



PLOTS, SOM'S,
KNN'S AND
**CONFUSION
MATRIX**

RESULTS AND
ANALYSIS

OBSERVATIONS

TEAM BRAVO

Raju Romeo Gomes (2302646)
Shyam Gund (2218417)
Titilope Madeenah Seriki (2130844)
Walter Andrés Paz Callizo (2239884)

DIAMONDS DATASET

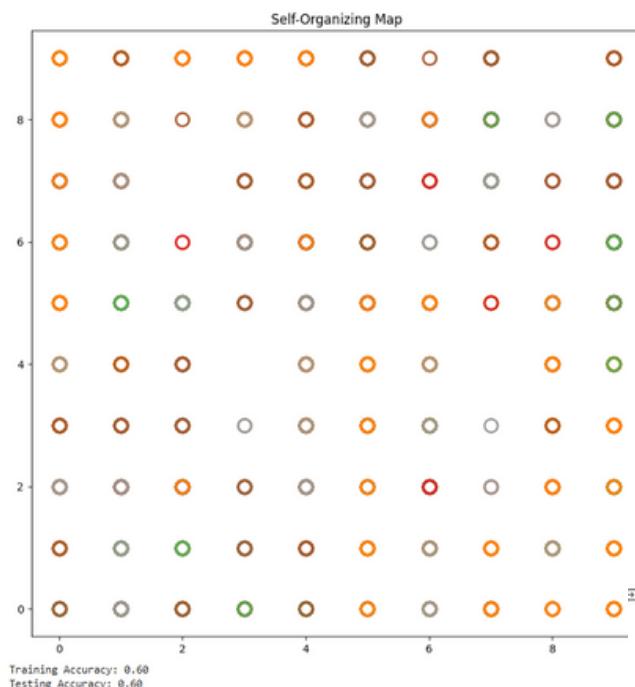
SELF-ORGANIZING MAPS

WHAT IS AN SOM?

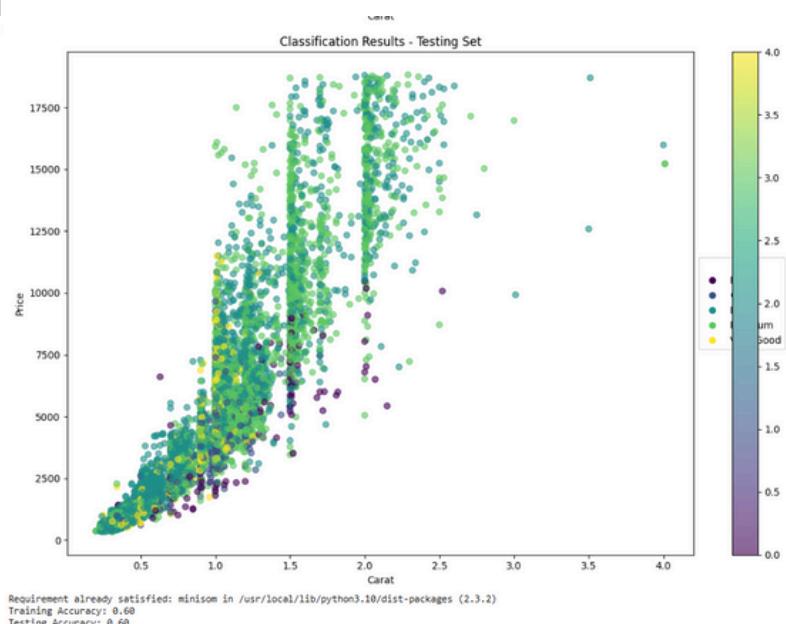


- KOHONEN NETWORKS
- PRIMARY FUNCTION IS TO CONVERT HIGH NON-FEATURES INPUT DATA TO A MAP AND CHART, WHICH RETAINS THE TOPOLOGICAL SIMILARITY OF THE DATA POINTS

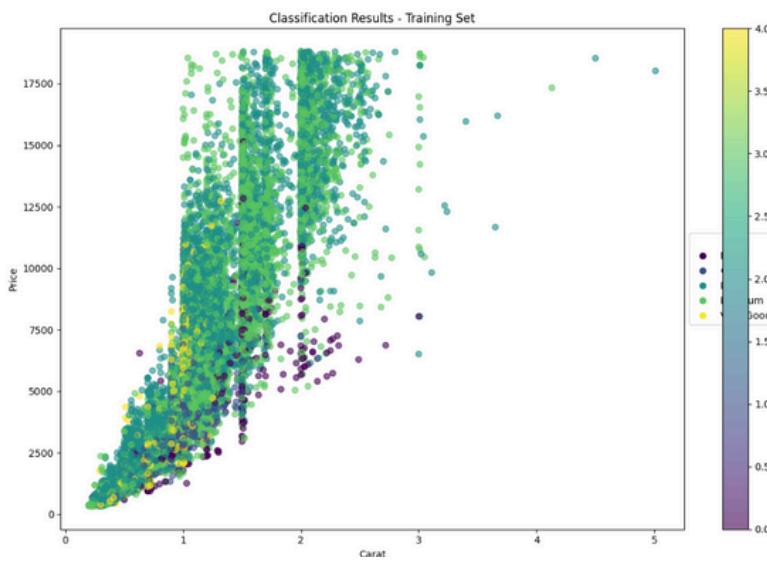
PLOTS, SOM'S, KNN'S AND CONFUSION MATRIX



