5th International Conference on Smart Electronics and Communication (ICOSEC 2024)
IEEE Xplore Part Number: CFP24V90-ART; ISBN: 979-8-3315-0440-3

Diamond Price Prediction using Machine Learning Techniques

Kapil Amadavadi
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, India
kapilamadavadi@gmail.com

Rashmi Rane
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, India
rashmi.rane@mitwpu.edu.in

Ruhi Patankar
Department of Computer Engineering and Technology
Dr. Vishwanath Karad MIT World Peace University
Pune, India
ruhi.patankar@mitwpu.edu.in

2024 5th International Conference on Smart Electronics and Communication (ICOSEC) | 979-8-3315-0440-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICOSEC61587.2024.10722317

Abstract — **Diamonds, composed of carbon formed through natural processes, stand as one of the most coveted and valuable materials globally, particularly among women. The intricate nature of diamond pricing, impacted by diverse factors including weight, cut grade, and dimensions, poses a challenge for accurate valuation. When choosing a technique for a research study, It is important to take into the nature of the data, including its type, distribution, and any missing values or outliers. The technique should align with the study's objectives, balancing model complexity with the need for interpretability, especially if the results must be understood by stakeholders. In this study, multiple machine learning techniques, including Linear Regression, Random Forest, Decision Tree, KNeighbors, XGB Regressor, Extra tree regressor, Gradient boosting regressor, and multi-layer perceptron were employed to forecast diamond prices. Through a comparative analysis of supervised machine learning models, focusing on accuracy and performance metrics such as R2 score, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), insights were gained into the efficacy of each model. Experimental findings revealed Extra tree as the most proficient model, exhibiting high R2 scores and low RMSE and MAE values, indicative of its superior predictive accuracy. This research contributes to the ongoing efforts to optimize diamond price projection, offering valuable insights for the diamond industry and beyond.**

Keywords— Machine Learning Algorithms, Diamond Price Prediction, Linear Regression, Extra tree.

I. INTRODUCTION

Diamonds, considered as one of the most prized gems globally, stand unparalleled in market value and desirability among gemstones. Renowned for their optical brilliance and durability, diamonds command a price tag several times higher than any other gemstone. Their allure extends beyond mere aesthetics, finding significance in cultural traditions, fashion, and various industrial applications owing to their exceptional hardness and light-manipulating properties. Traditionally, the value of diamonds has been determined by factors such as carat weight, cut, clarity, and color, collectively known as the "4Cs of Diamond Quality." However, the dynamics of diamond pricing have changed recently due to a number of variables such as market trends and technical developments. As a result, the traditional metrics may not fully capture the nuances of diamond valuation in today's market. In order to better estimate diamond values, researchers have resorted to sophisticated machine learning algorithms. By leveraging datasets such as the Kaggle diamond dataset and employing supervised regression algorithms, researchers aim to detect the complex relationships between diamond attributes and market prices. Through comparative analyses of regression models such as Linear Regression, Decision Tree, Random Forest, and others, researchers seek to determine which method is best for predicting prices.This paper synthesizes insights from multiple studies in the field, introducing a comprehensive approach to predicting diamond prices using machine learning. By examining the interplay of various diamond attributes and market dynamics, this research endeavors to enhance our understanding of diamond valuation in contemporary contexts. The subsequent sections present a review of related literature, the proposed methodology, experimental findings, and conclusions, offering valuable insights into the complex domain of diamond pricing.

II. LITERATURE SURVEY

In this survey paper various regression algorithms are used such as Linear Regression, Support Vector Regression, Random Forest Regression, Decision Tree, Huber, Passive Aggressive, Bayesian, Extra Tree, K-neighbor, XGBoost and CatBoost. The CatBoost Regression is the most suitable Algorithm for it [1]. Two machine learning algorithms k-NN and LASSO are used in this work. The authors have tested the k-NN and LASSO models with various alpha and k values. The k-NN approach outperforms the LASSO approach in terms of results [2].

The best features from the data have been selected for this survey through feature selection techniques such principal component analysis, recursive feature selection, and the Chi-square test. For forecasting, regression models such as random forest and linear regression are utilized [3]. The study focuses on using various machine learning approaches, including Cat Boost Regressor, Random Forest, Decision Tree, and XGB Regressor, to the problem of diamond price predictions. They used cross-validation to determine the degree to which the model varies from the real [4].

