DIAMOND PRICE PREDICTION

ANVAS P, B191123CE

ABSTRACT

As a type of carbon, diamond is one of the toughest and most expensive compounds produced naturally. However, unlike gold and silver, establishing the price of a diamond is extremely difficult due to the numerous factors that must be considered. I want to predict the price of a diamond based on its quality, colour, cut, depth, table, length, and other properties using a real-world dataset.

INTRODUCTION

One of the most well-known and sought-after gemstones is the diamond. Since ancient times, they have been utilised as decorative ornaments. Diamond's hardness and high light dispersion make it ideal for cutting. The diamond's unique "fire" makes it suitable for both industrial and decorative purposes. Because diamonds are one of the strongest minerals on the planet, they are utilised in a variety of machines and other forms of equipment for cutting and slicing.

Metals, glass, and other hard materials are sliced using diamonds. The majority of diamonds are employed in real-world applications such as surgery equipment, high-quality drill machines, and the aerospace and auto industries. Due to their high heat conductivity, diamonds are also employed in several high-end semiconductor products.

Even while elements such as polishing, cutting, and mining define the diamond's initial price, other characteristics are also important in determining the correct price. Clarity, carat weight, cut, breadth, length, colour, percentage of depth, and table width are some of these characteristics. The four basic qualities of a diamond, cut, clarity, carat weight, and colour, are considered the most important elements in deciding the price of a diamond, and are referred to as the 4Cs.

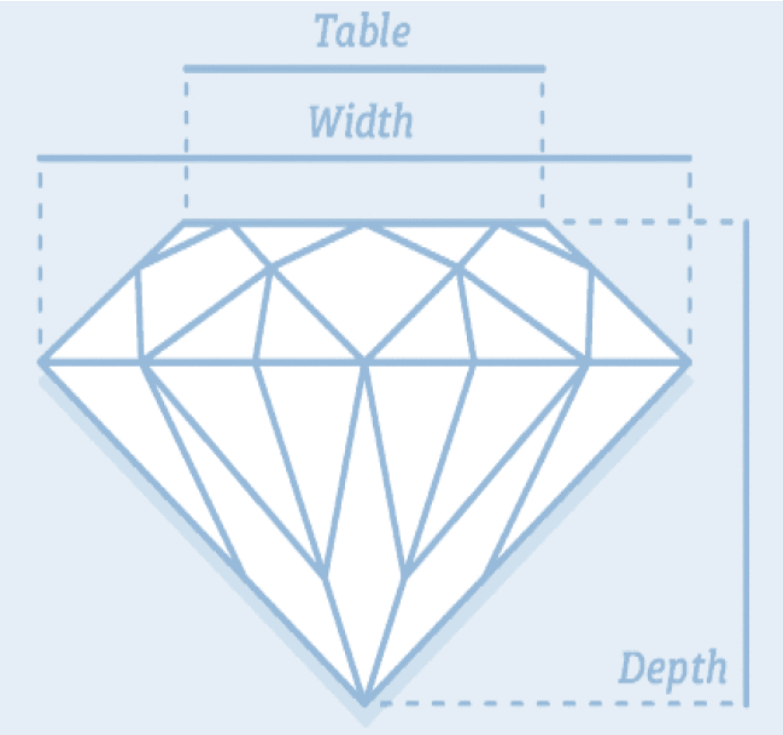
We came across a number of academics in this sector who introduced and tested a long selection of Machine Learning algorithms in order to determine the best method for training a diamond price prediction model. Finally I got a

conclusion best algorithm for finding diamond price prediction is ANN(artificial neural network)