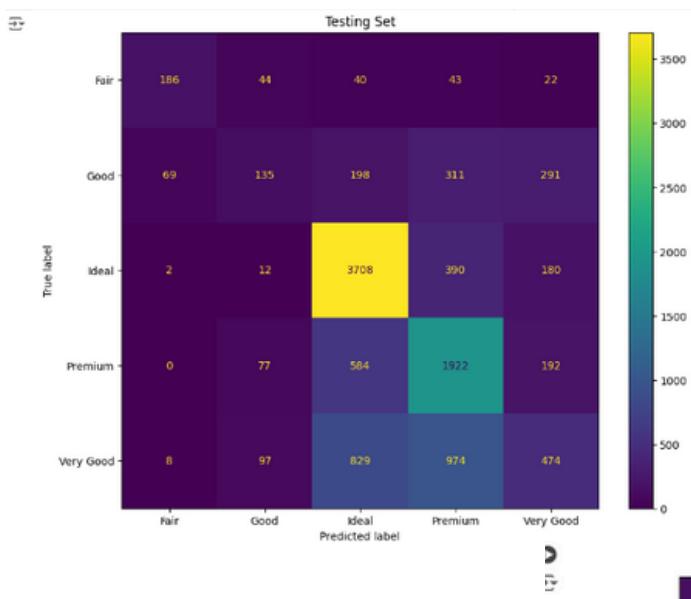
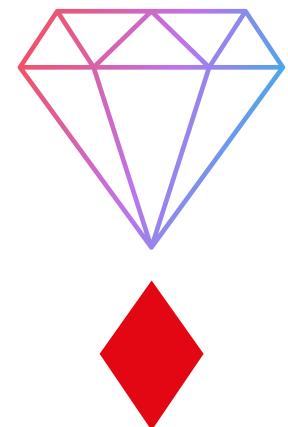
	precision	recall	f1-score	support
Fair	0.70	0.56	0.62	335
Good	0.37	0.13	0.20	1004
Ideal	0.69	0.86	0.77	4292
Premium	0.53	0.69	0.60	2775
Very Good	0.41	0.20	0.27	2382
accuracy			0.60	10788
macro avg	0.54	0.49	0.49	10788
weighted avg	0.56	0.60	0.56	10788



STUDIES AND OPINIONS

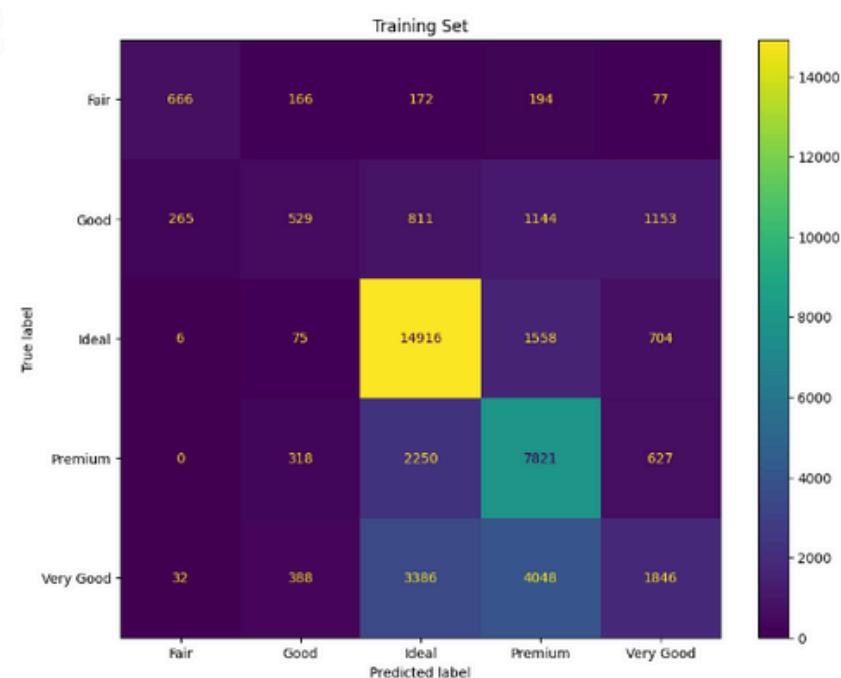


Shyam Gund (2218417)



Confusion Matrix (Testing Set):

```
[[ 186  44  40  43  22]
 [ 69 135 198 311 291]
 [  2  12 3708 390 180]
 [  0  77 584 1922 192]
 [  8  97 829 974 474]]
```