Numerous techniques, such as M5P, Random Forest, Multilayer Perceptron, Decision Stump, REP Trees, and M5Rules, are highlighted in this survey [5]. Eight distinct supervised models, including decision trees, random forests, lasso regression, ridge regression, linear regression, ElasticNet, AdaBoost Regressor, and Gradient-Boosting Regressor, were studied in this study. When compared to other supervised learning algorithms, the Random Forest yielded a superior overall outcome [6].

Similar to other previous papers, this work also employed supervised learning techniques, such as decision trees, linear regression, and KNN, which are helpful in estimating diamond prices. The Decision Tree Regressor, with an accuracy range from 87.49% to 88%, was the best algorithm. [7]. Various machine learning models, such Gradient Boosting Regressor, Support Vector Regression, Lasso, Random Forest, Decision Tree, KNeighbors, XGBRegressor, ElasticNet, and RidgeCV, have been used in this work. The results showed that the XGBoost Regressor model performed more effectively than the other models [8].

Several machine learning algorithms, including as support vector machines, neural networks, decision trees, random forests, and linear regression, can be employed in this analysis to predict diamond costs. With a 98% accuracy rate, Random Forest uses many trees to generate the most accurate results [9]. The present investigation report Select machine learning methods that are appropriate for regression challenges. Gradient Boosting, Random Forest, Decision Trees, Support Vector Machines (SVM), and Linear Regression are popular options [10].

This study introduces a multiclass classification approach using Extreme Learning Machines (ELMs), optimizing hyperparameters to address dataset imbalance and improve accuracy. The Regularized ELM model demonstrated the highest performance, achieving 83.75% accuracy with efficient training times [11]. The study uses machine learning models, particularly ensemble methods like Bagging with REP Tree, in order to forecast diamond prices using a variety of factors. It shows that these advanced techniques significantly improve prediction accuracy compared to simpler models, highlighting the potential of machine learning to refine diamond valuation despite some challenges [12].

This research compares Decision Tree and Random Forest models for diamond classification, with Random Forests outperforming due to better handling of data noise and overfitting. The study suggests future improvements using image-based data and advanced techniques like Deep Learning for more precise predictions [13]. This work explains that while Russian-made diamond drilling tools, like those from Terekalmaz, are technically comparable to foreign brands, their market share suffers due to inconsistencies in quality and a lack of focus on customer service and modern manufacturing methods. To regain competitiveness, Russian companies need to improve product consistency and better adapt to market demands [14]. The study identifies nine key factors influencing online diamond jewelry purchases: online shopping attitudes, price and quality, sensory risk, merchant credit, goods display, shopping process, website design, after-service, and professional services. The most critical factors are online shopping attitudes, price and quality, sensory risk, and merchant credit. Businesses should focus on these areas to improve consumer trust and satisfaction, enhancing their online marketing strategies [15].

III. METHODOLOGY

A. *Workflow of Proposedmodel*

Below Fig. 1 is represent the block diagram of working this regression model.

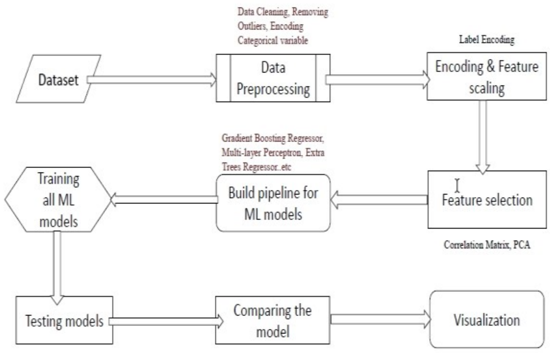

Fig. 1: Workflow of Proposed model

B. *Dataset*

Kaggle serves as a platform where individuals engage in the exchange and discourse surrounding machine learning and data science methodologies. It offers convenient access to numerous datasets, facilitating the training of machine learning models. Users not only share their models but also have the opportunity to develop and experiment with various tools. Our supervised machine learning models, for instance, were trained using the Diamond file collection.

