This python notebook was downloaded from:

nanoHub

Using neural networks to predict and classify crystal structures of elements

Outline:

- 1. Getting a dataset
- 2. Processing and Organizing Data
- 3. Creating the Model
- 4. Plotting

1. Getting a dataset

Datasets containing properties for the elements in the periodic table are available online; however, it would be thematic to create our own, using the tools from the first tutorial on MSEML Query_Viz. In this section we will query both Pymatgen and Mendeleev to get a complete set of properties per element. We will use this data to create the cases from which the model will train and test.

In this first snippet of code we will import all relevant libraries, the elements that will be turned into cases and the properties that will serve as the attributes for the cases. We will get 47 entries (which is a small dataset), but should give us a somewhat accurate prediction. It is important to note that more entries would move the prediction closer to the real value, and so would more attributes.

The elements listed were chosen because querying them for these properties yields a dataset with no unknown values, and because they represent the three most common crystallographic structures.

```
In [ ]: import tensorflow as tf
    from tensorflow import keras
    from keras import initializers
    from keras.layers import Dense
```

```
from keras.models import Sequential
import pymatgen.core as pymat
import mendeleev as mendel
import pandas as pd
import numpy as np
import random
%matplotlib inline
import matplotlib.pyplot as plt
import sys
fcc elements = ["Ag", "Al", "Au", "Cu", "Ir", "Ni", "Pb", "Pd", "Pt", "Rh", "Th", "Yb"]
bcc_elements = ["Ba", "Ca", "Cr", "Cs", "Eu", "Fe", "Li", "Mn", "Mo", "Na", "Nb", "Rb", "Ta", "V", "W" ]
hcp_elements = ["Be", "Cd", "Co", "Dy", "Er", "Gd", "Hf", "Ho", "Lu", "Mg", "Re",
                "Ru", "Sc", "Tb", "Ti", "Tl", "Tm", "Y", "Zn", "Zr"]
elements = fcc elements + bcc elements + hcp elements
random.Random(1).shuffle(elements)
#Note that different properties have been added from the two databases.
querable mendeleev = ["atomic number", "atomic volume", "boiling point", "en ghosh", "evaporation heat", "heat of formation",
                     "lattice constant", "melting point", "specific heat"]
querable_pymatgen = ["atomic_mass", "atomic_radius", "electrical_resistivity", "molar_volume", "bulk_modulus", "youngs_modulus"
                     "average ionic radius", "density of solid", "coefficient of linear thermal expansion"]
querable values = querable mendeleev + querable pymatgen
```