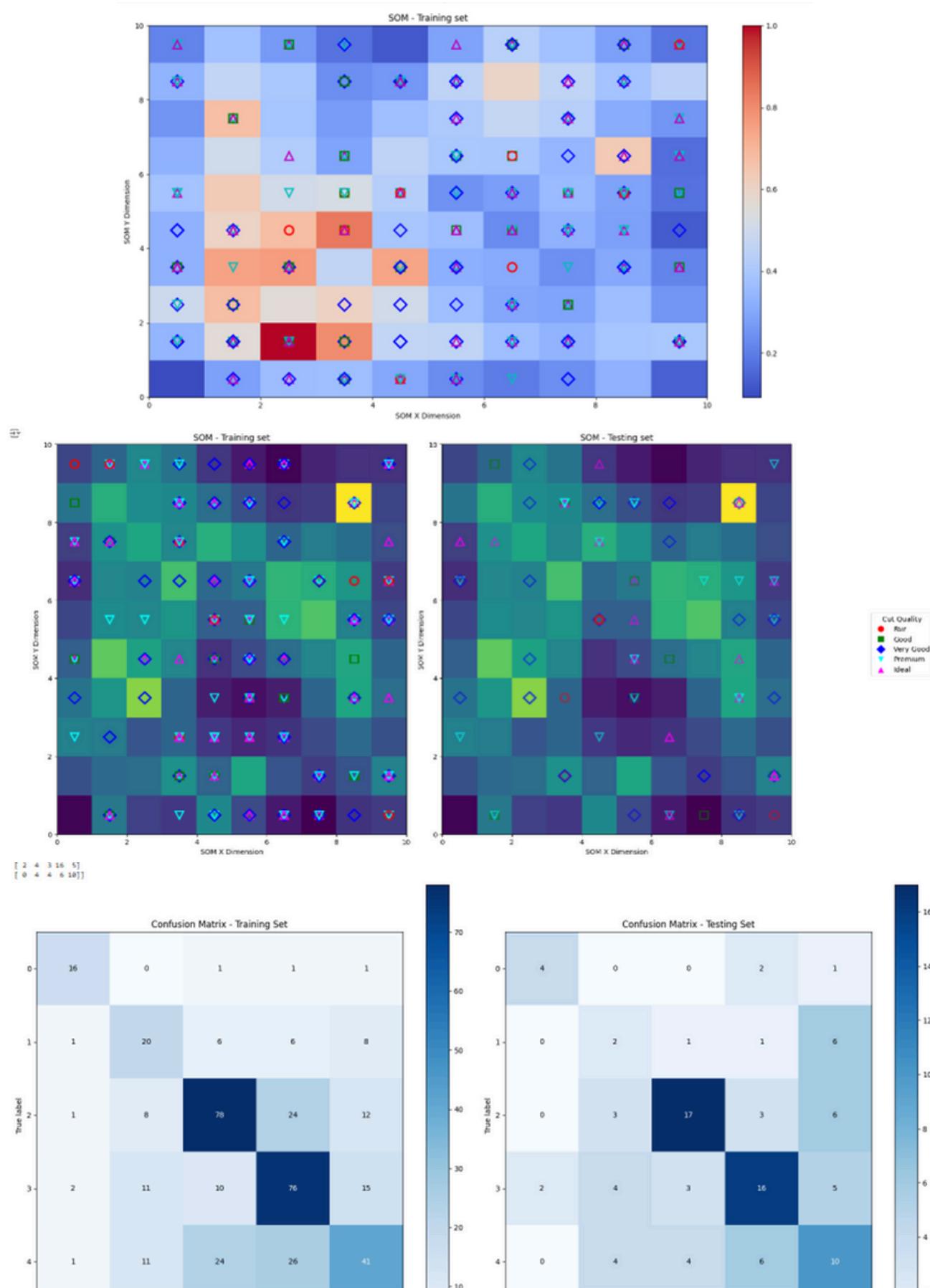


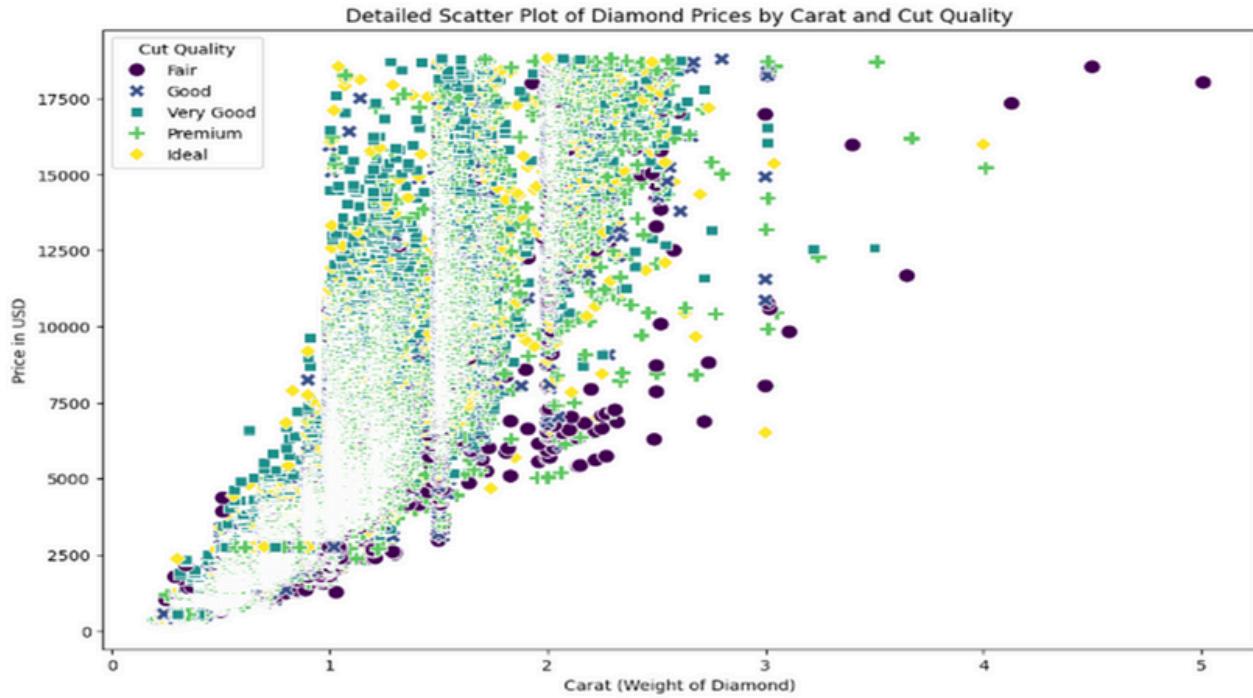
Confusion Matrix (Training Set):

```
[[ 666 166 172 194 77]
 [ 265 529 811 1144 1153]
 [  6  75 14916 1558 704]
 [  0  318 2250 7821 627]
 [ 32 388 3386 4048 1846]]
```



STUDIES AND OPINIONS





SOMS INTERPRETATION BY AI, AS WELL AS THE KNOWLEDGE FROM ACADEMIC SOURCES, IMPLIES THAT SOMS ARE VERY SIGNIFICANT IN THE FIELD OF DATA VISUALIZATION AND DIMENSIONALITY.