The dataset comprises 53,940 entries, each describing different aspects of diamonds. It contains the carat weight, which ranges from 0.2 to 5.01, and the price in US dollars, which fluctuates between $326 to $18,823. While the color is rated from J (worst) to D (best), the cut quality is categorized as Fair, Good, Very Good, Premium, or Ideal. On a scale ranging from I1 (worst) to IF (best), clarity is evaluated. The diamond's measurements are also included in the dataset: depth (z) = 0 to 31.8 mm, width (y) = 0 to 58.9 mm, and length (x) = 0 to 10.74 mm. Additionally, the total depth percentage, calculated as 2 times the depth divided by the sum of length and width, ranges from 43% to 79%, and the table, which shows how big the diamond's top is in relation to its widest point, spans from 43 to 95%.

The dataset also includes a market demand variable, categorized as high, medium, or low.

C. Data Description

The 53,940 distinct items in the dataset for diamond collecting each have unique characteristics listed in Table 1. By leveraging the qualities present in the data, we predict the best estimate of diamond prices. These qualities are commonly referred to as the 4Cs of diamonds, a terminology widely used by experts in the diamond and jewelry industry. Fig. 2 is described dimension of diamond.

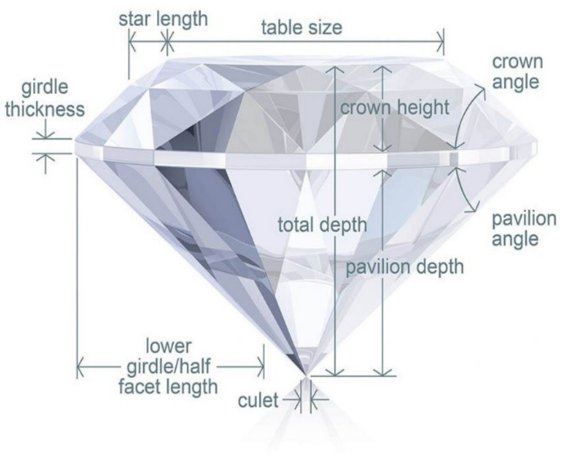

Fig 2: Diamond Dimensions

TABLE1: 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
Market demand	(high, low, Medium)

Table 1 indicates that the weight of the diamonds in carat whose ranges from 0.21 to 5.01 kilograms. The cut quality is categorized into five levels: excellent, premium, decent, and fair. A range of hues, from the lowest (J) to the brightest (D) on the scale, may be seen in diamonds. The clarity trait encompasses eight distinct values, spanning from the least clarity (I1) to the greatest (IF). The values of attributes like dimensions (x, y, and z), table, depth, and price can be either integer or floating-point. The evaluation of a diamond's table, width, and depth is depicted in Table 1. Considering that a diamond's carat(weight), cut(quality), color, and clarity primarily determine its price. It emphasizes how important they are.

D. Data Pre Processing

The data preprocessing steps include data cleaning, which involves correcting or removing inaccuracies and missing values; identifying and removing outliers to reduce the impact of extreme data points; and encoding categorical variables into numerical values to enable their use in machine learning models. These actions guarantee that the dataset is correct and prepared for analysis.

In a few essential steps were made during our research's data preprocessing phase to guarantee the dataset's quality and applicability for machine learning analysis. Data cleaning techniques were employed to eradicate discrepancies, absent values, and mistakes, thus augmenting the dependability of the dataset. To lessen the impact of erroneous or extreme data points on model performance and training, outlier detection and removal approaches were also used. Moreover, methods like Label Encoding were used to encode category information into numerical representations, facilitating the integration of these features into machine learning algorithms. These preprocessing steps collectively contributed to optimizing the dataset for subsequent analysis and model training, ultimately enhancing the robustness and accuracy of our research outcomes.

E. Outlier

Data points known as outliers differ considerably from the remainder of the dataset; these anomalies are frequently the result of errors or discrepancies in the data that were collected.

Their presence can adversely affect statistical testing and model training. Various methods, including the Z-score, Interquartile Range (IQR), and DBSCAN clustering, are employed to detect outliers. We create a scatter plot with a regression line using Seaborn and Matplotlib, visualizing the relationship between "price" with different parameter of the dataset. This plot helps identify outliers and trends between these two variables.

