

aaa

by Aa Aa

Submission date: 08-Jun-2023 08:47AM (UTC-0500)

Submission ID: 2111771947

File name: sing_Machine_Learning_Algorithms-soft_computing_j_component.docx (1.43M)

Word count: 4270

Character count: 23960

Diamond Price Prediction and Classification using Machine Learning Algorithms

8

PROJECT REPORT

Submitted in fulfilment for the J Component of ITA 6004 – Soft Computing

in

M.C.A

By

Hrishikesh S G (22MCA0162)

Rajat Singh (22MCA0139)

Srijan Paria(22MCA0266)

8

Under the guidance of

Dr. BRIJENDRA SINGH

SITE



VIT[®]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

8

School of Information Technology and Engineering

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
1.	Abstract	3
2.	Introduction	3
3.	Literature Review	4-8
4.	Objectives of the project	9
5.	Dataset Description	9
6.	Proposed Methodology	10-13
7.	Experimental results	14-16
8.	Discussion and conclusion	17
9.	References	18

Abstract - Diamonds are one of the most precious and valuable commodities, and their prices are influenced by several factors, including carat, cut, clarity, and color. With the help of machine learning algorithms, it is possible to predict diamond prices more accurately [2]. In this study, classification algorithms, like XgBoost, AdaBoost, Decision Tree, Gradient Boosting, Random Forest, Logistic Regression, Naïve Bayes were utilized to classify diamond prices. The performance metrics used include accuracy, precision, recall, and F1-score. This paper throws light on the use of machine learning in the “fluctuative” world diamond industry and provides insights into the importance of selecting the appropriate algorithm for accurate price classification. Overall, this paper adds to the existing literature on the use of machine learning in the diamond industry and highlights the potential of this technology for improving the accuracy of diamond price prediction.

Keywords - Diamonds, Machine learning algorithms, Price prediction, Classification algorithms, Accuracy, Diamond industry Performance.

INTRODUCTION

The price of diamonds is determined by various factors, namely color, clarity, cut and carat [4]. The process of diamond price determination might be complex and time-consuming, requiring a thorough understanding of the diamond grading system and market trends [1]. ML algorithms have been increasingly used in the diamond industry to automate and improve the process of diamond price classification. ML algorithms by nature are good at analyzing large data and identify patterns that are not apparent to human experts [5]. This can help diamond traders and appraisers make more accurate and informed decisions about the value of a diamond.

This research paper focuses on the use of machine learning algorithms for diamond price classification.

Specifically, the performance of four popular algorithms is explored: XGBoost, AdaBoost, Random Forest, and Gradient Boosting. These algorithms were trained on a dataset of over 10,000 diamonds with known prices and various features such as carat weight, color, clarity, and cut.

Literature Review:

Ref. No.	Title	Algorithms Used	Evaluation Metric	Data Set
1	⁶ Comparative Analysis Of Supervised Models For Diamond Price Prediction	¹⁵ Linear regression, Random Forest, Lasso Regression, DecisionTree, Ridge Regression, ElasticNet, AdaBoostRegRessor, and GradientBoostingRegr essor	RMSE, Accuracy	Kaggle repository
¹⁰ 2	Machine Learning Algorithms For Diamond Price Prediction	Linear regression, Forest regression, Gradient Boosting ²⁵ gressor, Polynomial regression, Neural Network.	MAE(Mean Absolute Error), RMSE(Root-mean-square deviation)	Kaggle repository
3	¹ Comparative Study Of Predicting Diamond Ring Prices In Online Retail Shop	¹ Multiple Linear Regression, Random Forest, Deep Neural Network	MAE, MAPE(Mean absolute percentage error)	Kaggle repository
4	² Comparison Of Machine Learning Algorithms For Predicting Diamond Prices Based On Exploratory Data Analysis	² Ridge Regression, LASSO Regression, ElasticNet, Random Forest Regression, XGBoost, Support Vector Regression	²² Mean Absolute Error, Mean Squared Error	Online Diamond Store
5	¹ Subjectivity Of Diamond Prices In Online Retail: Insights From A Data Mining Study	¹³ Multiple linear regression, Decision forest , Boosted regression trees, Artificial neural network	⁹ Mean absolute error, Mean absolute percent error	Online Diamond Store
6	⁵ Fatigue Driving Detection With Modified Ada-Boost And Fuzzy Algorithm	Ada-Boost, Fuzzy		NA