STUDIES AND OPINIONS

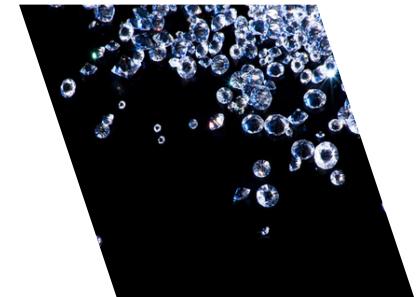
Titilope Madeenah Seriki
(2130844)

Training Set:

Accuracy: 0.6625

Classification Report:

	precision	recall	f1-score	support
Fair	0.74	0.89	0.81	19
Good	0.67	0.29	0.41	41
Ideal	0.80	0.80	0.80	123
Premium	0.63	0.69	0.66	114
Very Good	0.52	0.56	0.54	103
accuracy			0.66	400
macro avg	0.67	0.65	0.64	400
weighted avg	0.67	0.66	0.66	400

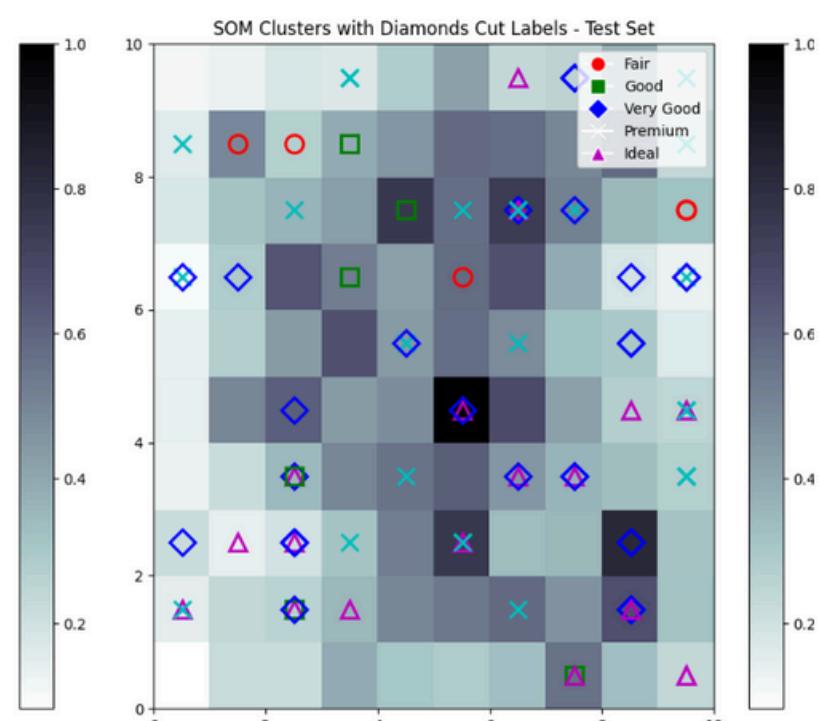
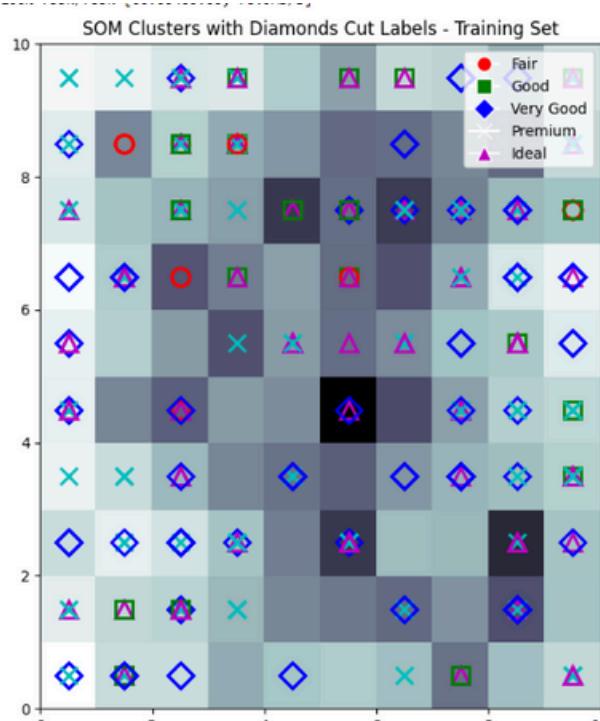


Test Set:

Accuracy: 0.72

Classification Report:

	precision	recall	f1-score	support
Fair	0.70	1.00	0.82	7
Good	0.67	0.60	0.63	10
Ideal	0.86	0.66	0.75	29
Premium	0.74	0.77	0.75	30
Very Good	0.61	0.71	0.65	24
accuracy			0.72	100
macro avg	0.72	0.75	0.72	100
weighted avg	0.73	0.72	0.72	100