Methods

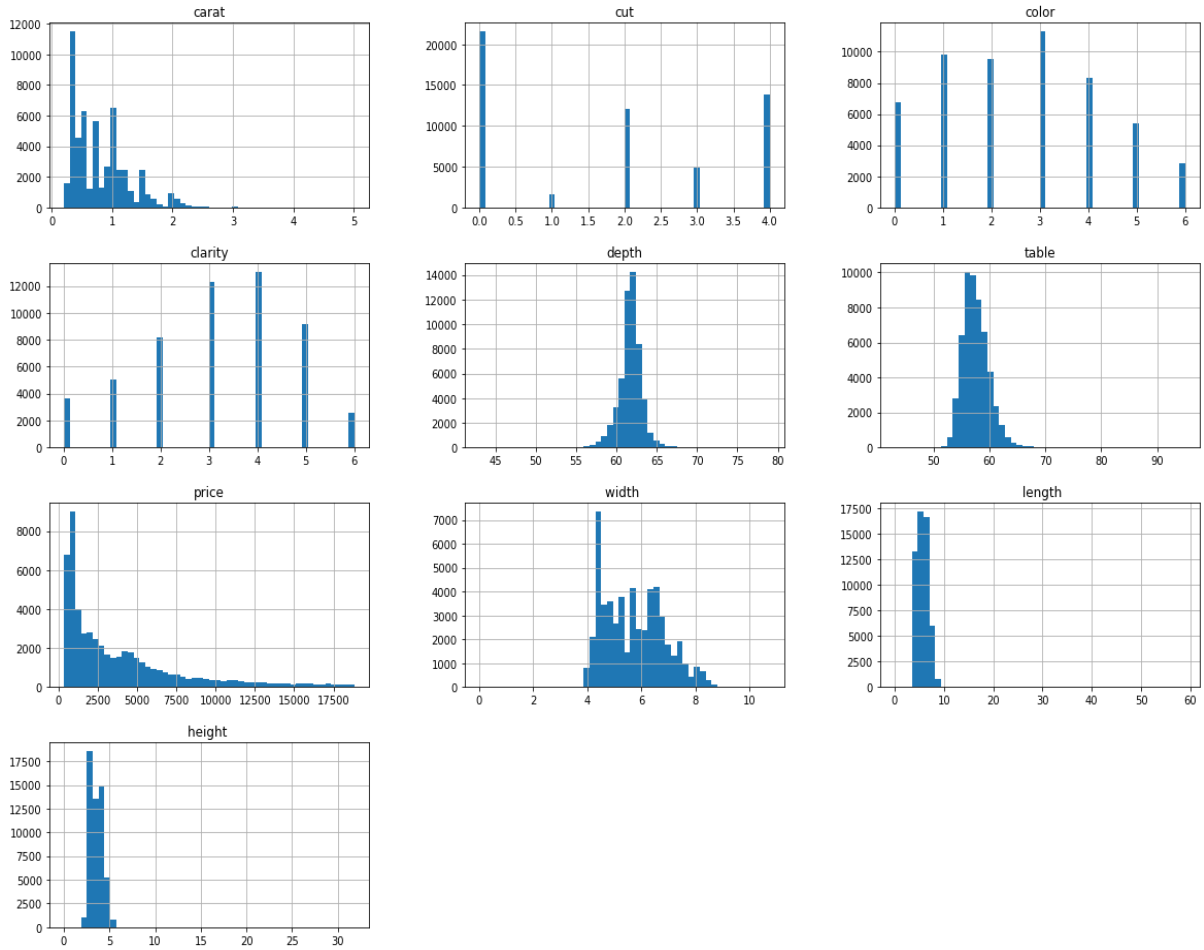
The ANN approach is a key component of this research. A neural network (artificial neuron network) is a computational model that simulates the way nerve cells in the brain work. An artificial neural network is made up of three or more interconnected layers. Input neurons make up the first layer. These neurons send input to deeper layers, which then deliver the final output data to the final output layer**.** To make it more efficient for training,