7	⁵ Comparison Of Gradient Boosting And Extreme Boosting Ensemble Methods For Webpage Classification	Gradient Boosting, Extreme Gradient Boosting	Accuracy, Recall, Precision, and F1 score	NA
8	⁷ An Improved Xgboost Model Based On Spark For Credit Card Fraud Prediction	XGBoost	³ Recall, F1-Score, and AUC (Area under the ROC Curve)	NA
²¹ 9	Comparison Of Classification Techniques Used In Machine Learning As Applied On Vocational Guidance Data	NAIVE BAYES, ONER, KSTAR, JRIP	³¹ Mean absolute Error, Root mean squared error, Kappa statistic	NA
10	⁷ Improved Ada Boost Classifier For Sports Scene Detection In Videos: From Data Extraction To Image Understanding	Ada Boost	Accuracy	NA (Video)

⁶

Comparative Analysis of Supervised Models for Diamond Price Prediction:

The cost of diamonds varies depending on their characteristics. In this paper, they have conducted a comparison investigation and implemented a number of supervised models⁶ to forecast the diamond's price. In the study, they have compared the performance of eight different supervised models, including linear regression, lasso regression, ridge regression, decision tree, random forest, ElasticNet, AdaBoost Regressor, and Gradient Boosting Regressor. The model that performed the best overall and produced the most accurate results was highlighted. The goal of this study is to preprocess data, identify correlations between dataset attributes, train the aforementioned models, assess their correctness, evaluate their results, and ultimately determine which model performs the best, which is the Random Forest Regression Model.

32

Machine Learning Algorithms For Diamond Price Prediction:

The ability of the diamond dealers to estimate the price accurately is crucial. The enormous variety of the sizes and qualities of diamond stones makes the prediction procedure challenging. Many machine learning methods, including Linear regression, Random forest regression, polynomial regression, Gradient descent, and Neural networks, were utilised in this study to aid in the prediction of diamond price. After developing numerous models, evaluating their efficacy, and evaluating the outcomes, it turns out that the random forest regression is the most effective one.

25

Comparative Study Of Predicting Diamond Ring Prices In Online Retail Shop:

By leveraging data from online retail diamond ring stores, this study intends to create models for estimating the retail prices of jewellery. There are 187,821 records for loose diamonds and 2,206 records for rings. This study develops and compare a performance of three models consist of Multiple Linear Regression (MLR), Random Forest (RF), and Deep Neural Network (DNN). The evaluation metrics used for comparing algorithms are accuracy of prediction using MAE and MAPE. The results show that MAE for the ring price prediction of MLR, RF, and DNN are \$688.36, \$235.33, and \$273.00, respectively. In addition, MAE for a diamond price prediction of MLR, RF, and DNN are \$3254.03, \$450.44, \$445.94, respectively.

13

Comparison Of Machine Learning Algorithms For Predicting Diamond Prices Based On Exploratory Data Analysis:

A sizable dataset of loose diamonds scraped from an online diamond store is exposed to data mining in order to explore how a diamond's physical characteristics could predict its price. The results show that diamond weight, colour, and clarity are the most significant drivers of diamond pricing. So, submit a proposal for an exploratory data analysis that contains a section that uses LASSO regression, ElasticNet regression, and Random Forest regression to analyse various aspects of news articles. This technique uses historical data to estimate diamond prices while maintaining a simple to understand trading strategy. In terms of prediction accuracy and interpretability, the suggested approach outperforms state-of-the-art techniques like extreme learning machines that use deep learning. The information suggests that news impact is important for forecasting.

20

Subjectivity Of Diamond Prices In Online Retail: Insights From A Data Mining Study:

Diamonds are under a special class of goods whose perceived value is heavily influenced by socially formed ideas. They have utilised data mining on a sizable dataset of loose diamonds scraped from an online diamond store to examine the extent to which the physical characteristics of a diamond can be used to predict the diamond price. They discovered that the main factors affecting diamond prices are diamond weight, colour, and clarity. The data mining findings also point to a significant level of subjectivity in diamond pricing, which may be a result of diamond dealers' use of price obfuscation techniques.

5

Fatigue Driving Detection With Modified Ada-Boost And Fuzzy Algorithm:

The most significant and evident factors for detecting fatigued driving are facial features. In this study, the modified Ada-Boost algorithm is used to precisely recognise faces and pinpoint their eyes and mouths. The parameters of the eyes and mouth state are extracted using the adaptive threshold. Finally, a fuzzy algorithm paired with PERCLOS principles is employed to assess the level of weariness. The proposed method has proven to be more resilient, faster, more precise, and able to handle real-time demands, according to experiments.

5

Comparison Of Gradient Boosting And Extreme Boosting Ensemble Methods For Webpage Classification:

Classifying web pages is a crucial effort in many fields, including web content screening, contextual advertising, and maintaining or growing web directories, among others. Web page classification using machine learning techniques has been found to be effective, and ensemble models have been used to enhance the output of single classifiers. In this work, binary classification is performed using the ensemble models of Gradient Boosting and Extreme Boosting. The dataset including web page URLs was compiled by hand. Extreme boosting's gain in accuracy and speed was confirmed by a comparison of the two boosting algorithms. The speed and accuracy of extreme boosting have been shown to be around ten times faster than those of gradient boosting.