F. Label Encoding

Preprocessing methods such as Label Encoding are used to boost the dataset for training before training various machine learning algorithms on the real dataset. By allocating an integer to each distinct category, the label encoding approach transforms categorical data into numerical values. This makes it possible for machine learning algorithms to handle and comprehend categorical input effectively. To transform the dataset's category attributes into numerical values, label encoding is applied, thereby enhancing training efficiency for machine learning algorithms. Table II demonstrates the Label Encoding process, where attributes such as Carat, Cut, Color, Clarity, Depth, Price, Table, and dimensions (x, y, z) are encoded.

This transformation assigns meaningful numerical identifiers ranging from 0 to n, where n represents the number of categories. In the diamond dataset, categorical characteristics like cut, color, and clarity are encoded using label encoding. To ensure an equitable distribution, additionally 20% of the dataset is put aside for testing and the remaining 80% for training. To maintain consistency, a random state of '7' is maintained throughout the investigation, ensuring consistent sampling from the dataset.

Fig. 3 is showing the data description after label encoding.

	carat	cut	color	clarity	depth	table	price	x	y	z	market_demand
0	0.23	2	1	3	61.5	55.0	326	3.95	3.98	2.43	1
1	0.21	3	1	2	59.8	61.0	326	3.89	3.84	2.31	1
2	0.23	1	1	4	56.9	65.0	327	4.05	4.07	2.31	1
3	0.29	3	5	5	62.4	58.0	334	4.20	4.23	2.63	1
4	0.31	1	6	3	63.3	58.0	335	4.34	4.35	2.75	1

Fig. 3: Data description after encoding

G. *Correlation of Features*

Understanding the interrelationships among various variables is pivotal, and correlation serves as a valuable tool for uncovering these connections. By employing statistical methods, such as correlation analysis, we gain insights into how different variables evolve over time and their impact on predicting the outcome variable. Utilizing a heatmap representation of the correlation matrix, visually identification of highly correlated components is done. Significant relationships between factors like x, y, z, carat, and the dependent parameter, cost, were found by correlation matrix analysis. Consequently, these variables were prioritized during model training. Conversely, attributes such as depth, cut, and table exhibited minimal correlation and were deemed less influential. Despite the dataset's limited number of characteristics, we chose to retain it for analysis purposes. Below Fig. 4 shows the Correlation Matrix

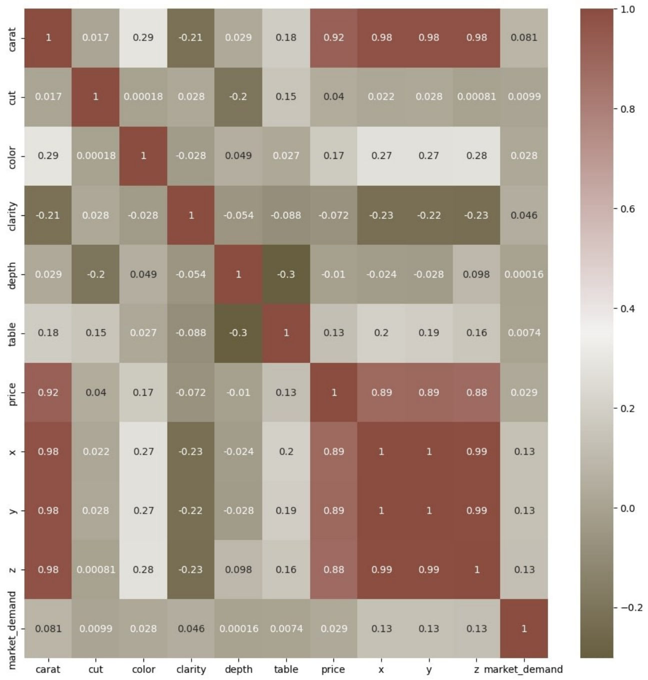

Fig 4. Correlation matrix

H. *Traditional ML Regression Algorithms*

In this methodology section, several machine learning algorithms are employed to develop predictive models for this research.

- Linear Regression: A basic strategy that works well for regression jobs is Linear Regression, It aims to make a straight line connection between the goal variable and the input data. Because linear regression is simple to use and interpret, it facilitates understanding of the connections between variables. The relationship's strength and direction are shown by the coefficients, which offer a clear picture of how changes in the independent factors affect the variable that is reliant..

