[ICAETBM-2024]
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Diamond Price Prediction Using Machine Learning Algorithm

Prof. Vivek Nagargoje[1], Gunjan Parate[2], Vedant Kokane[3], Kalpak Kondekar[4] , Muskan Hannure[5], Dr. Chandrakant Kokane[6]

[1,2,3,4,5,6] Nutan Maharashtra Institute of Engineering and Technology, Pune

ABSTRACT

A set up natural process conflation of carbon known as Diamond, is one of the toughest and most immensely precious material known to men and women. Investments in precious gems like diamonds are in immensely demand. The rate of a diamond, nevertheless, is not that as easily calculated as the value of either platinum or gold, since so numerous factors must be taken into consideration. Because there is such a wide range of diamond qualities and rates; as a result, being suitable to make reliable price prognostications is vital for the diamond sedulity. Although, making accurate foretell is challenging. In this research article, we executed multiple machine knowledge ways applied to the challenge of diamond price auguring's analogous as Random Forest, Linear Retrogression, Decision Tree Random Forest. This configuration thing is to develop an accurate model for estimating diamond prices predicated on its attributes analogous as cut grade, weighting property, and enclose. We evaluated the sum of estimated values and test values of predicted values with overestimated, underrated and exact estimations.

Keywords: Machine Learning Algorithm , Retrogression, Diamond, styling, insert.

I. INTRODUCTION

Diamonds are the most valued precious monuments in the world. It's the most valuable gemstone. Its worth is influence by factors such as its composition, precious of cut, purity level, weight in carats, and a variety of other characteristics. Diamonds are used in numerous places, similar as it's in diligence, and they're effective in cutting polishing and drilling. Based on the concept of the "four Cs" diamond undergo grading and certification. These include color, cut, clarity, and weight. This metric facilitates a consistent understanding for consumers globally when purchasing diamonds, forecasting a consistent understanding for consumers globally when purchasing diamond, fostering trade and ensuring value for their purchases. Diamond prices are typically established daily and are exchanged in US Bones. To enhance the prediction of diamond prices, the Kaggle diamond dataset is employed to explore the interrelationships among factors such as carat weight, price, and color. This analysis utilizes scatter plots to visualize these metrics.. A notable finding is the strong correlation between carat weight and price. However, it's noted that this correlation seems to be wavering currently, resulting in increased uncertainty regarding diamond prices. This machine literacy model analyzes further than 4 characteristics and it can therefore produce a more accurate result. Machine literacy is an advancing technology that empowers computers to autonomously learn from historical data. Machine learning employs various

algorithms to generate mathematical models and make predictions using historical data and information. At present, machine learning is actively employed to accomplish various tasks, including but not limited to image recognition, speech recognition, email filtering, Facebook auto-tagging, and recommender systems. What is machine literacy? In real life, humans have the capacity to learn from their past experiences and their literacy skills.
Can machines learn from experience and historical data in a manner akin to humans? This is where the concept of machine learning comes into play. " Machine literacy enables machine to automatically learn from once data and to ameliorate performance from its experience and to prognosticate new effects without being programmed. Machine literacy is a component of artificial intelligence focused on creating algorithms that enable computers to learn from experience and past data autonomously. Machine literacy predicts the affair grounded on former data set. How does machine literacy work? A machine learning system learns from historical data, constructs a predictive model, and upon receiving new data, generates predictions for it. And the delicacy of prognostications depends upon the quantum of data, the huge quantum of data helps to make good model and ameliorate the delicacy. Then we're using different-different machine learning algorithms like direct retrogression, lariat, Ridge, Elastic Net retrogression and arbitrary timber retrogression. But arbitrary timber gives us to 97 of delicacy which is high delicacy of model. Random Forest Regression Random timber retrogression is a unique machine literacy algorithm that combines the generalities of arbitrary timbers and retrogression analysis.

II. PROBLEM STATEMENT

The thing is to prognosticate price of given diamond (Retrogression Analysis).