7

An Improved Xgboost Model Based On Spark For Credit Card Fraud Prediction:

For numerous financial organizations, credit card theft results in substantial financial losses. An enhanced XGBoost model based on Spark is suggested due to the unbalanced dataset and vast amount of data in the field of credit card fraud. The Smote algorithm was employed in this project to balance the training set. The fraud detection system also employed the Spark-based XGBoost classifier. The test sets were then simultaneously classified at the end. The model provided in this project is contrasted with the logistic regression model, decision tree model, random forest model, and the original XGBoost model in the experiment comparing models.

19

Comparison Of Classification Techniques Used In Machine Learning As Applied On Vocational Guidance Data:

The computerization of commercial operations by firms and recent advancements in information systems have made data analysis quicker, simpler, and more accurate. In more and more applications, including those in medicine, finance, education, and energy, data analysis methods like data mining and machine learning are being applied. Data processed through data mining can be used to derive useful additional information using machine learning techniques. Such valuable and important data enables firms to develop their future policies on a more solid foundation and to realise considerable time and cost savings. This study uses data mining and machine learning approaches to apply classification algorithms to data collected from individuals during the vocational guidance process.

12

Improved Ada Boost Classifier for Sports Scene Detection In Videos: From Data Extraction To Image Understanding

17

This study proposes an improved Ada Boost classifier for the sports scene detection in movies, from data extraction to picture understanding. Virtual reality technology is the prerequisite for the creation of a VR sports simulation system, and it is this technology that allows for precise reproduction of all competitive sports. This method can serve as a technical guide for coaches and athletes, on the basis of which the training tools are continually improved, the training effect is enhanced, and physical injury to the athletes is avoided. The low-level features or high-level features are typically used in conventional scene identification techniques. Although these methods have straightforward advantages that are simple to put into practice, this study employs the Ada Boost model for an effective examination.

Objectives of project:

- To develop a robust and accurate model for predicting diamond prices based on their attributes such as carat weight, cut, clarity, and color.
- To explore and compare the performance of different machine learning algorithms in predicting diamond prices.
- To evaluate the effectiveness of soft computing techniques in handling the complex interrelationships between diamond attributes and price.
- To investigate the impact of feature selection and feature engineering techniques on the accuracy of diamond price prediction.
- To identify the most influential attributes in determining diamond prices and their relative importance.
- To develop a classification model that accurately assigns diamonds to appropriate price ranges.
- To compare the performance of different classification algorithms in categorizing diamonds based on their prices.
- To analyze the strengths and weaknesses of each classification algorithm in the context of diamond price classification.
- To determine the optimal combination of features and algorithm parameters for achieving the highest classification accuracy.
- To provide insights and recommendations for diamond industry professionals on the selection and implementation of machine learning algorithms for price prediction and classification.
- To contribute to the existing literature on the application of machine learning in the diamond industry, specifically in the context of price prediction and classification.

Dataset Description:

This classic dataset originally contained the prices and other attributes of almost 54,000 diamonds. However, 14184 of those seem to be the same diamonds, measure from a different angle. This can be found out but checking for duplicated value when disregarding the variables x, y, z, depth and table, which are dependent on the angle.

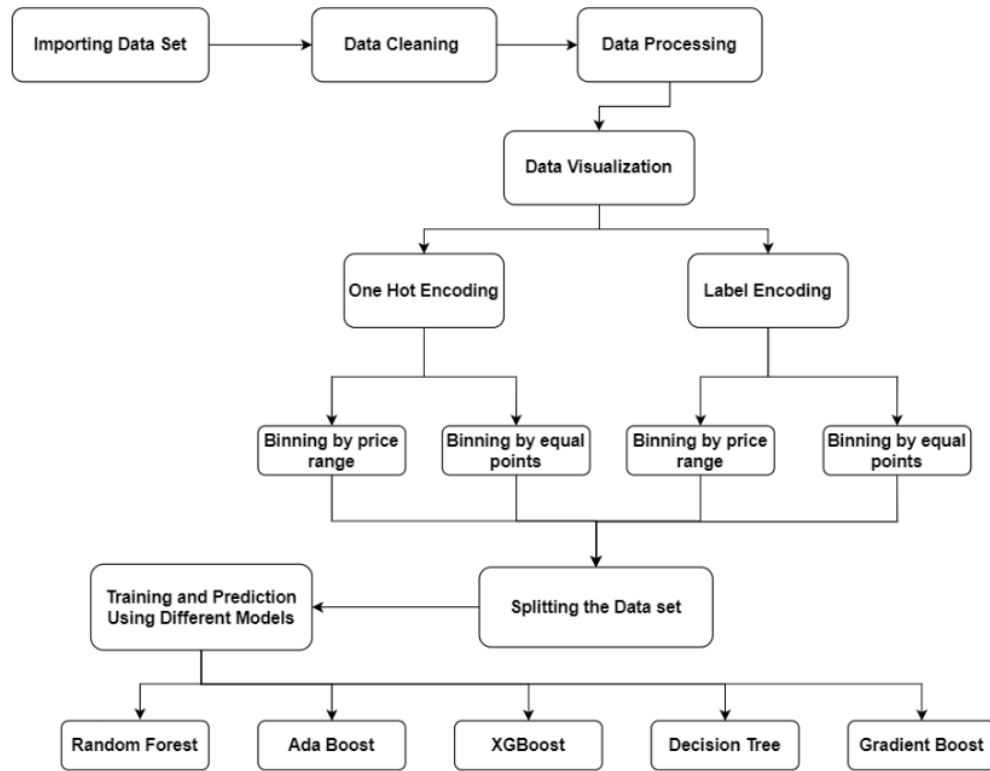
Attributes : "carat", "cut", "color", "clarity", "depth", "table", "price", "x", "y", "z"

