CENG 3522 Applied Machine Learning
Final Project Description

Team Members

Mehmet Kadri GOFRALILAR
İrem KIZILIRMAK
Mehmet Enes KIZILKAYA

Project Description (Alternative 2 - Data to problem)

This project is based on implementing and comparing various machine learning algorithms that will be used for predicting a numerical feature, “price”, in a dataset downloaded from Kaggle. In this process, research on features of the data, exploratory data analysis (EDA), data visualization, imputation (if necessary) and regression will be used. The data will be understood as best as possible. If the results are insufficient, discretization and classification could be tried as well.

~ Problem

Addressing whether the prediction of the stone price based on given attributes that contain information (size, color, clarity etc.) accurately is possible or not. Meanwhile, deciding on the best machine learning algorithm for this purpose on this dataset.

~ ML algorithms

Various machine learning algorithms that are developed for regression (Linear Regression, k-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine, Lasso Regression) will be implemented and compared.

~ Dataset (Gemstone dataset)

~~ Dictionary

- Unnamed ~ Index
- Carat ~ Carat weight of the cubic zirconia.
- Cut ~ Describe the cut quality of the cubic zirconia. Quality is increasing order Fair, Good, Very Good, Premium, Ideal.
- Colour ~ Colour of the cubic zirconia.With D being the best and J the worst.
- Clarity ~ Cubic zirconia Clarity refers to the absence of the Inclusions and Blemishes. (In order from Best to Worst, FL = flawless, I3= level 3 inclusions) FL, IF, VVS1, VVS2, VS1, VS2, SI1, SI2, I1, I2, I3
- Depth ~ The Height of a cubic zirconia, measured from the Culet to the table, divided by its average Girdle Diameter.
- Table ~ The Width of the cubic zirconia's Table expressed as a Percentage of its Average Diameter.
- Price ~ The price of the “cubic zirconia”.
- X ~ Length of the cubic zirconia in mm.
- Y ~ Width of the cubic zirconia in mm.
- Z ~ Height of the cubic zirconia in mm.

```
gem_data.head()
```
✓ 0.0s

	Unnamed: 0	carat	cut	color	clarity	depth	table	x	y	z	price
0	1	0.30	Ideal	E	SI1	62.1	58.0	4.27	4.29	2.66	499
1	2	0.33	Premium	G	IF	60.8	58.0	4.42	4.46	2.70	984
2	3	0.90	Very Good	E	VVS2	62.2	60.0	6.04	6.12	3.78	6289
3	4	0.42	Ideal	F	VS1	61.6	56.0	4.82	4.80	2.96	1082
4	5	0.31	Ideal	F	VVS1	60.4	59.0	4.35	4.43	2.65	779

```
gem_data.describe()
```
✓ 0.0s

	Unnamed: 0	carat	depth	table	x	y	z	price
count	26967.000000	26967.000000	26270.000000	26967.000000	26967.000000	26967.000000	26967.000000	26967.000000
mean	13484.000000	0.798375	61.745147	57.456080	5.729854	5.733569	3.538057	3939.518115
std	7784.846691	0.477745	1.412860	2.232068	1.128516	1.166058	0.720624	4024.864666
min	1.000000	0.200000	50.800000	49.000000	0.000000	0.000000	0.000000	326.000000
25%	6742.500000	0.400000	61.000000	56.000000	4.710000	4.710000	2.900000	945.000000
50%	13484.000000	0.700000	61.800000	57.000000	5.690000	5.710000	3.520000	2375.000000
75%	20225.500000	1.050000	62.500000	59.000000	6.550000	6.540000	4.040000	5360.000000
max	26967.000000	4.500000	73.600000	79.000000	10.230000	58.900000	31.800000	18818.000000