STUDIES AND OPINIONS

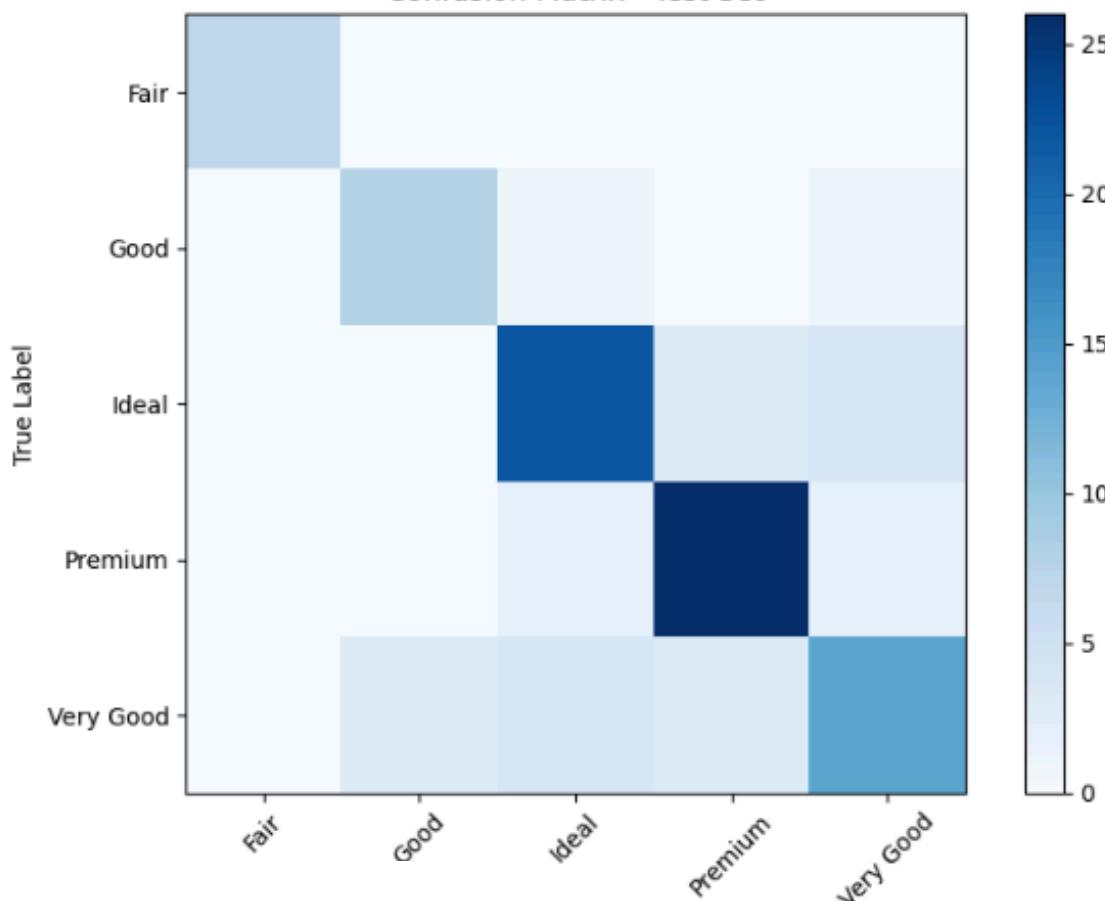
```

. Confusion Matrix (Test Set):
[[ 7  0  0  0  0]
 [ 0  8  1  0  1]
 [ 0  0 22  3  4]
 [ 0  0  2 26  2]
 [ 0  3  4  3 14]]

```

**Titilope Madeenah Seriki
(2130844)**

Confusion Matrix - Test Set



Training Set:

Accuracy: 0.67

Classification Report:

	precision	recall	f1-score	support
Fair	0.93	0.74	0.82	19
Good	0.60	0.59	0.59	41
Ideal	0.77	0.80	0.79	123
Premium	0.63	0.63	0.63	114
Very Good	0.58	0.57	0.58	103
accuracy			0.67	400
macro avg	0.70	0.67	0.68	400
weighted avg	0.67	0.67	0.67	400

Test Set:

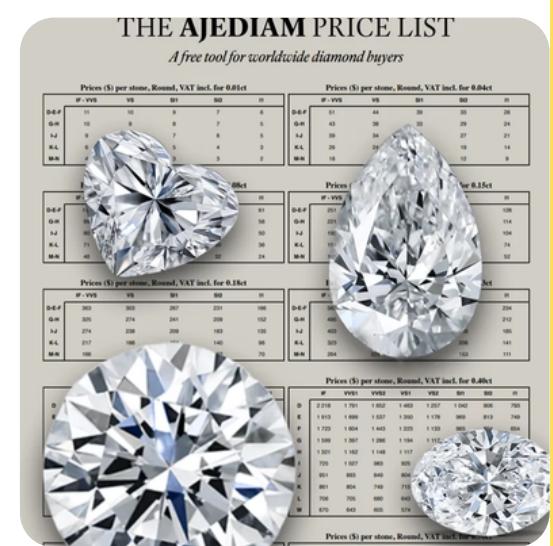
Accuracy: 0.77

Classification Report:

	precision	recall	f1-score	support
Fair	1.00	1.00	1.00	7
Good	0.73	0.80	0.76	10
Ideal	0.76	0.76	0.76	29
Premium	0.81	0.87	0.84	30
Very Good	0.67	0.58	0.62	24
accuracy			0.77	100
macro avg	0.79	0.80	0.80	100
weighted avg	0.77	0.77	0.77	100