Kaggle Repository : <https://www.kaggle.com/datasets/ulrikthgepedersen/diamonds>

Github Code: <https://github.com/rajat-singh1999/diamonds-price-prediction-and-classification>

Proposed Methodology:

Flow chart:



[Fig. 1] Flow Chart

Machine Learning:

The procedure of using computational methods to learn information directly from the data. This learning should be done without relying too much on a predetermined equation. This whole process is called Machine learning. The data samples used to train the model is very important. When the data models increase, the accuracy of this model might also increase. Since the process of Machine learning largely involves a lot of statistical computations, the data used to train such a model needs to be clean and relevant to the intended goals of the project. The data used should be consistent, that is, it should not have too many null values and

understandingly the tabulated data should be correct. The relevance of the data used is important because if the data is not relevant than the model will predict wrongly for real world cases. For example, if we need a model to predict a disease in an individual, we need the medical data of that person rather than the data about his TV watching habits. The TV habits data can be useful for the model where subscription to a TV plan is predicted. There are many machine learning algorithms that use different statistical computation techniques to train a model. We have used some of these models during this research.

Hyperparameter Tuning

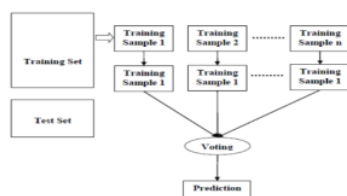
Machine Learning models are mathematical models associated with several parameters. These parameters are to be learned by the same machine learning model. These parameters are learnt by the model by training the model with the existing data. This process is called fitting the data to the model. There is another kind of parameters that cannot be directly learned by a model through the regular training process, known as Hyperparameters. These parameters are fixed even before the actual training process begins. Significant properties of a model are expressed by such parameters. The complexity or how fast a model must learn is also ascertained by Hyperparameters. In practice, there are two main strategies that are used to implement Hyperparameter tuning, Randomised Search and Grid search. For grid search technique, all the possible hyperparameters including the intermediate combinations of the hyperparameters are tried, that is, for each hyperparameter a new model is created. This is the reason why grid search is more computationally expensive. The Randomised search Tuning strategy is better than the “GridSearch” strategy because it tries only a fixed number of hyperparameters for the model. For this research, the “RandomizedSearchCV” python module is used.

Decision Tree

In each decision tree, the internal nodes stand for the features of the data set, leaf nodes for outcomes and branches for decision rules. All in all, this classifier is an ML algorithm which generates a model that can classify using a tree data structure. It is easy to implement and comprehend because tree structures are easy to understand. A tree structure also mirrors a human being’s decision-making process, that is, the process of choosing one of the options at each step until the goal is reached or final decision is made. A decision tree starts with a question and based on the possible outcomes; it splits the tree at every level. The main issue in the decision tree algorithm is to choose the best attributes for the root node and the sub nodes. To do this task, we use the technique called ASM. With ASM the best possible attribute can easily be selected for the nodes of the tree.

Random Forest

In some cases, the problem is such that a single decision tree might not be enough to solve the problem. Random Forest consists of multiple decision trees associated with logical parts of the dataset. The Average is taken at the end and then checked if we are getting better accuracy. In most cases we get better accuracy. Random Forest is best suited for problems related to regression and classification. Random forest comes under the umbrella of ensemble learning. The process of using different classifiers together to tackle a complex problem and to improve the overall accuracy of the solution is called Ensemble learning. This is obvious that a greater number of trees in the forest will lead to better accuracy but also take a toll on performance. Random Forest is useful because it takes less time to train, and predicts the output with high accuracy. It performs well even when a part of the data is missing.