- Decision Trees: Decision trees are adaptable models that efficiently handle both regression and classification issues by recursively partitioning the feature space according to the most informative qualities. The results are easily interpreted. The tree structure makes it simple to communicate to stakeholders the thinking behind forecasts by visually representing decision routes based on many criteria.

- Random Forest: it is an group approach, aggregates forecasts from many of decision trees to improve generalization and minimize overfitting, offering robust performance across various datasets. Random Forest has the ability to manage missing values and outliers and is capable of handling a broad variety of data types, including those with both category and numerical properties.

- K-Nearest Neighbors: A non-parametric technique called KNN uses the majority vote of its closest neighbors to classify data items, making it particularly useful for classification tasks with complex decision boundaries. By leveraging these diverse algorithms, we aimed to explore different modeling approaches and select the most suitable one for our specific research objectives.

I. *Advanced ML Regression Algorithms*

In the research methodology used, advanced regression algorithms have been incorporated to enhance the accuracy and predictive power of models.

- XGBRegressor: XGBRegressor high-performance implementation of gradient boosting, utilizes an optimized gradient boosting framework, offering superior speed and efficiency.

- Extra Tree Regressor: Extratree Regressor or Extremely Randomized Trees, is an extension of Random Forest that adds more unpredictability to the process of creating trees, Improved generalization and decreased variance are the outcomes of an extension of Random Forest.

  An ensemble learning technique called the Extra Trees Regressor creates several decision trees, introducing extreme randomness by selecting both features and split points randomly at each node. Unlike traditional decision trees that choose the best split based on criteria like Gini impurity, Extra Trees increases diversity among trees by this randomization, reducing overfitting and variance. This method is more advanced than traditional algorithms because it combines high efficiency, robustness, and lower variance, making it appropriate for complex and big datasets.

- Gradient Boosting Regressor: The Gradient Boosting Regressor is an ensemble machine learning approach that builds a model gradually to fix errors in previous models by combining weak learners, often decision trees. Each new tree is added to minimize the residual errors of the combined prior trees, leading to a strong predictive model. This iterative process allows Gradient Boosting to achieve high accuracy, making it effective for complex regression tasks.

- Multi-layer Perceptron (MLP): Artificial neural networks come in the form of multi-layer perceptron that can recognize complex non-linear relationships in data because it is composed of numerous layers of interconnected neurons.

  It is composed of an input layer, one or more hidden layers, and an output layer, among other layers of cell. All neurons within a layer are linked to all neurons within the following layer, and During training, weights between neurons are adjusted to minimize the discrepancy between predicted and actual outcomes.

The methodology was enhanced by incorporating advanced regressor algorithms, allowing for the exploration of sophisticated modeling techniques and leveraging their capabilities to effectively address the complexities of the research problem.

IV. RESULTS

Since the Extra Tree Regressor can handle data variability and noise well, it probably fared better in predicting diamond prices than other algorithms. Its random splitting of data points and features helps prevent over fitting while capturing complex non-linear relationships among features like carat, cut, and clarity.

Extratree regression emerges as the best-performing model with a 98.6882% R-squared value, achieving the highest accuracy among the regression techniques analyzed. This result highlights the effectiveness of Extratree regression in accurately predicting outcomes based on the input features. Additionally, the root mean square error (RMSE) values further support the superior performance of Extratree regression, with the lowest RMSE among all algorithms analyzed.

Fig. 5 represents mean squared error, mean absolute error, square root of mean squared error, and r^2 score of all regression models.

	Model	MAE	MSE	RMSE	R Squared
0	LinearRegression	868.394566	1.690523e+06	1300.201081	0.890147
1	DecisionTree	306.324151	3.592800e+05	599.399717	0.976653
2	RandomForest	235.310540	2.036025e+05	451.223322	0.986770
3	KNeighbors	336.992116	3.463113e+05	588.482167	0.977496
4	XGBRegressor	243.022850	2.072289e+05	455.223983	0.986534
5	ExtraTreeRegressor	232.813817	2.018685e+05	449.297824	0.986882
6	Gradient Boosting Regressor	350.715475	3.446717e+05	587.087489	0.977603
7	Multi-layer Perceptron	437.575150	5.250700e+05	724.617146	0.965880

Fig 5. Table for visualization

Below Fig. 6 is represent the bar chart of all regression models vs it's r2 value and through which we analyze the accuracy of all models.