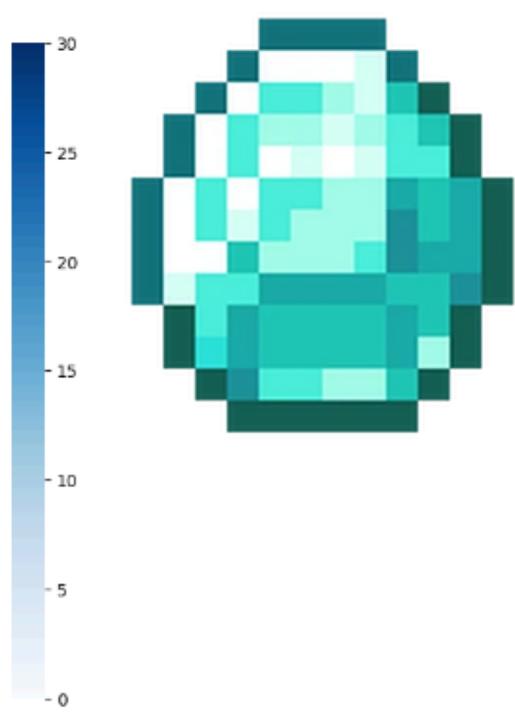
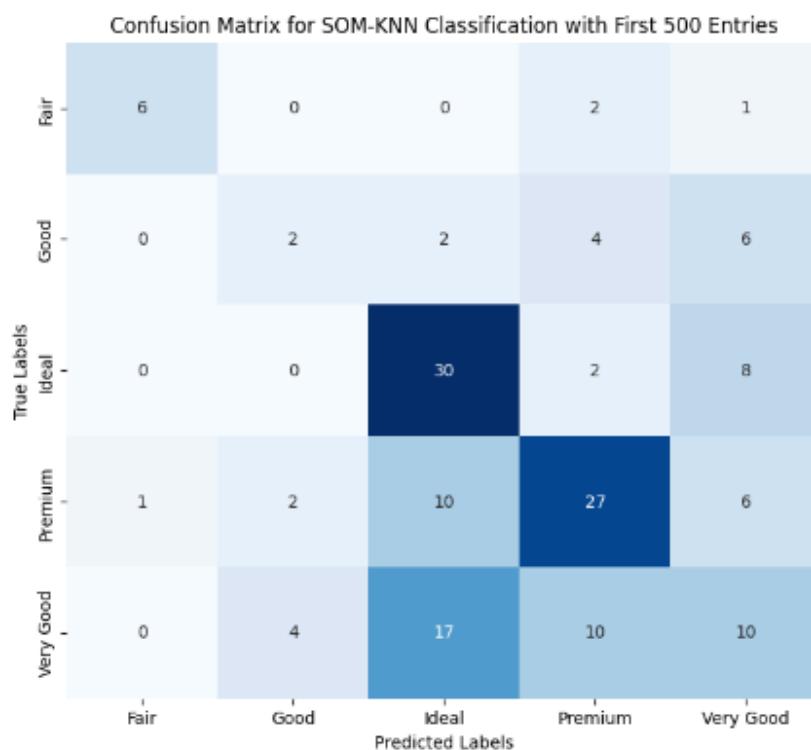
Confusion Matrix (Test Set):

```
[[ 7  0  0  0  0]]
```



STUDIES AND OPINIONS

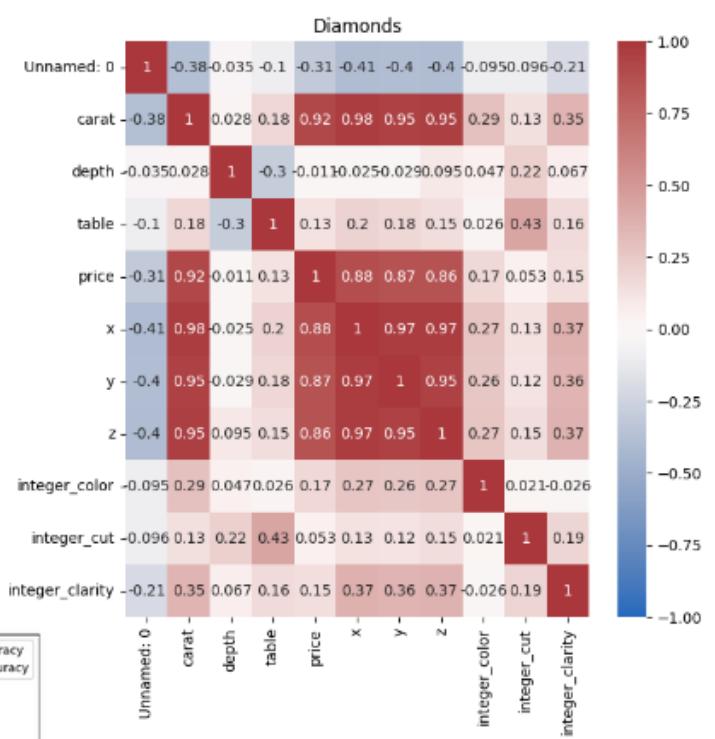
[4]



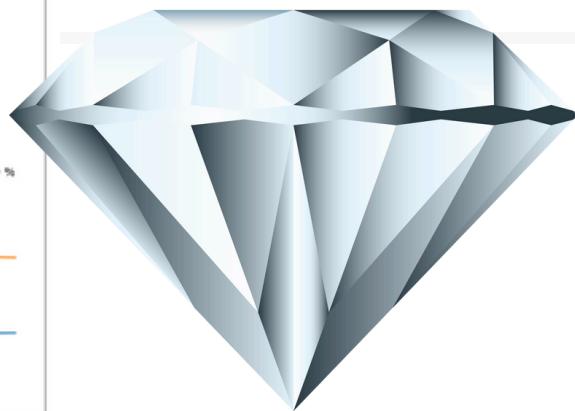
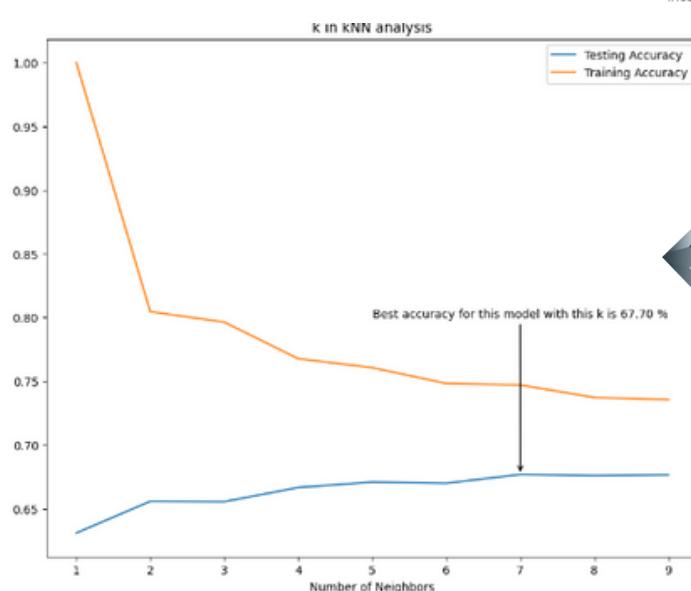
Classification Report:

	precision	recall	f1-score	support
0	0.86	0.67	0.75	9
1	0.25	0.14	0.18	14
2	0.51	0.75	0.61	48
3	0.60	0.59	0.59	46
4	0.32	0.24	0.28	41
accuracy			0.50	150
macro avg	0.51	0.48	0.48	150
weighted avg	0.48	0.50	0.48	150

