DIAMOND PRICE PREDICTOR

Submitted to:
Dr. Sarwan Singh

Submitted by:
Komal Yadav
Rajni
Archna

INDEX

LOADING DATA

- Dataset is taken from github.com
- This dataset contains the prices and other attributes of almost 54,000 diamonds. There are 10 attributes included in the dataset including the target i.e. price.

```
# read dataset
url="https://raw.githubusercontent.com/sarwansingh/Python/master/ClassExamples/data/diamond.csv"
df1=pd.read_csv(url)
df1.head()
```

	Unnamed: 0	carat	cut	color	clarity	depth	table	price	x	y	z
0	1	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
1	2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
2	3	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
3	4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
4	5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75

➢**Feature description**:

- **price** price in US dollars ($326--$18,823)This is the target column containing tags for the features.

➢**The 4 Cs of Diamonds:-**

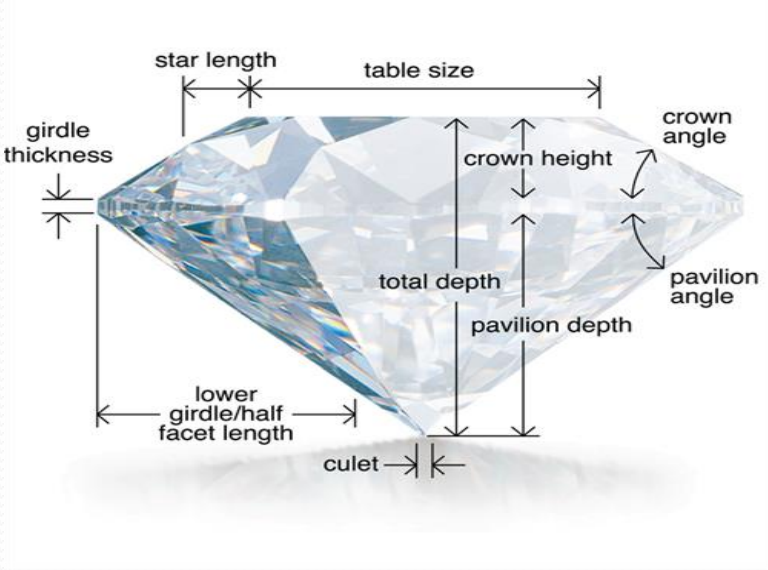
- **carat (0.2--5.01)** : Carat is the primary indicator of the size of a diamond.
- **cut (Fair, Good, Very Good, Premium, Ideal)** : Cut determines the brilliance or sparkle of the diamond.
- **color, from J (worst) to D (best)** : Color rating range from colorless to slightly tinted.
- **clarity (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best)):** The amount of inclusions or flaws in the diamond. Can range from flawless to included.

Dimensions:-

- x length in mm (0--10.74)
- y width in mm (0--58.9)
- z depth in mm (0--31.8)
- **depth total depth percentage = z /mean(x,y) = 2 * z / (x + y) (43--79)** The depth of the diamond is its height (in millimetres) measured from the culet (bottom tip) to the table (flat, top surface).
- **table width of the top of the diamond relative to widest point (43--95)**

A diamond's table refers to the flat facet of the diamond seen when the stone is face up.