[Fig. 2]

XG-Boost

XgBoost stands for Extreme Gradient Boosting. Its library is written in C++, so it optimises the Gradient boost training process. As the name suggests, XgBoost attempts to boost the Gradient Boosting model. Boosting is an ensemble modelling technique where several weak classification models are combined to form a stronger classification model. This process simply adds weak models one after the other. At the beginning a model is created on the data and then a second model is added in series. The second model tries to correct the errors made by the model previous to the current one. This cycle repeats itself until the maximum number of models is added or the complete data set is predicted. For XgBoost, the multiple decision tree creation follows a sequence. Weights are an important part of this process. The independent variables that are fed to the decision tree which predicts results are assigned with weights first. If an independent variable is predicted wrongly, the weight of that variable is increased and then fed to the next decision tree. These individual trees or classifiers are then ensembled together to obtain a more precise and stronger model.

Measure of Performance: Confusion Matrix

The table that lists the number of correct as well as incorrect guesses. The effectiveness of a classification model is required to judge its performance. The Confusion Matrix(CM) shows the recall, F1_score, accuracy, and precision to judge the performance of a classification model. True positives are the values where the both the predicted and actual values are true. On the other hand, true negatives are the values where the actual value is false, and the predicted value is also false. False Positives are the values where the predicted value is false, and the actual

value is true. False negatives are the value where the actual value is true and predicted value is false.

Actual	Predicted		
		Negative	Positive
	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

[Fig. 3] CM for binary classification (2x2 matrix)

Accuracy is simply the frequency of correct predictions. It is the proportion between the number of accurate predictions and all predictions combined.

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision indicates the level of accuracy attained in real predictions. Out of all the samples that really belong to the positive class, the proportion of samples that were accurately predicted.

$$\text{Precision} = \frac{\text{Predictions Actually Positive}}{\text{Total Predicted Positive}} = \frac{TP}{TP + FP}$$

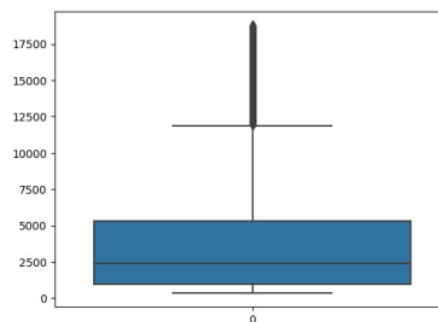
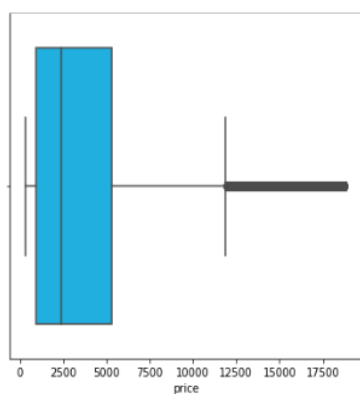
Recall measures how well actual observations match predictions. It is also referred to as sensitivity.

$$\text{Recall} = \frac{\text{Predictions Actually Positive}}{\text{Total Actual Positive}} = \frac{TP}{TP + FN}$$

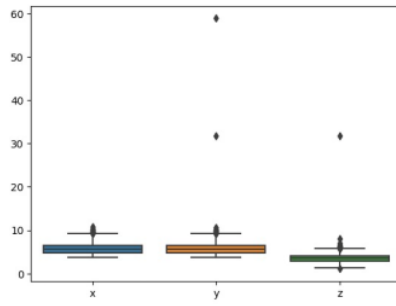
The harmonic mean(HM) of recall and precision is the F1 score. The F1 score is responsible to keep precision and recall the classifier in balance.

$$\text{F1-Score} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Data visualization :