Accuracy: 0.5



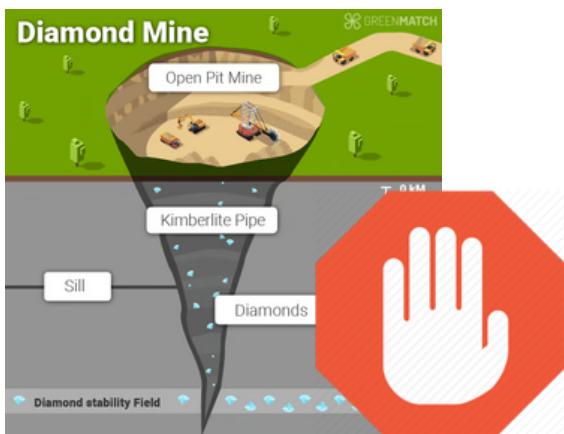
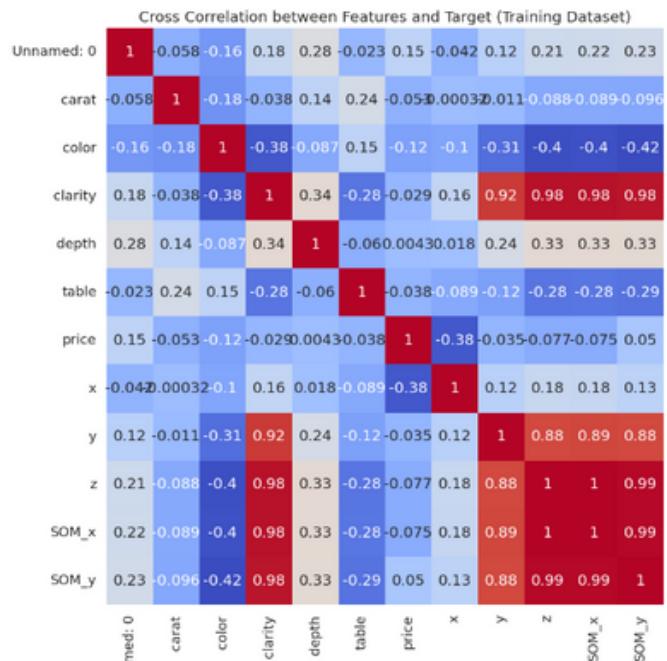
Walter Paz (2239884)



STUDIES AND OPINIONS

Index	carat	cut	color	clarity	depth	table	price	x	y	z
91	0.96	Fair	E	SI2	55.1	69.0	2757	6.45	6.33	3.52
204	0.98	Fair	H	SI2	67.9	60.0	2777	6.05	5.97	4.08
227	0.84	Fair	G	SI1	55.1	67.0	2792	6.39	6.20	3.47
296	0.70	Fair	G	VVS1	58.8	66.0	2797	5.81	5.90	3.44
298	1.01	Fair	E	SI2	67.4	60.0	2797	6.19	6.05	4.13
314	0.76	Fair	G	VS1	59.0	70.0	2800	5.99	5.90	3.46
352	0.90	Fair	I	SI1	67.3	59.0	2804	5.93	5.84	3.96
356	0.57	Fair	E	VVS1	58.7	66.0	2805	5.34	5.43	3.16
359	0.74	Fair	F	VS2	61.1	68.0	2805	5.82	5.75	3.53
385	0.99	Fair	I	SI2	68.1	56.0	2811	6.21	6.06	4.19
423	0.99	Fair	J	SI1	55.0	61.0	2812	6.72	6.67	3.68
440	0.95	Fair	F	SI2	56.0	60.0	2815	6.62	6.53	3.68
443	0.96	Fair	E	SI2	53.1	63.0	2815	6.73	6.65	3.55

OUTLIERS



Requirement already satisfied: minisom in /use/local/lib/python3.10/dist-packages (2.3.2)

Unique values for 'clarity': ['SI2', 'SI1', 'VS1', 'VS2', 'VVS1', 'I1', 'IF']

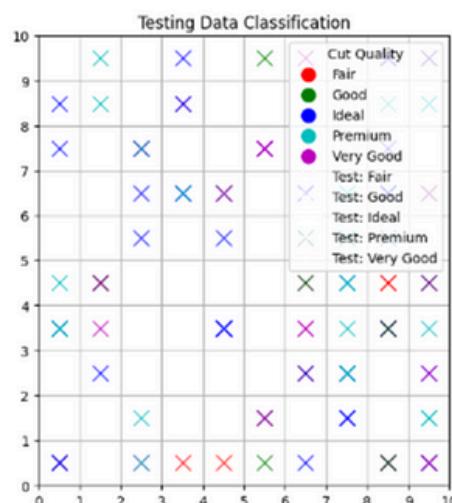
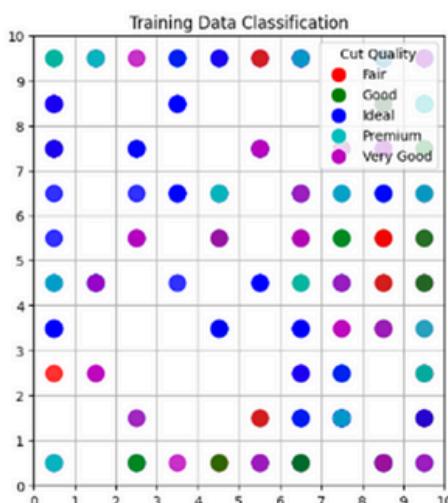
Price range: 326 - 2822

Classification Report:

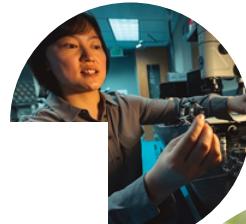
	precision	recall	f1-score	support
0	0.33	0.56	0.42	9
1	0.23	0.21	0.22	14
2	0.58	0.42	0.46	48
3	0.53	0.52	0.53	46
4	0.37	0.39	0.38	41
accuracy			0.43	150
macro avg	0.39	0.42	0.40	150
weighted avg	0.44	0.43	0.43	150

Accuracy: 0.4333333333333333

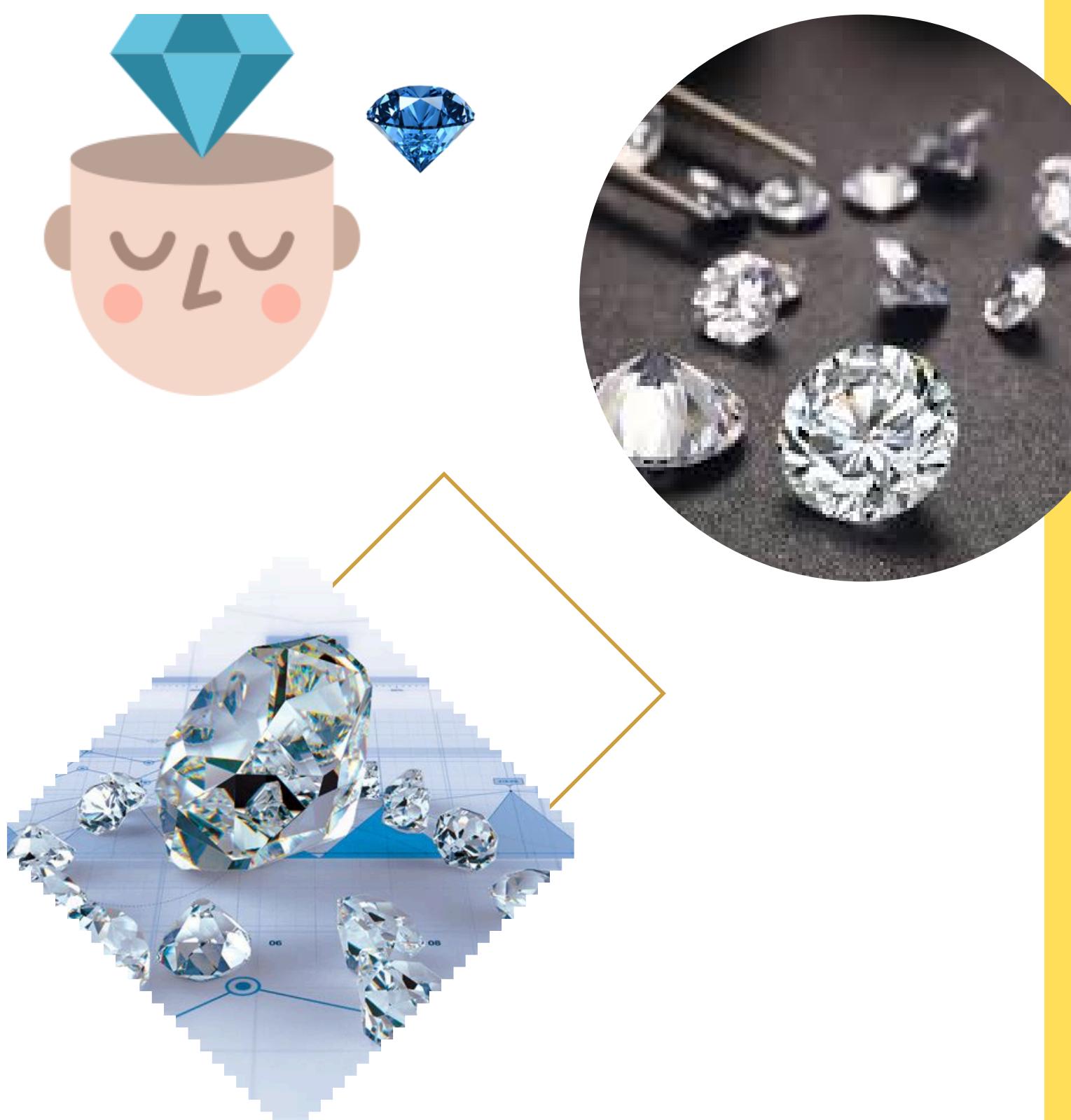
SOM Classification Results



Walter Paz (2239884)



FINAL THOUGHTS?



Raju Romeo Gomes (2302646)
Shyam Gund (2218417)
Titilope Madeenah Seriki (2130844)
Walter Andrés Paz Callizo (2239884)

REFERENCES

Christopher, V. (2020, August 8). How does self-organizing algorithm work? Analytics Vidhya. <https://medium.com/analytics-vidhya/how-does-self-organizing-algorithm-works-f0664af9bf04>

Gemini. (2024). <https://gemini.google.com/>

Google Colab. (2024). <https://colab.research.google.com/>

Kohonen, T. (2013). Essentials of the self-organizing map. Neural networks, 37, 52-65.

<https://www.sciencedirect.com/science/article/pii/S0893608012002596>

OpenAI. (2024). ChatGPT. <https://chatgpt.com/>

University of California, Berkeley. (2018). Berkeley.edu. <https://www.berkeley.edu/>

Manning College of Information & Computer Sciences. (n.d.). Manning College of Information & Computer Sciences. <https://www.cics.umass.edu/>

Massachusetts Institute of Technology. (n.d.). The Massachusetts Institute of Technology (MIT). Massachusetts Institute of Technology. <https://www.mit.edu/>

Stanford University. (2019, February 14). Stanford University. Stanford University; Stanford University. <https://www.stanford.edu/>