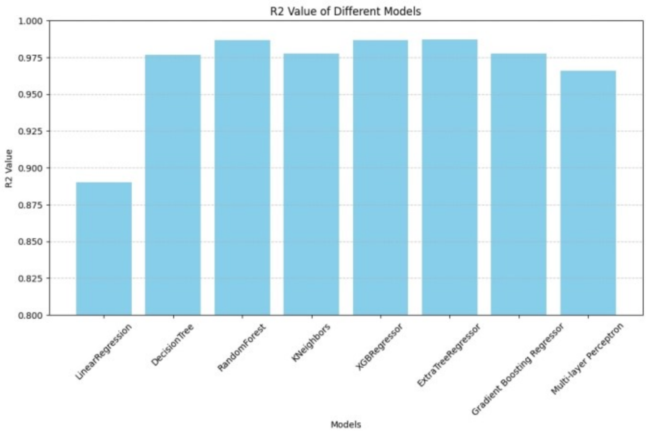

Fig 6. Bar chart of models vs R2 score

Below Fig. 7 is represent the bar chart of all regression models vs it's r2 value

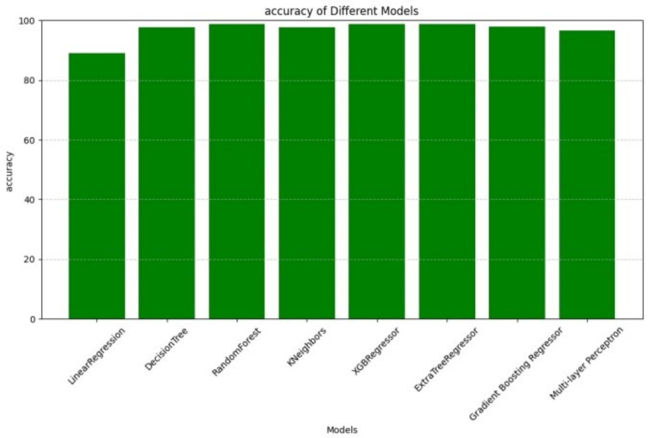

Fig. 7 Bar chart of Models vs Accuracy

V. FUTURE SCOPE

In the future, we can increase accuracy of diamond price prediction model by using following terms:

- Expanded Parameter Set: Including additional diamond attributes like fluorescence and symmetry for a more comprehensive analysis.
- Bias Mitigation: Implementing strategies to address database bias and ensure model fairness.
- Real-Time Data Integration: Incorporating up-to-date market data from online sources to capture dynamic pricing trends.
- Advanced ML Techniques: Employing cutting-edge algorithms like Neural Networks and SVMs for enhanced predictive accuracy.

VI. CONCLUSION

With an R-squared value of 98.6882%, Extratree regression is clearly the best-performing model in this research, which analyzes many regression techniques. It also achieves the highest level of precision. This result highlights the effectiveness of Extratree regression in accurately predicting outcomes based on the input features. Additionally, the root mean square error (RMSE) values further support the superior performance of Extratree regression, with the lowest RMSE among all algorithms analyzed. In conclusion, Extratree regression emerges as the preferred choice for regression tasks due to its exceptional accuracy and lower prediction error rates compared to other algorithms. These findings underscore the significance of algorithm selection in achieving accurate and reliable predictive models. In the future, more investigation and verification of Extratree regression in a variety of datasets and practical applications may offer insightful information on its resilience and efficacy in a range of contexts.

REFERENCES

[1] Mihir, Harshvadan, et al. "Diamond price prediction using machine learning." 2021 2nd International Conference on Communication, Computing and Industry 4.0 (C2I4). IEEE, 2021.

[2] Fitriani, Shafilah Ahmad, Yuli Astuti, and Irma Rofni 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 (ISMODE). IEEE, 2022.

[3] Pandey, Avinash Chandra, Shubhangi Misra, and Mridul Saxena. "Gold and diamond price prediction using enhanced ensemble learning." 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE, 2019.

[4] Basha, Md Shaik Amzad, et al. "An efficient machine learning approach: Analysis of supervised machine learning methods to forecast the diamond price." 2023 International Conference for Advancement in Technology (ICONAT). IEEE, 2023.

[5] Basha, Md Shaik Amzad, et al. "Exploring Machine Learning Models to Predict the Diamond Price: A Data Mining Utility Using Weka." 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2023.