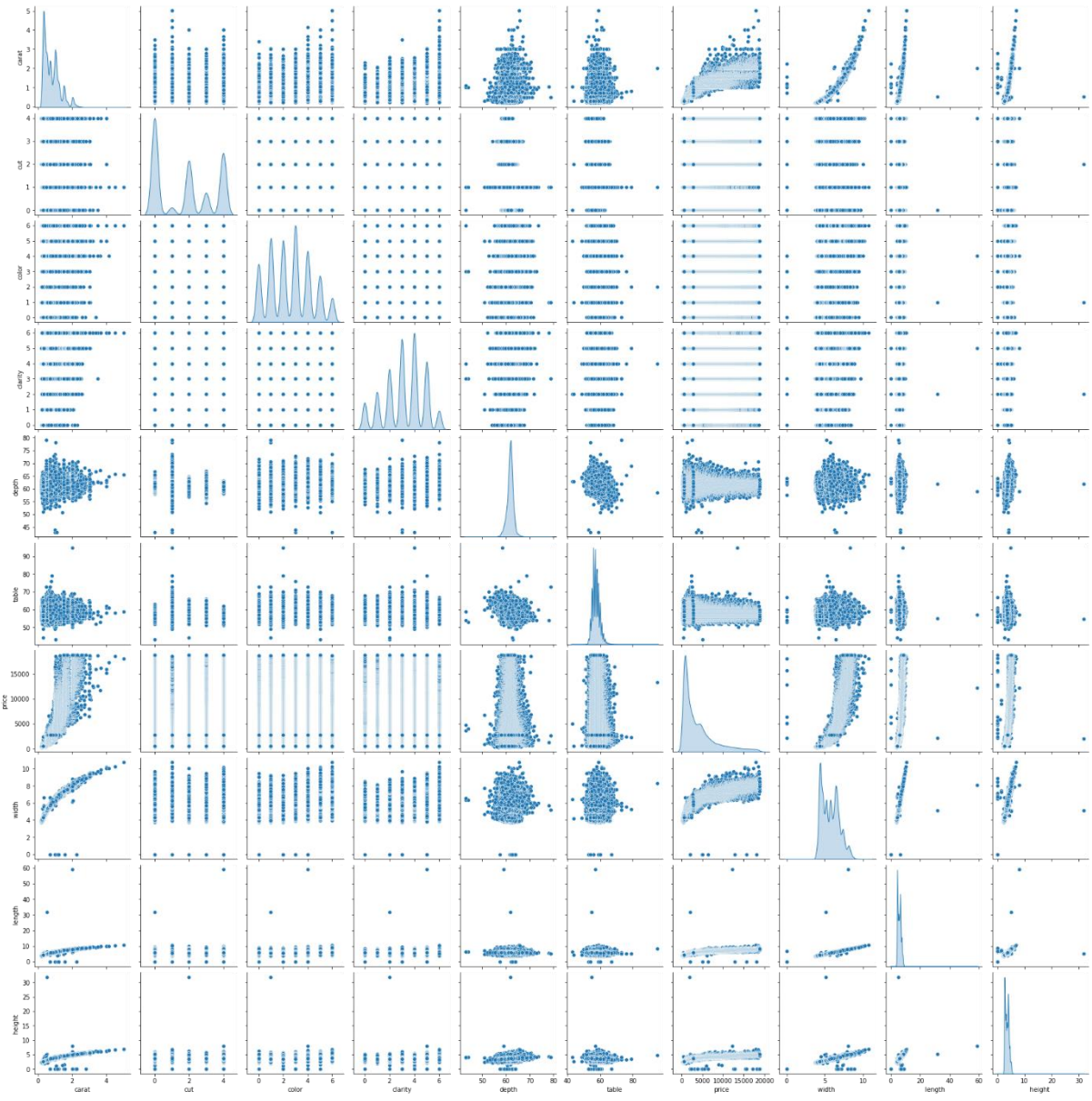
the dataset is pre-processed and optimised. The dataset is divided into two parts: training data and testing data. Training data is used to train the models, while testing data is used to test the models and determine the values of performance parameters. The performance parameter values for all of the models are acquired after training the models and using the test data. Based on these parameters, the optimal regression model is determined, and the same model is utilised for price prediction. The price forecast is made in US dollars**.**



DATA SET

There are 53,940 total unique samples in the diamond dataset. Carat, cut, colour, clarity, depth, table, price, length, height, and width are the attributes of the dataset. The characteristics of the data are critical for estimating the best-estimated diamond price. All of the diamonds in the databases have a carat weight ranging from 0.2Kg to 5.01Kg. A categorical variable with five distinct values is the cut. ideal(0), premium(1), very good(2), good(3), and fair(4) are all terms that can be used to describe a situation . The diamonds' colour varies from 0 to 6, with 6 representing the worst and 0 representing the finest. The clarity attribute is categorical, with eight different values: IF(0), VVS1(1), VVS2(2), VS1(3), VS2(4), SI1(5), SI2(6), and I1(7), with I1 being the worst and IF being the greatest. Depth, table, price, length, height, and width are all continuous properties with different integer and floating-point values. The table, width, and depth of the diamond are all measured.




carat
cut
color
clarity
depth
table
price
width
length
height

RESULTS AND discussion

METHODOLOGY

Google colab is used for obtaining,processing,training,testing and predicting The price of the diamond with the help of ANN.

The following python library files are used for this.

TENSOR FLOW

Tensorflow is an open-source software library for numerical computation. It's the most famous deep learning library in recent years. 7 A practitioner using TensorFlow can build any deep learning structure, like a simple artificial neural network.

Keras

Keras is a high-level neural networks library written in python that works as a wrapper to Tensorflow. It is used in cases where the neural network is to be quickly built and tested with minimal code lines. It contains implementations of commonly used neural network elements like layers, objectives, activation functions, optimizers, and tools to make working with images and text data easier.

PANDAS

Pandas is a open source library for data manipulation in python

MATPLOTLIB

It is a library used for creating static,animated and interactive visualisation in python

ANN network

80 % data is used for training and 20% data is used for testing the network.