There exist 10 independent variables: Id, Carat, Cut, Color, Clarity, Depth, Table, x, y, and z.

ID: - unique identifier of each diamond.

Carat denotes the distinctive unit of weight measurement specifically used for weighing gemstones and diamonds.

Cut: Reflects the quality of diamond cutting.

Color – Color of diamond.

Clarity: - Diamond clarity is measure of chastity and oddity of the gravestone, Graded by the transparency of these attributes under - power exaggeration.

Depth: refers to the height of a diamond, measured in millimeters from the Culet (lowest point) to the table (flat, top surface).Table diamond's table is the hand which can be seen when the gravestone is viewed face up.

X: Dimension of the diamond along the X-axis.

Y: Dimension of the diamond along the Y-axis.

Z:- Dimension of the diamond along the Z-axis.

Target Variable Price of diamonds.

III. LITURATURE SURVEY

Numerous studies have tried to prognosticate diamond prices using colorful ways. For illustration, José (2) employed data mining ways similar as M5P, direct retrogression, and neural networks, using the M5P model showing a high position of delicacy. Singh et al. [3] employed multiple linear regression (MLR) to analyze the associations between diamond prices and the four Cs: carat weight, cut, color, and clarity. Multiple Linear Regression (MLR) is widely acknowledged as a suitable data mining model for analyzing diamond datasets.. Numerous machine learning algorithms have been employed to forecast diamond and gold prices. However, the main challenge lies in choosing the most suitable model,

utilizing pre-processing techniques and correlation analyses (4). generally, prices diamonds are expressed in US bones, but the dependency between price and carat weight isn't always direct. This is because weighted diamonds are generally highly precious than lighter bones, and the drift of a high connection between carat weight and price seems to be grow feeble (5).

To gain a deeper understanding of this correlation and the fluctuation in diamond prices concerning heavier specimens, we utilize a scatter plot visualization of the Kaggle diamond dataset, which aids in comprehension.

The Significance of Automated Diamond Price Prediction

1. Delicacy Automated diamond price vaticination harnesses the power of data and algorithms to offer largely accurate price estimates. This is especially vital in an assiduity where small differences in grading can restate into significant variations in price.

2. Efficiency Traditional diamond pricing styles are frequently time- consuming and reliant on mortal moxie. robotization accelerates the process, making it more effective and cost-effective.

3. Translucency by using machine literacy models, pricing factors and their influences come more transparent. Buyers and merchandisers gain perceptivity into how different characteristics impact a diamond's value.

4. Threat Mitigation Predictive models can help alleviate pitfalls associated with diamond deals, icing that both buyers and merchandisers have a more objective understanding of a diamond's worth.

5. Request perceptivity the data- driven nature of machine literacy can uncover request trends and pricing patterns that may scape mortal spectators. This information can inform pricing strategies and request opinions.

Structure of This Study

This exploration paper embarks on a trip through the world of automated diamond value vaticination using machine literacy. It encompasses the following crucial factors.

Data Collection A comprehensive dataset of diamond characteristics, containing the fresh factors and Four Cs , serves as the foundation for this article.

Point Engineering The dataset undergoes scrupulous point engineering to prize meaningful perceptivity from the raw data. This process involves transubstantiating, homogenizing, and opting applicable features to feed into the machine literacy models.

Machine Learning Models colorful machine learning algorithms, including retrogression and ensemble styles, are applied to prognosticate diamond prices. Model selection and tuning are performed to enhance predictive accuracy and mitigate potential biases.

Evaluation The performance of the machine literacy models is strictly assessed using applicable evaluation criteria, icing that the prophetic delicacy aligns with assiduity norms.

Interpretation The prophetic models are interpreted to understand the relative significance of different attributes in determining diamond prices. This analysis offers perceptivity into the pricing dynamics of diamonds.

Practical operations the counteraccusations of automated diamond price vaticination are explored, containing their eventuality to improve pricing strategies, request translucency, and threat operation.