[6] Sharma, Garima, et al. "Comparative analysis of supervised models for diamond price prediction." 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021.

[7] Mankawade, Amruta &Kokate, Chinmay & Soman, Koustubh& Mohite, Atharva &Vispute, Ayush & More, Omkar. (2023). Diamond Price Prediction Using Machine Learning Algorithms. International Journal for Research in Applied Science and Engineering Technology.

[8] Zhang, Hairong. (2023). Prediction and Feature Importance Analysis for Diamond Price Based on Machine Learning Models. Advances in Economics, Management and Political Sciences.

[9] T. Ramaswamy , S.Sanjana , T Meghana, MACHINE LEARNING ALGORITHMS FOR DIAMOND PRICE PREDICTION ,European Chemical Bulletin,

[10] Suryadevara, Chaitanya Krishna. "Sparkling Insights: Automated Diamond Price Prediction Using Machine Learning." *A Journal of Advances in Management IT & Social Sciences* (2016).

[11] Ramírez, José, et al. "Extreme learning machines for predict the diamond price range." 2023 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON). IEEE, 2023.

[12] Basha, Md Shaik Amzad, et al. "Gems of Prediction: From Clarity to Carats-Unveiling Diamond Prices with Machine Learning in Waikato Environment for Knowledge Analysis." 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT). IEEE, 2024.

[13] Swain, Debabrata, Avani Chaturvedi, and Vishesh Rathi. "GemInsight: Unleashing Random Forest for Diamond Quality Forecasting." 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2023.

[14] Ataev, Pavel L., Oxana V. Islamova, and Zakir Zh Berov. "Determination of the Technical Level and Quality of Diamond Drill Bits, Represented in the World Market." 2018 IEEE International Conference" Quality Management, Transport and Information Security, Information Technologies"(IT&QM&IS). IEEE, 2018.

[15] Yang, Yan-ping, and Qi Li. "A Quantitative Research on Influencing Factors of the Behavior That Consumers Buying Diamond Jewelry on Line." 2010 International Conference on Internet Technology and Applications. IEEE, 2010.

[16] Kumar, GV Saatwik, and K. Jaisharma. "Improve the Accuracy for Flight Ticket Prediction using XGBRegressor Optimizer in Comparison with Extra TreeRegressor Performance." 2023 6th International Conference on Contemporary Computing and Informatics (IC3I). Vol. 6. IEEE, 2023.

[17] Keprate, Arvind, and RM Chandima Ratnayake. "Using gradient boosting regressor to predict stress intensity factor of a crack propagating in small bore piping." 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE, 2017.

[18] Kumar, D. Sumith, and N. P. G. Bhavani. "Improving the Accuracy for Predicting Solar Power Using the Novel Gradient Boosting Regressor Algorithm in Comparison With the RANSAC Regressor Algorithm." 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS). IEEE, 2023.

[19] Xu, Xueli, et al. "Towards prediction of bus arrival time using multi-layer perceptron (mlp) and mlp regressor." 2023 8th International Conference on Business and Industrial Research (ICBIR). IEEE, 2023.

[20] Rao, N. Sri Sai Venkata Subba, and S. John Justin Thangaraj. "Flight Ticket Prediction using Random Forest Regressor Compared with Decision Tree Regressor." 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM). IEEE, 2023.

[21] Swain, Debabrata, Avani Chaturvedi, and Vishesh Rathi. "GemInsight: Unleashing Random Forest for Diamond Quality Forecasting." 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2023.

[22] Alsuraihi, Waad, et al. "Machine learning algorithms for diamond price prediction." Proceedings of the 2020 2nd international conference on image, video and signal processing. 2020.

[23] Kigo, Samuel Njoroge, Evans Otieno Omondi, and Bernard Oguna Omolo. "Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model." Scientific Reports 13.1 (2023): 17315.

[24] OuYang, Zhe. "Research on the Diamond Price Prediction based on Linear Regression, Decision Tree and Random Forest." Highlights in Business, Economics and Management 24 (2024): 248-257.

[25] Acharya, Kalpa Nigam, and Manas Ranjan Senapati. "Diamond Price Prediction Using Data Mining Techniques." International Conference on Research in Management & Technovation. Singapore: Springer Nature Singapore, 2023.