Dimensional Proportion of Diamond



DATA PREPROCESSING

Steps involved in Data Pre-processing

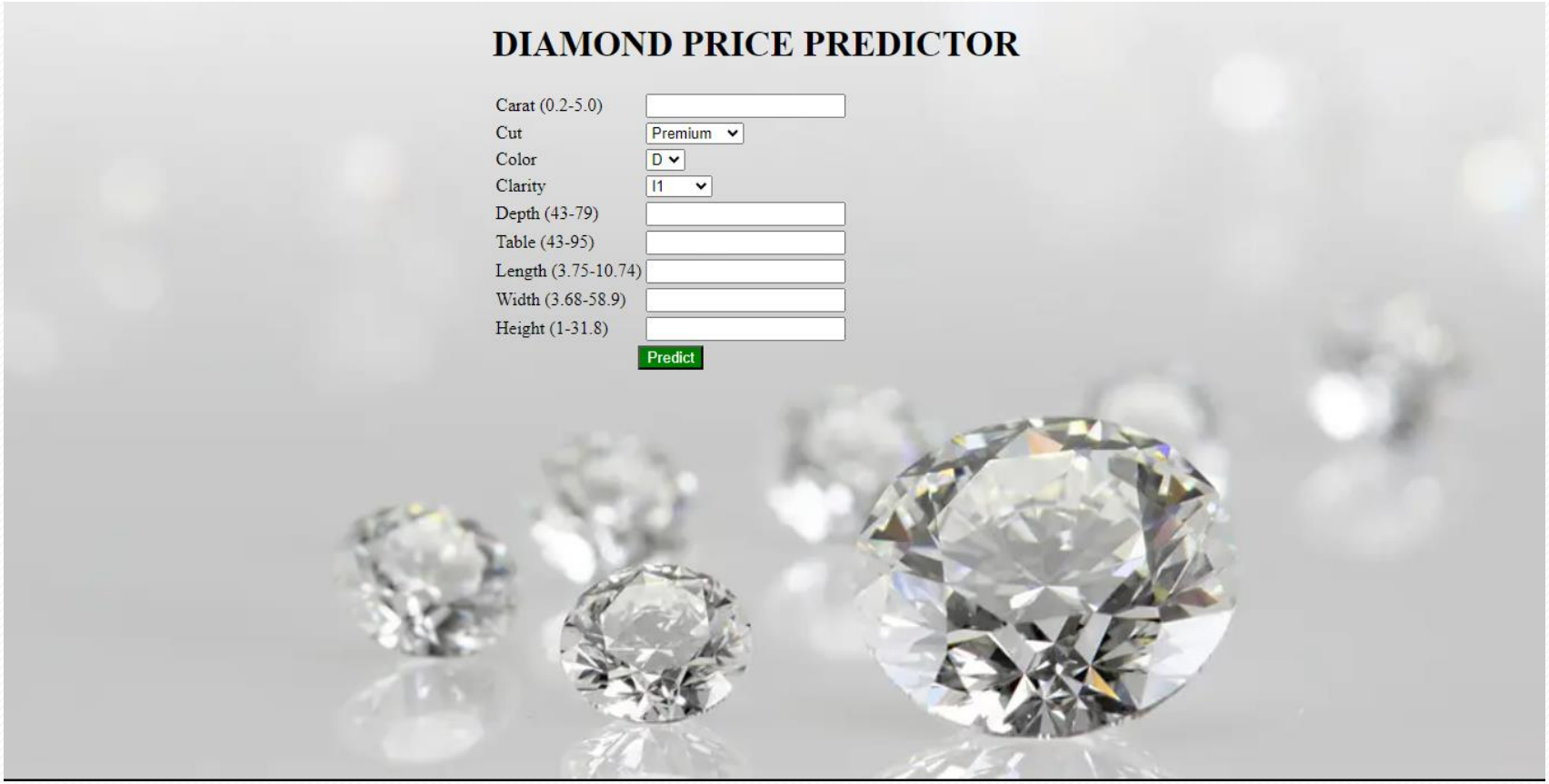
- Data cleaning
- Identifying and removing outliers
- Encoding categorical variables

ALGORITHM USED

- MODEL-Random Forest classifier

Model Deploy	Negative root mean error
Random Forest Classifier	-543.47
K-neighbors Classifier	-722.92
Decision Tree	-756.89
Linear Regression	-1215.30
XGB Regressor	-544.7

DEPLOYMENT


DIAMOND PRICE PREDICTOR
Carat (0.2-5.0)
Cut
Premium
Color
D
Clarity
I1
Depth (43-79)
Table (43-95)
Length (3.75-10.74)
Width (3.68-58.9)
Height (1-31.8)
Predict

- THANK YOU -