I. METHODOLOGY

1.LINEAR REGRESSION

The relationship between dependent and independent variables is modeled using a forward equation of the following form.

$Y = \beta_0 \beta_1$ - Y stands for dependent variable. - X_1, X 2,., X_k are independent variables. - β_0, β_1, β 2,., β_k are parts of the model that represent the intercept and stage of the independent variable. - ε stands for an error term that allows for unexplained changes.

simple. multiple regression - Simple forward regression includes one independent variable, and multiple forward regression has two or more independent variables.

2. DESICION TREE ALGORITHM:

This algorithm is supervised machine learning algorithm. It is a tree like structure and by using decision tree algorithm we can make decisions.By using decision tree we can solve the problems like classification and regression appropriately.

Overfitting introduced in decision tree. Strategies such as pruning and setting maximum depth can help reduce overfitting.

Evaluation - Decision tree models are evaluated using criteria such as Gini contamination, entropy or mean square error (for regression). Depending on your problem, you can evaluate performance based on sensitivity, superiority, recall, or F1 score.

3.SUPPORT VECTOR MACHINE ALGORITHM

Support vector machine is also used for solve the problems like classification and regression. still, it's primarily used for bracket problems in machine literacy. Support vector machine efficient when we have a small number of data. To solve the classification problem we use SVM classifier while for regression problem we use SVM regressor.

4.RANDOM FOREST ALGORITHM

1) Random Forest is an ensemble learning algorithm widely used in machine literacy for both bracketing and regression problems. This is a complex and important algorithm that can improve the sensitivity and reliability of characterization models. Random forests are based on the idea of combining predictions from multiple decision trees to produce a more reliable and accurate final confirmation.

The methodology for automated diamond price vaticination using machine literacy is designed to give a structured and methodical approach to erecting prophetic models for predicting diamond prices. It involves the following way

1. Data Collection Data Sources Acquire a advanced dataset of diamonds that includes attributes similar as Carat weight, Cut, Color, Clarity, and other applicable features. The dataset should also include the corresponding diamond prices.

2. Data Quality Assurance Conduct data cleansing and preprocessing to handle any missing values, outliers, and inconsistencies within the dataset. Insure that the dataset is in a format suitable for machine literacy.

3. Point Engineering point Selection Identify and elect applicable features for price vaticination. This may include Carat weight, Cut, Color, Clarity, depth chance, table chance, and other attributes that impact diamond prices.

4. Point Transformation Perform any necessary metamorphoses on the data, similar as garbling categorical variables, spanning numerical features, and normalizing data.

5. Data unyoking Train- Divide the dataset into a training set and a testing set. Generally, a 70- 30 or 80-20 split is used, where the training set is used to train machine literacy models, and the testing set is used for evaluation.

6. Model Selection Machine Learning Algorithms Choose suitable machine learning algorithms for retrogression tasks. Common choices include Linear Retrogression, Decision Trees, Random timber, Support Vector Machines (SVM), and grade Boosting. Model Tuning Fine- tune hyperparameters of the named machine literacy models using ways like grid hunt or arbitrary hunt to optimize their performance

7. Model Training the Models Train the named machine literacy models on the training dataset using the optimized hyperparameters. This involves fitting the models to the training data to learn the connections between the features and diamond prices.

8. Model Evaluation Performance Metrics evaluate the effectiveness of trained models using relevant regression criteria, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and R-squared (R2) score. Visual examination Visually check the prognosticated diamond prices compared to the factual prices to gain perceptivity into model delicacy and implicit impulses.