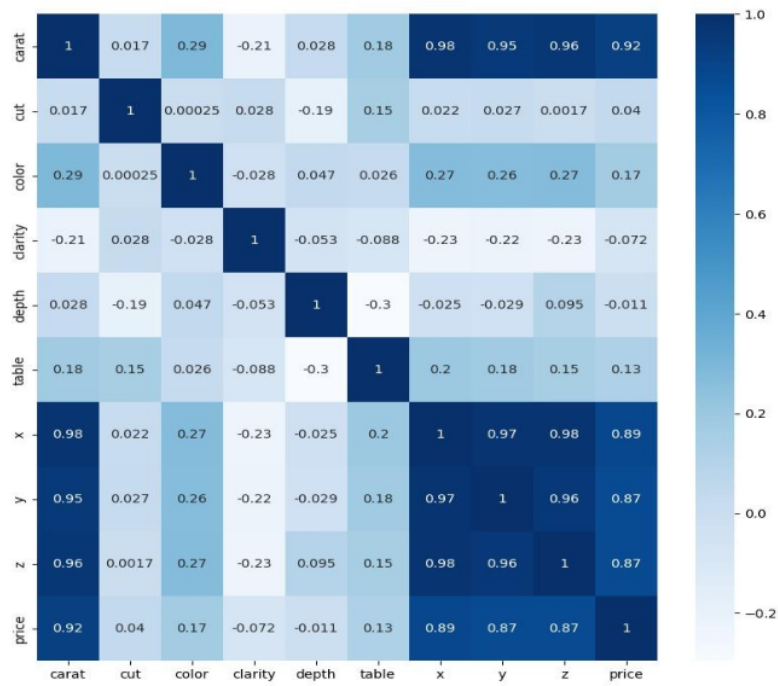


[Fig. 4] Price boxplot



[Fig. 5] Price distribution of diamonds in the dataset

[Fig. 6] Data Distribution of x,y,z



[Fig. 7] Correlation Matrix

Experimental results:

Finding from Exploratory data analysis of the dataset indicate that most of the diamonds listed are in the price range of 800 to 5000 dollars, as shown in Fig. 4, whereas the entire price class for the data set ranges from 100 to 18000 dollars. So, there is a heavy portion of the data belonging to a specific range. This finding helped identify that the binning of data based on equal price range and binning of data based on equal number of points(frequency) in a bin shall give us 2 different results. As mentioned in the flowchart (Fig. 1) above, One Hot Encoding [6] has been used on the dataset and then performed binning based on equal price range and then by equal number of data points. Similarly, the data was label encoded and then classified(binned) them by equal price ranges or again by equal number of data points. As a result, 4 different combinations of data were found that were fed into the machine learning algorithms for training. Below is the table that has best performing algorithms and their accuracy.

Random forest gives accuracy of 87.425% and XgBoost classifier gives 87.073%

Various classification algorithms:

TABLE I.

<i>Method Used</i>	Label Encoding Binning By Equal Points	Label Encoding Binning By Equal Prices	One Hot Encoding Binning By Equal Data Points	One Hot Encoding Binning By Equal Price
<i>Algorithms</i>				
<i>Gaussian NB</i>	51.05	70.92	53.54	69.96
<i>Decision Tree</i>	77.58	83.53	77.38	83.85
<i>XgBoost</i>	83.76	87.07	83.52	87.18
<i>Random Forest</i>	83.10	87.16	82.10	87.51
<i>Gradient Boosting</i>	77.18	84.03	75.69	83.73
<i>Adaboost</i>	30.12	60.57	30.12	60.57

The hyperparameter findings are not satisfactory as shown in the below table. It showed an improvement, but that improvement is not worth the time taken to train the hyperparameters.

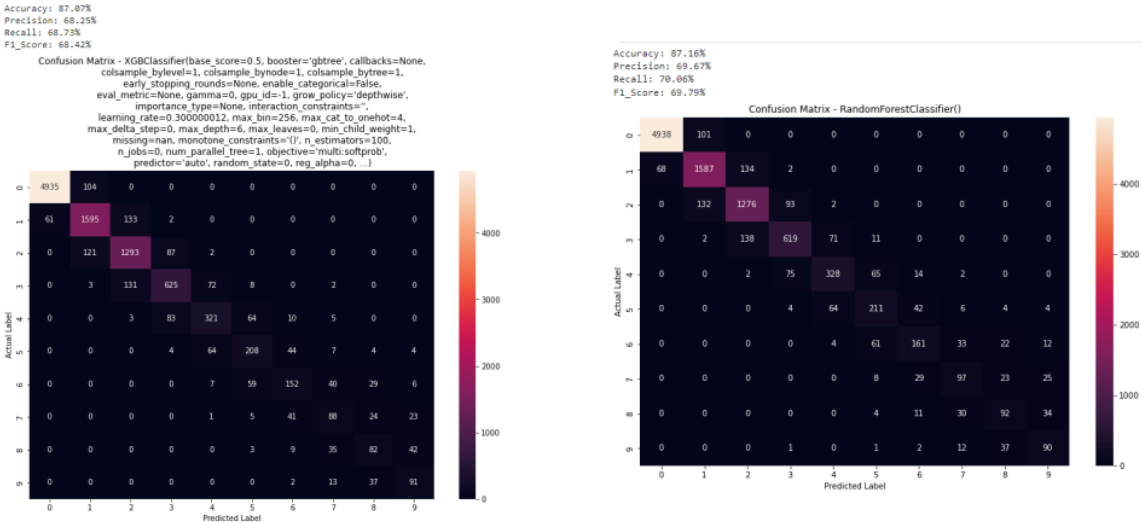
So, in the final model building, the hyperparameter tuned models shall be discarded. Reasons for these unsatisfactory results may be the varied nature of the dataset. Hyperparameter tuning works well for data which has linear correlation throughout, so in such datasets, one changed hyperparameter can work well for all tuples in that data. For the current dataset, a single parameter adjustment could not be found as it is simply not possible, i.e., there is no such set of same parameters that can accommodate for all the trend changes in the dataset. Therefore, the hyperparameter tuning extension is not worth doing.