X_train.shape=(43152,9)

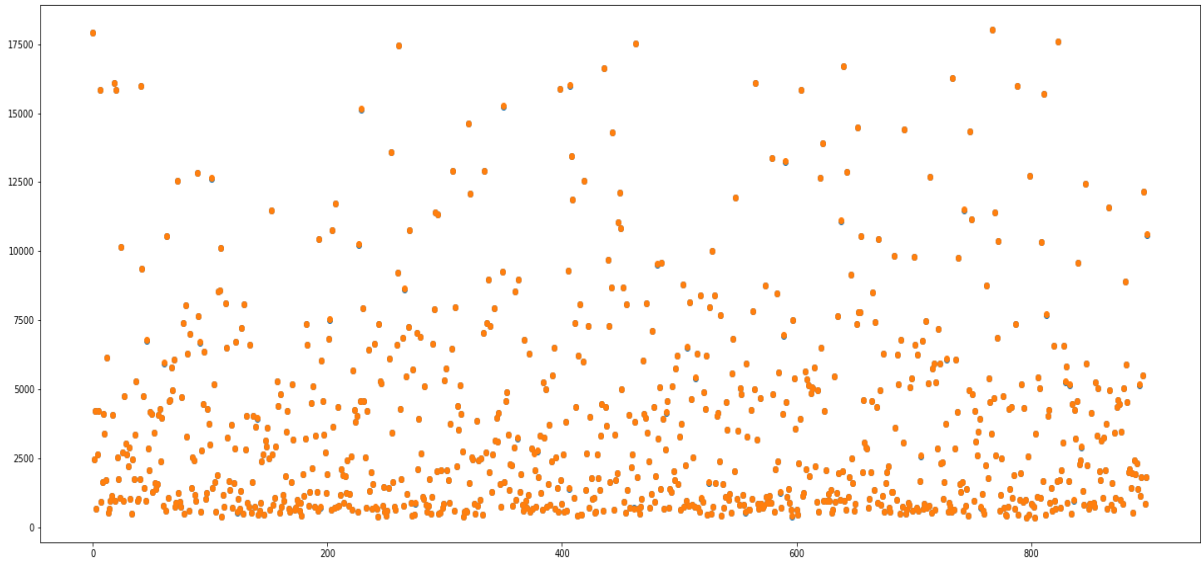
X_test.shape=(10788,9)

Ann network consisting of 4 layers 1 output layers 1 input layers 2 hidden layers. Input layers consist 9 units and activation unit is “relu”. Two hidden layers Of 32 units activation units is “relu”.output layer of unit 1 and activation unit is None.

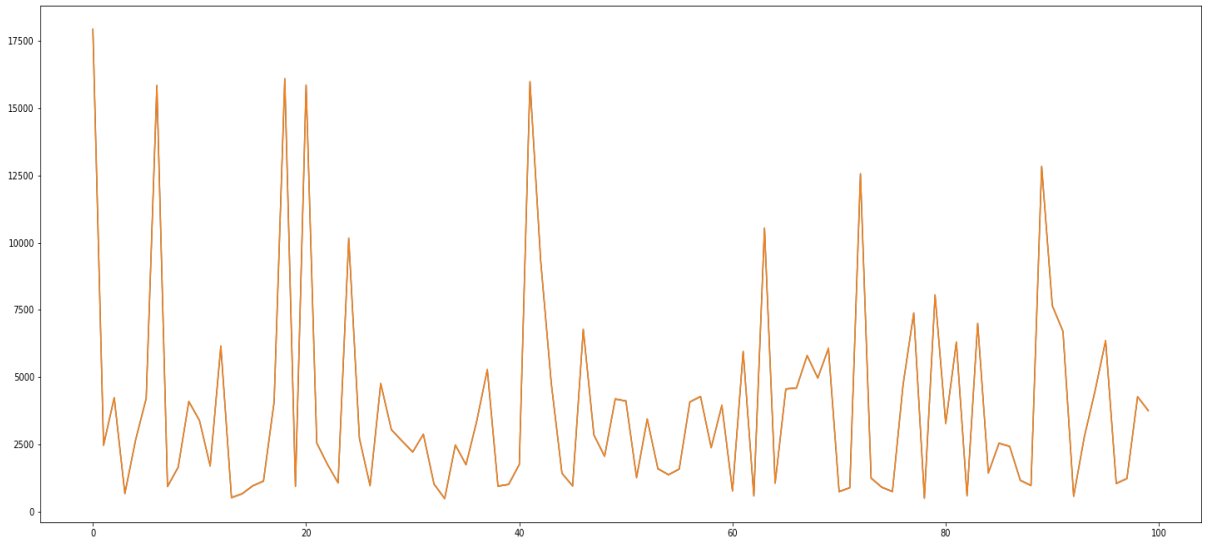
RESULTS

When we passed the test set for predicting the values using the

Artificial neural network model, we got mean absolute error =3.21



The prediction of first 1000 test cases are giving



Graph between actual price and predicted price

CONCLUSION

The model is capable to predict the values of price from different parameters of the diamond such as carat, cut, color, clarity, depth, table,width,length,

Height,. We should also introduce the number of features such as shape,polish,

Symmetry to obtain more accurate results.

REFERENCES

Data link

https://www.kaggle.com/ritikmaheshwari/diamond-price-prediction