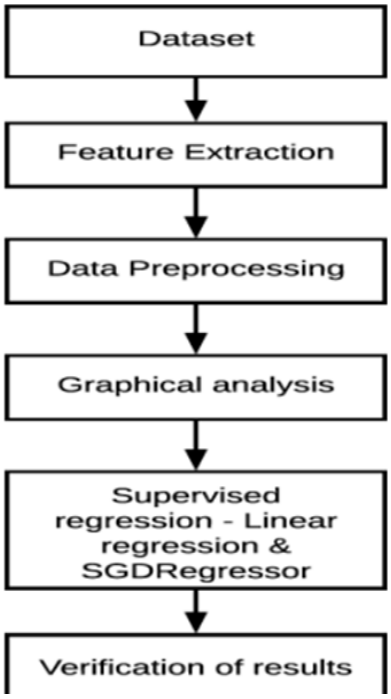
9. Model Interpretation point significance dissect the point significance scores of the models of machine learning understand which attributes have the most significant influence on diamond prices.

10. Model Deployment (Optional) Real- World operation If applicable, emplace the trained model in a real- world setting to automate diamond price prognostications. This may involve creating a stoner-friendly interface or integrating the model into being systems.

11. Obstacles and Restrictions Data Restriction Admit any restrictions or difficulties with the approach that arose, such as problems with the quality of the data or the requirement for specialized knowledge in the field.

12. Unborn Directions Enhancements bandy implicit advancements or unborn directions for the automated diamond price vaticination system, similar as incorporating fresh features or exploring deep literacy approaches.

By following this methodology, experimenters and interpreters can totally develop and estimate machine literacy models for automated diamond price vaticination. The performing models offer a data-driven approach to estimating diamond prices, furnishing precious perceptivity and delicacy in the diamond pricing process.



	carat	cut	color	clarity	depth	table	x	y	z	price
0	1.52	Premium	F	VS2	62.2	58.0	7.27	7.33	4.55	13619
1	2.03	Very Good	J	SI2	62.0	58.0	8.06	8.12	5.05	13387
2	0.70	Ideal	G	VS1	61.2	57.0	5.69	5.73	3.50	2772
3	0.32	Ideal	G	VS1	61.6	56.0	4.38	4.41	2.71	666
4	1.70	Premium	G	VS2	62.6	59.0	7.65	7.61	4.77	14453

In this data set, we have neither missing value or nor duplicate value present. After performing some EDA correlation between variables:

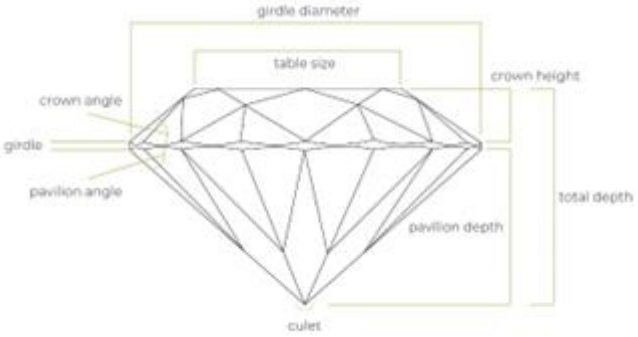
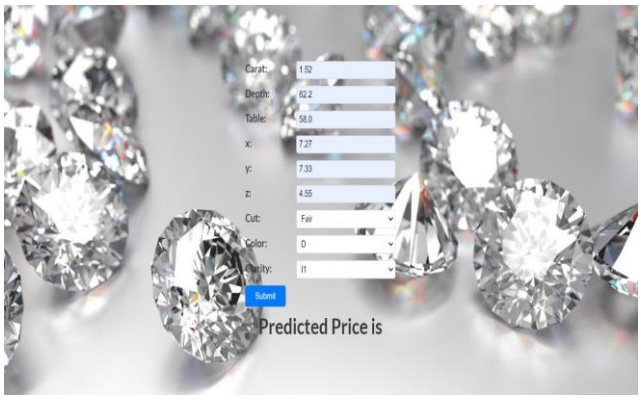


