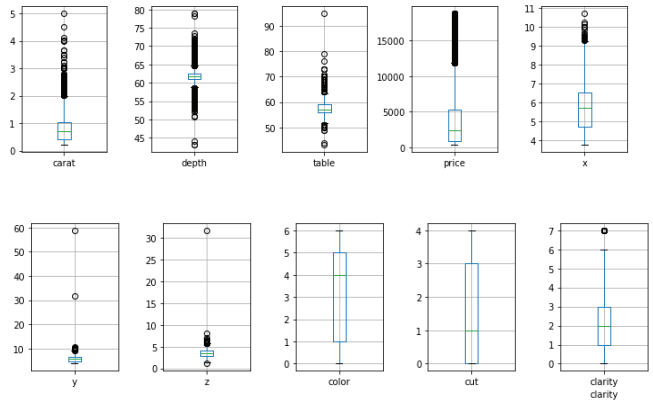


Fig. 8. Removal of outliners

### F. *Linear Regression:*

To forecast target value for inputs that are not included in the data set we have, we apply a technique called linear regression, which involves creating a line that best matches the measured values visible on the plot. A more straightforward Linear regression model makes it less of a challenge to explain how the method fits and what the findings demonstrates. By using regression analysis, we could learn how significant associations are among both predictors. It is possible for us to predict a distinct factor for each attribute, but we also have the capability of obtaining a spectrum of factors along with a threshold of certainty that the value is in that scope.

### G. *Decision trees*

Statistical tests could be quickly and easily used to verify the algorithm's accuracy. Because of this, Decision Trees could be considered a trustworthy structure. Decision trees are useful for understanding the likelihood of various outcomes by breaking down a huge dataset into subgroups that include cases with comparable values [12]. Targeted changeable classes or values could well be predicted using decision trees. There is less work involved in data preprocessing for decision trees than for other approaches. A decision tree could be constructed without first normalizing the data. The data in a decision tree does not need to be scaled in any way. Decision tree construction is not significantly impacted by the presence or absence of missing data [13].

### H. *Random Forest*

The random forest approach utilizes several decision trees to arrive at a categorization. It uses bagging and characteristic pseudo random to create a statistically independent forest of trees whose collective prediction is more accurate than that of any classification algorithm. Utilizing base classifiers from the train data and a random characteristic choice in graph formation, a Random Forest is a collection of unpruned categorization or regressed networks [14].

### I. *XG Boost*

Both a linear regression modeling solution and a tree retraining method are available inside XG Boost. The ability to do several calculations simultaneously on a particular machine is what makes it so quick [15]. Cross-validation could be conducted, and key variables can be identified using its supplementary characteristics. XG Boost excels on organized, medium-sized sets of data with a reasonable number of characteristics and subcategories. This strategy is highly recommended since XG Boost excels at regression and classification which account for most real concerns. The prototype capabilities it provides are delivered quickly because of the efficiency of its boosting method.

### J. *CAT Boost*

In Cat boost, we used parameters to enhance circumstances based on measurable criteria. Error probability monitoring could be done rapidly as well. Cat Boost [16] allows for the incorporation of non-numerical elements, which reduces the need for processing of the data and yields better training results.

## IV. Result Analysis

In order to determine which approach was accurate, we trained machine learning methods on the same data and compared their individual parameters. In order to attain the required effectiveness of the approach, Cross-Validation, a variable selection approach, is used to swap out subsets of the training data repeatedly and arbitrarily, such as verification and analysis, in required to training and validate the algorithm. It is calculated by averaging the total disparity between observed and forecasted outcomes. To put it another way, the MAE indicates that the model's predictions are, on average, off by a larger margin than the true value. For any given model, the lower the MAE, the more accurate the forecast. The average variance between the observed values and the projected values is expressed as a squared error. The advantage of MSE is that it allows us to perform squared, which prevents the elimination of negative terms. Root-mean-squared error (RMSE) is calculated by averaging the squared disparity across the anticipated and observed results. Since RMSE squared the mistakes before averaging them, it is a more effective productivity statistic, especially for penalizing very egregious mistakes. The Cross validation, R2, MAE, RMSE scores of all the models employed for forecasting of diamond price on the same data are shown in a comparison seen in Table. III, IV, V

TABLE III. ANALYSIS

| ***Machine learning Models*** | ***CV SCORE*** | ***R2 score*** | ***MAE*** | ***RMSE*** |
|---|---|---|---|---|
| Linear Regression | 1344.79 | 0.9731761 | 311.7599 | 648.69748 |
| Decision Tree | 49.9695 | 0.9999176 | 2.986569 | 35.944145 |
| Random Forest | 35.9515 | 0.9999508 | 3.236079 | 27.759516 |
| XGB Regressor | 205.509 | 0.9973517 | 126.6803 | 203.82597 |
| Cat Boost | 70.157 | 0.9997169 | 39.14737 | 66.637138 |

Table III shows that Random Forest has the greatest R2 score (0.9999). In terms of Cross validation score 35.9515 lowest observed, and RMSE value is lowest of all models. In terms of accuracy, other models like XG-Boost, K-Nearest, Cat-Boost are all quite close. We have utilized Random Forest to estimate diamond prices based on user input since it outperforms competing methods

