Diamond Price Prediction using Machine Learning Algorithm

Prof. Vivek Nagargoje, Muskan Hannure, Vedant Kokane, Kalpak Kondekar, Gunjan Parate, Vilas Deotare

Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra, India

ARTICLEINFO

ABSTRACT

Diamond, a set up natural process emulsion of carbon, is one of the hardest and most immensely precious material known to men, especially further to women. Investments in precious gems like diamonds are in significant demand. The rate of a diamond, nonetheless, isn't as fluently calculated as the value of either gold or platinum since so numerous factors must be taken into account. Because there's such a broad range of diamond confines and rates; as a result, being suitable to make dependable price prognostications is pivotal for the diamond assiduity. Although, making accurate prognostications is grueling . In this study, we enforced multiple machine literacy ways employed to the challenge of diamond price soothsaying's similar as Linear Retrogression, Random Forest, Decision Tree Random Forest. This composition's thing is to develop an accurate model for estimating diamond prices grounded on its characteristics similar as weighting factor, cut grade, and confines. We compared the sum of estimated values and test values of prognosticated values with overrated, undervalued and exact estimations.

Keywords :— Machine Learning, Regression, Diamond, Styling, Insert

I. INTRODUCTION

In terms of request value, diamonds are by far the most sought- after gem kind. Diamonds demand a price label that's knockouts of times advanced than that of any other rocks . Diamond's optic characteristics, or how it interacts with light, contribute to its wide sun. Diamonds are popular for a variety of reasons outside their egregious durability and life, including the fashion assiduity, artistic significance, and directors' patient advertising. When it comes to jewelry, diamonds are at the veritably top. the world's most precious particulars, as it possesses the astounding capability of spreading light far and wide. The worth of a diamond is determined by colorful factors, including its shape, cut, eliminations(contaminations), and weight in carats. Diamonds have colorful artificial operations due to their superior slicing, buffing, and piercing capabilities. The tremendous worth of rocks has assured their continued worldwide trade for glories. Faceted diamonds are estimated grounded on their color, cut, clarity, and carat weight to assess their overall quality. In the 1950s, the Gemological Institute of America created a standardized fashion for assessing diamonds quality known as the" 4Cs of Diamond Quality." As a rule, the weight of accoutrements similar

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 applicable. Since diamonds are so precious, indeed 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 product 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 booby-trapped and cut. When a diamond is fated for use as an garnishment, it goes through a number of procedures to enhance its appearance and change it into a stunning decoration mortal practice of adorning oneself. The original cost of a diamond depends on its oddity, size, and the time and trouble needed to polish, cut, and mine them. The right price depends on several factors, not just one from the gems. The four Cs relate to the color, clarity, carat, cut, and donation of the rock . confines like depth, range, 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.

II. RELATED WORK

The objective of this study is to conduct a comprehensive examination of various supervised machine learning models employed for the prediction of diamond prices. Predicting diamond prices is a challenging task due to the multifaceted nature of factors influencing their value, including carat weight, cut, color, clarity, and market trends. This research explores the performance of multiple supervised learning algorithms, such as linear regression, decision trees, random forests, support vector machines, and neural networks, in the context of diamond price prediction. The paper likely includes a comparative analysis of these models in terms of predictive accuracy, robustness, and computational efficiency. It may

employ a dataset containing historical diamond price data and various diamond attributes.[1]

Published in 2023, the second paper focuses on the significance of land mapping and its regression, driven by the growing demand for precise geographical information. In this study, the authors aim to evaluate the efficiency and effectiveness of diverse supervised machine learning techniques for forecasting diamond prices. The research likely employs a dataset featuring historical diamond price information and associated attributes. It assesses the performance of various supervised machine learning models, including linear regression, decision trees, random forests, support vector machines, and potentially others, in terms of their predictive accuracy, computational efficiency, and robustness. The authors may also investigate the impact of feature engineering and model hyperparameter tuning to improve prediction quality. This research provides valuable insights for those seeking to utilize machine learning in addressing diamond price prediction challenges.[2]

The paper titled 'Least Absolute Shrinkage and Selection Operator (LASSO) and k-Nearest Neighbors (k-NN) Algorithm Analysis Based on Feature Selection for Diamond Price Prediction' explores the application of LASSO and k-NN machine learning algorithms in the prediction of diamond prices. It commences with an introduction highlighting the importance of diamond price prediction and the necessity for effective feature selection techniques.