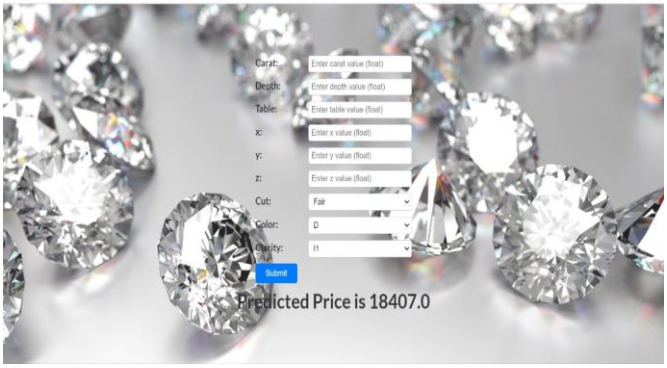
Fig 1: Diamond Features

IV. RESULT ANALYSIS

After conducting a range of experiments and analyzing the outcomes, it can be deduced that supervised learning techniques such as linear regression, decision trees, and K-nearest neighbors (KNN) can be effectively employed for estimating diamond prices.The Decision Tree Regressor algorithm exhibited outstanding performance, achieving an accuracy of approximately 87.49% to 88. In unborn work, it would be helpful to incorporate unsupervised models to further improve the delicacy and wholesome of diamond price prognostications using the dataset.

Output:





I. CONCLUSION

In the realm of diamond pricing, the integration of machine literacy has steered in a new period of delicacy, effectiveness, and translucency. This exploration ventured into the intricate world of self-operating diamond price vaticination using machine literacy, and the trip uncovered several crucial findings and achievements

1.Prophetic delicacy Machine literacy models demonstrated their capability to prognosticate diamond prices with a high degree of delicacy. They exercised the power of data, effectively landing the intricate connections between diamond attributes and pricing.

2.Translucency The operation of machine literacy offered translucency in the pricing process. It revealed the relative significance of attributes like Carat weight, Cut, Color, and Clarity in determining diamond prices, furnishing precious perceptivity for assiduity stakeholders.

3. Effectiveness Automated diamond price vaticination streamlined the valuation process. What was formerly a time- consuming and expert-dependent task can now be fulfilled efficiently, reducing the time and cost associated with price determination.

4.Threat Mitigation Machine literacy models contributed to threat mitigation in diamond deals. Both buyers and merchandisers gained a more objective understanding of a diamond's value, reducing the eventuality for controversies and misconstructions.

VII. Future Scope

Unborn compass While this exploration has made significant strides in automated diamond price vaticination, there is ample room for unborn disquisition and advancement

1. Deep literacy Integration Incorporating advanced techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) could potentially enhance predictive accuracy even more. Deep literacy models can uncover complex patterns within diamond data.

2. Market Dynamics Future exploration can claw into the analysis of request dynamics and external factors impacting diamond prices, similar as profitable conditions, consumer preferences, and global trends

3. Real- Time Pricing Developing real- time pricing models that can acclimatize to changing request conditions and give up- to- the- nanosecond valuations for diamonds. **4. instrument and Authentication** Integrating machine literacy into instrument and authentication processes to corroborate the authenticity and origin of diamonds, addressing enterprises related to conflict diamonds and ethical sourcing.

5. Marketplace Integration Integration of automated price vaticination models into online commerce, enabling buyers and merchandisers to gain instant valuations for diamonds listed for trade.

6. Blockchain Technology Investigating the application of blockchain technology to create transparent and unalterable records of diamond characteristics and transactions, thereby bolstering confidence in the industry's diligence.

7. Ethical Considerations Incorporating ethical considerations into price vaticination models, icing that ethical and sustainable practices are reflected in diamond valuations.

8. Global Expansion Extending exploration and operations beyond traditional diamond requests to arising requests where diamonds are gaining fissionability.

As technology continues to advance and the diamond assiduity evolves, the community between machine literacy and diamond pricing will continue to shape the future of this age-old request. The hunt for perfection, translucency, and sustainability in the diamond assiduity will drive further invention, icing that diamonds, with their dateless beauty and appeal, remain both a symbol of enduring love and a testament to the power of data- driven perceptivity.