Various classification algorithms with hyperparameter tuning:

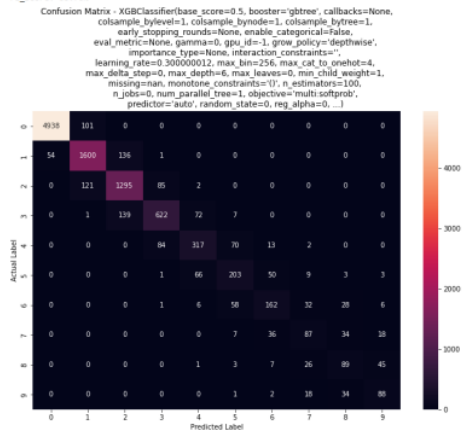
TABLE II.

Algorithms	Best Score
Decision Tree Classifier	84.618%
Random Forest Classifier	72.952%
XgBoost Classifier	46.726%

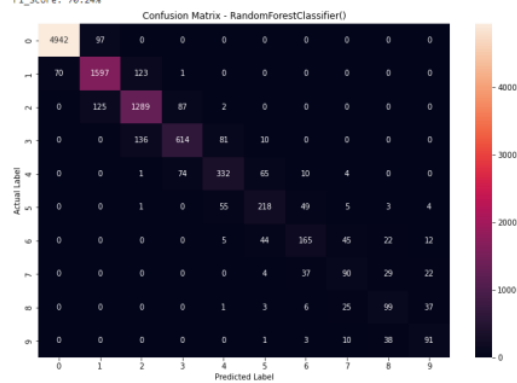
Confusion Matrix of best performing models :



Accuracy: 87.18%
Precision: 68.63%
Recall: 69.08%
F1_Score: 68.76%



Accuracy: 87.51%
Precision: 69.97%
Recall: 70.67%
F1_Score: 70.24%



Discussion and conclusion:

Random Forest Classifier is the best performing model on this dataset, with accuracy: 87.425%. As mentioned above, regression analysis could have achieved higher MSE, but in the real world, a range of price would be more beneficial to the stakeholders rather than an exact price. Tuning the hyperparameters in some models is not improving the performance as such, this is a major finding. This model now can be used in APIs to predict the price range of a newly found diamond in the current market. For future work, there are places of improvement in this paper, like in the hyperparameter tuning. The hyperparameters can be better tuned to achieve a higher result. Also, the training and testing of the dataset was done only through Hold-out cross validation method. The other types of cross validation have not been explored in this paper.

References:

- [1] Sharma, G., Tripathi, V., Mahajan, M., & Srivastava, A. K. (2021, January). Comparative analysis of supervised models for diamond price prediction. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 1019-1022). IEEE.
- [2] Alsuraihi, W., Al-hazmi, E., Bawazeer, K., & AlGhamdi, H. (2020, March). Machine learning algorithms for diamond price prediction. In Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing (pp. 150-154).
- [3] Chaijunla, T., & Taninpong, P. Comparative Study of Predicting Diamond Ring Prices in Online Retail Shop.
- [4] Basysyar, F. M., & Dwilestari, G. Comparison of machine learning algorithms for predicting diamond prices based on exploratory data analysis.
- [5] Mamonov, S., & Triantoro, T. (2018). Subjectivity of diamond prices in online retail: insights from a data mining study. *Journal of theoretical and applied electronic commerce research*, 13(2), 15-28.
- [6] Yong, Z., Jianyang, L., Hui, L., & Xuehui, G. (2018, June). Fatigue driving detection with modified ada-boost and fuzzy algorithm. In 2018 Chinese Control And Decision Conference (CCDC) (pp. 5971-5974). IEEE.
- [7] Dutta, J., Kim, Y. W., & Dominic, D. (2020, November). Comparison of gradient boosting and extreme boosting ensemble methods for webpage classification. In 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) (pp. 77-82). IEEE.
- [8] Chen, H., Ai, H., Yang, Z., Yang, W., Ye, Z., & Dong, D. (2020, September). An Improved XGBoost Model Based on Spark for Credit Card Fraud Prediction. In 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS) (pp. 1-6). IEEE.
- [9] Bulbul, H. I., & Unsal, Ö. (2011, December). Comparison of classification techniques used in machine learning as applied on vocational guidance data. In 2011 10th International Conference on Machine Learning and Applications and Workshops (Vol. 2, pp. 298-301). IEEE.
- [10] Yue, Y., & Yang, Y. (2020, February). Improved Ada Boost Classifier for Sports Scene Detection in Videos: from Data Extraction to Image Understanding. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 1-4). IEEE.