III. METHODOLOGY/EXPERIMENTAL CONCLUSION

These tools are Jupyter lab, Matplotlib, Numpy, Seaborn and Sklearn. We use Jupyter Labs' interact ive development platform to write code and visualize data. J upyter Labs is a simple place to prepare and organize data s uitable for machine learning and data science. The main adv antage is writing plugins

and new products and integration with Crontab. Matplotlib is a python library for visualization and animation to analyze given data. Nump y is a Python library that works faster for processing arrays t han lists. Pandas is a powerful tool for data analysis. It is als o a flexible program developed by Python for data cleaning, merging, organizing, selecting and organizing. Seaborn is a l ibrary developed by matplotlib for making plots and visualizations. Built on Matplotlib and tightly integrated with the pandas data model. We have tested 4 machine learning algorithm to get the most accurate result and compare which one is the best, the 4 machine learning models are as follows

1. LINEAR REGRESSION

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

$Y = \beta_o \beta_1$ - Y stands for dependent variable. - X_1, X_2,., X_k are independent variables. - β_o, 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 allowsfor unexplained changes.

Simple. multiple regression - Simple forward regressionincludes one independent variable, and multiple forward regression includes 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. Techniques 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

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.
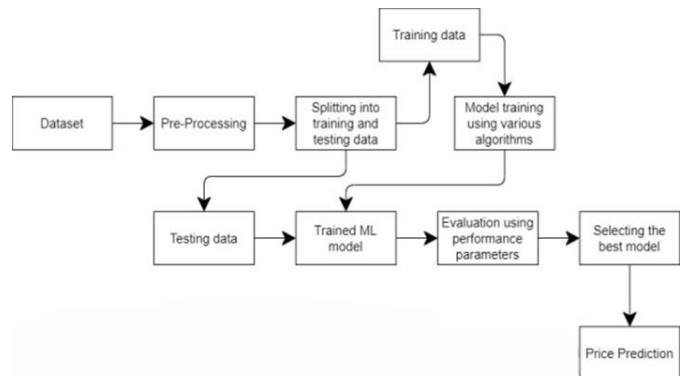
IV.CONCLUSION

When investing, knowing how to forecast diamond prices is critical.This work contains an in-depth evaluation along with studies estimating diamond pricing using criteria more compared to the renowned 4Cs.

However, setting the parameter value affects the accuracy results. In addition, the selected features: carat, x, y,z, are not sufficient to predict the price of diamonds due to various factors, such as the quality of

diamonds and reactions from social media that affect investment. Therefore, to get the best results from the model, the dataset should constantly be updated and added data.

V. FUTURE SCOPE

Talking about the future prospects, further efforts will be put into solving difficulties like blurry image of certificate, incorrect values of attributes, etc. We would also try to introduce the number of features such as shape, table value, polish, symmetry, etc. to obtain more accurate results.

VI. REFERENCES

[1]. G. Sharma, V. Tripathi, M. Mahajan, and A. K. Srivastava, "Comparative analysis of supervised models for diamond price prediction," in 2021

[2]. A. C. Pandey, S. Mishra, and M. Saxena, "Gold and diamond price prediction using enhanced ensemble learning," in 2019.

[3]. Kotak, Pranav, and Hiral Modi. "Enhancing the data mining tool WEKA." In 2020.

[4]. G. Sharma, V. Tripathi, M. Mahajan and A. Kumar Srivastava, "Comparative Analysis of Supervised Models for Diamond Price Prediction," IEEE

[5]. Kokane, Chandrakant D., and Sachin D. Babar. "Supervised word sense disambiguation with recurrent neural network model." Int. J. Eng. Adv. Technol.(IJEAT) 9.2 (2019).

[6]. Kokane, Chandrakant D., Sachin D. Babar, and Parikshit N. Mahalle. "Word Sense Disambiguation for Large Documents Using Neural Network Model." 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT). IEEE, 2021.

[7]. Kokane, Chandrakant, et al. "Word Sense Disambiguation: A Supervised Semantic Similarity based Complex Network Approach."

International Journal of Intelligent Systems and Applications in Engineering 10.1s (2022): 90-94.

[8]. Kokane, Chandrakant D., et al. "Machine Learning Approach for Intelligent Transport System in IOV-Based Vehicular Network Traffic for Smart Cities." International Journal of Intelligent Systems and Applications in Engineering 11.11s (2023): 06-16.

[9]. Kokane, Chandrakant D., et al. "Word Sense Disambiguation: Adaptive Word Embedding with Adaptive-Lexical Resource." International Conference on Data Analytics and Insights. Singapore: Springer Nature Singapore, 2023.