TABLE IV. ESTIMATIONS

| ***Machine learning Models*** | ***Over Estimated*** | ***Under Estimated*** | ***Exact Estimation*** | ***Sum of Estimated Values*** | ***Test Values*** |
|---|---|---|---|---|---|
| Linear | 6976 | 6375 | 126 | 13477 | 13477 |
| Decision Tree | 2013 | 2592 | 8872 | 13477 | 13477 |
| Random Forest | 1182 | 6291 | 6004 | 13477 | 13477 |
| XGB Regressor | 7366 | 6072 | 39 | 13477 | 13477 |
| Cat Boost | 6512 | 6822 | 143 | 13477 | 13477 |

Table IV shows Decision Trees model predicted exact estimation of 8872 values. Under estimated, overestimated values of 2013 and 2592

TABLE V. PREDICTED VALUES

| | Linear Regression | XGB | Decision Tree | Random Forest | CAT Boost |
|---|---|---|---|---|---|
| y_test | predicted | predicted | predicted | predicted | predicted |
| 6426 | 6137 | 6639.794 | 6426 | 6425.99 | 6458.79957 |
| 2771 | 2919.2 | 2809.7625 | 2770 | 2770.61 | 2785.255335 |
| 3903 | 4006.2 | 3996.2593 | 3900 | 3901.63 | 3993.801772 |
| 3678 | 3409 | 3745.6956 | 3678 | 3677.08 | 3633.597001 |
| 660 | 742 | 710.1307 | 660 | 660.03 | 711.6079873 |

Table V shows that out of 100 values decision tree predicted prices exact closely and other models like Linear Regression have low Predicted prices. Random forest predicted prices closely and XGB, Cat boost predicted overpriced values

## V. CONCLUSIONS

In this investigation, researchers employed machine learning strategies to anticipate future diamond prices. Following are some findings from this study: We trained, tested, and evaluated different machine learning approaches (Linear Regression, Random Forest, Decision Tree, Cat Boost Regressor, and XGB Regressor), after comparing the performance of different algorithms for predicting diamond prices random forest approach outperforms. The Experimental findings shows that Random Forest Regression predicted accurate values and high R2 score, low RMSE and MAE values. Decision Tree outperforms predicted values Side by side prices. From the literature review we found our model performed better results. Out of all Performance metrics Random Forest low RMSE value, and next close value achieved by Decision Tree.

## REFERENCES


[1] Diamond-The most popular gemstone", [online] Available: https://geology.com/minerals/diamond.shtml.

[2] S. A. Fitriani, Y. Astuti and I. R. Wulandari, "Least Absolute Shrinkage and Selection Operator (LASSO) and k-Nearest Neighbors (k-NN) Algorithm Analysis Based on Feature Selection for Diamond Price Prediction," 2021 International Seminar on Machine Learning, Optimization, and Data Science, 2022, pp. 135-139

[3] G. Sharma, V. Tripathi, M. Mahajan and A. Kumar Srivastava, "Comparative Analysis of Supervised Models for Diamond Price Prediction," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021.

[4] A. C. Pandey, S. Misra and M. Saxena, "Gold and Diamond Price Prediction Using Enhanced Ensemble Learning," 2019 Twelfth International Conference on Contemporary Computing (IC3), 2019.

[5] Singfat Chu (2001) Pricing the C's of Diamond Stones, Journal of Statistics Education, 9: 2.

[6] Chu Singfat (1996) Diamond Ring Pricing Using Linear Regression, Journal of Statistics Education.

[7] S. A. Fitriani, Y. Astuti and I. R. Wulandari, "Least Absolute Shrinkage and Selection Operator (LASSO) and k-Nearest Neighbors (k-NN) Algorithm Analysis Based on Feature Selection for Diamond Price Prediction," 2021 International Seminar on Machine Learning, Optimization, and Data Science, 2022, pp. 135-139

[8] W. Alsuraihi, E. Al-Hazmi, K. Bawazeer and H. Alghamdi, "Machine Learning Algorithms for Diamond Price Prediction", ACM International Conference Proceeding Series, pp. 150-154, 2020.

[9] Marmolejos, José M. Peña. "Implementing Data Mining Methods to Predict Diamond Prices." in 2018 conference of Data Science

[10] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[11] Diamonds dataset (2017). Kaggle datasets repository. https://www.kaggle.com/shivam2503/diamonds.

[12] E. Gyimah and D. K. Dake, "Using Decision Tree Classification Algorithm to Predict Learner Typologies for Project-Based Learning," 2019 International Conference on Computing, Computational Modelling and Applications (ICCMA), 2019, pp. 130-1304.

[13] X. Hu, Y. Yang, L. Chen and S. Zhu, "Research on a Customer Churn Combination Prediction Model Based on Decision Tree and Neural Network," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 129-132.

[14] H. V. Ramachandra, G. Balaraju, A. Rajashekar and H. Patil, "Machine Learning Application for Black Friday Sales Prediction Framework," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 57-61.

[15] H. V. Ramachandra, G. Balaraju, A. Rajashekar and H. Patil, "Machine Learning Application for Black Friday Sales Prediction Framework," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 57-61.

[16] X. Dou, "Online Purchase Behavior Prediction and Analysis Using Ensemble Learning," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 532-53.