ORIGINALITY REPORT

22%

SIMILARITY INDEX

16%

INTERNET SOURCES

14%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1

datascience.cmu.ac.th

Internet Source

3%

2

ijeast.com

Internet Source

2%

3

Hongwei Chen, He Ai, Zhihui Yang, Weiwei Yang, Zhiwei Ye, Dawei Dong. "An Improved XGBoost Model Based on Spark for Credit Card Fraud Prediction", 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), 2020

Publication

1%

4

J Dutta, Yong Woon Kim, Dalia Dominic. "Comparison of Gradient Boosting and Extreme Boosting Ensemble Methods for Webpage Classification", 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2020

Publication

1%

5	www.ijana.in Internet Source	1 %
6	academic-accelerator.com Internet Source	1 %
7	www.researchgate.net Internet Source	1 %
8	www.coursehero.com Internet Source	1 %
9	digitalcommons.montclair.edu Internet Source	1 %
10	Submitted to University of Westminster Student Paper	1 %
11	Submitted to Southern New Hampshire University - Continuing Education Student Paper	1 %
12	Yibo Yue, Yucheng Yang. "Improved Ada Boost Classifier for Sports Scene Detection in Videos: from Data Extraction to Image Understanding", 2020 International Conference on Inventive Computation Technologies (ICICT), 2020 Publication	1 %
13	Submitted to Aston University Student Paper	1 %

14

Jayakanth Kunhoth, Somaya Al Maadeed, Moutaz Saleh, Younes Akbari. "Exploration and analysis of On-Surface and In-Air handwriting attributes to improve dysgraphia disorder diagnosis in children based on machine learning methods", Biomedical Signal Processing and Control, 2023

Publication

1 %

15

Garima Sharma, Vikas Tripathi, Manish Mahajan, Awadhesh Kumar Srivastava. "Comparative Analysis of Supervised Models for Diamond Price Prediction", 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021

Publication

<1 %

16

Md Shaik Amzad Basha, Peerzadah Mohammad Oveis, C Prabavathi, Macherla Bhagya Lakshmi, M Martha Sucharitha. "An Efficient Machine Learning Approach: Analysis of Supervised Machine Learning Methods to Forecast the Diamond Price", 2023 International Conference for Advancement in Technology (ICONAT), 2023

Publication

<1 %

17

Xing Wang, Chong Jiang. "Computer-Aided Physical Training Sports Injury Risk Simulation

<1 %

Based on Embedded Image System", Microprocessors and Microsystems, 2021

Publication

18

ebin.pub

Internet Source

<1 %

19

ijsrcseit.com

Internet Source

<1 %

20

researchwith.montclair.edu

Internet Source

<1 %

21

www.ijeat.org

Internet Source

<1 %

22

Md Shaik Amzad Basha, M Martha Sucharitha, A. Uma Devi, M Ashok, Peerzadah Mohammad Oveis. "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), 2023

Publication

<1 %

23

B. Y. Ong, S. W. Goh, Chi Xu. "Sparsity adjusted information gain for feature selection in sentiment analysis", 2015 IEEE International Conference on Big Data (Big Data), 2015

Publication

<1 %

24 Submitted to University College for the Creative Arts at Canterbury, Epsom, Farnham, Maidstone and Rochester
Student Paper <1 %

25 science.cmu.ac.th
Internet Source <1 %

26 www.mdpi.com
Internet Source <1 %

27 "Ubiquitous Computing and Ambient Intelligence", Springer Science and Business Media LLC, 2017
Publication <1 %

28 Waad Alsuraihi, Ekram Al-hazmi, Kholoud Bawazeer, Hanan Alghamdi. "Machine Learning Algorithms for Diamond Price Prediction", Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing, 2020
Publication <1 %

29 eprints.arums.ac.ir
Internet Source <1 %

30 cran.stat.auckland.ac.nz
Internet Source <1 %

31 www.nabet.us
Internet Source <1 %

www.sciencegate.app

32

Internet Source

<1 %

33

Amir Hussein Ataei Monfared, Amir Hossein Avizheh. "Re-rank Search Engine Results via Machine Learning", 2023 9th International Conference on Web Research (ICWR), 2023

Publication

<1 %

34

link.springer.com

Internet Source

<1 %

35

"Advanced Computational Paradigms and Hybrid Intelligent Computing", Springer Science and Business Media LLC, 2022

Publication

<1 %

Exclude quotes On
Exclude bibliography On

Exclude matches Off