As before, we will use the database queries to populate lists which can be displayed by the Pandas library in a user-friendly table with the properties as the column headers.

```
In []: all_values = [] # Values for Attributes
all_labels = [] # Crystal structure labels (0 = fcc, 1 = bcc, 2 = hcp)

for item in elements:
    element_values = []

# This section queries Mendeleev
    element_object = mendel.element(item)
    for i in querable_mendeleev:
```

```
element values.append(getattr(element object,i))
    # This section queries Pymataen
    element object = pymat.Element(item)
    for i in querable pymatgen:
        element values.append(getattr(element object,i))
    all values.append(element values) # All lists are appended to another list, creating a List of Lists
    if (item in fcc elements):
        all labels.append([1, 0, 0]) # The crystal structure labels are assigned here
    elif (item in bcc elements):
        all labels.append([0, 1, 0]) # The crystal structure labels are assigned here
    elif (item in hcp elements):
        all labels.append([0, 0, 1]) # The crystal structure labels are assigned here
#print("What is inside element values")
#print(element values)
#print("What is inside all labels")
#print(all labels)
# Pandas Dataframe
df = pd.DataFrame(all values, columns=querable values)
# We will patch some of the values that are not available in the datasets.
# Value for the CTE of Cesium
index Cs = df.index[df['atomic number'] == 55]
df.iloc[index Cs, df.columns.get loc("coefficient of linear thermal expansion")] = 0.000097
# Value from: David R. Lide (ed), CRC Handbook of Chemistry and Physics, 84th Edition. CRC Press. Boca Raton, Florida, 2003
# Value for the CTE of Rubidium
index_Rb = df.index[df['atomic_number'] == 37]
df.iloc[index Rb, df.columns.get loc("coefficient of linear thermal expansion")] = 0.000090
# Value from: https://www.azom.com/article.aspx?ArticleID=1834
# Value for the Evaporation Heat of Ruthenium
index Ru = df.index[df['atomic number'] == 44]
df.iloc[index Ru, df.columns.get loc("evaporation heat")] = 595 # kJ/mol
# Value from: https://www.webelements.com/ruthenium/thermochemistry.html
```

```
# Value for the Bulk Modulus of Zirconium
index_Zr = df.index[df['atomic_number'] == 40]
df.iloc[index_Zr, df.columns.get_loc("bulk_modulus")] = 94 # GPa
# Value from: https://materialsproject.org/materials/mp-131/
df.head

c:\Users\manda\AppData\Local\Programs\Python\Python311\Lib\site-packages\pymatgen\core\periodic_table.py:212: UserWarning:
No data available for coefficient_of_linear_thermal_expansion for Cs
c:\Users\manda\AppData\Local\Programs\Python\Python311\Lib\site-packages\pymatgen\core\periodic_table.py:212: UserWarning:
```

c:\Users\manda\AppData\Local\Programs\Python\Python311\Lib\site-packages\pymatgen\core\periodic table.py:212: UserWarning:

No data available for coefficient of linear thermal expansion for Rb

No data available for bulk modulus for Zr

Out[]:	<body> bound method</body>	NDFrame.head	of	atomic_number	atomic_volume	boiling_point	en_ghosh	evaporation_heat	\
	0	27	6.70	3200.15	0.143236	389.1			
	1	69	18.10	2223.15	0.216724	232.0			
	2	39	19.80	3618.15	0.121699	367.0			
	3	75	8.85	5863.15	0.243516	704.0			
	4	28	6.60	3186.15	0.147207	378.6			
	5	67	18.70	2973.15	0.207795	301.0			
	6	79	10.20	3109.15	0.261370	340.0			
	7	21	15.00	3109.15	0.119383	332.7			
	8	45	8.30	3968.15	0.140838	494.0			
	9	74	9.53	5828.15	0.239050	824.0			
	10	64	19.90	3546.15	0.194400	398.0			
	11	65	19.20	3503.15	0.198863	389.0			
	12	72	13.60	4873.15	0.229987	575.0			
	13	70	24.80	1469.15	0.221190	159.0			
	14	55	70.00	944.15	0.154213	68.3			
	15	30	9.20	1180.15	0.155152	114.8			
	16	56	39.00	2118.15	0.158679	142.0			
	17	25	7.39	2334.15	0.135284	221.0			
	18	26	7.10	3134.15	0.139253	340.0			
	19	42	9.40	4912.15	0.131267	590.0			
	20	11	23.70	1156.09	0.093214	97.9			
	21	71	17.80	3675.15	0.225650	414.0			
	22	90	19.80	5058.15	0.102770	513.7			
	23	29	7.10	2833.15	0.151172	304.6			
	24	3	13.10	1615.15	0.105093	148.0			
	25	81	17.20	1746.15	0.173447	162.4			
	26	23	8.35	3680.15	0.127334	460.0			
	27	37	55.90	961.15	0.104686	75.8			
	28	40	14.10	4679.15	0.124889	567.0			
	29	24	7.23	2944.15	0.131305	342.0			
	30	41	10.80	5014.15	0.128078	680.0			
	31	47	10.30	2435.15	0.147217	254.1			
	32	4	5.00	2741.15	0.144986	309.0			
	33	44	8.30	4420.15	0.137649	595.0			
	34	13	10.00	2792.15	0.150078	284.1			
	35	22	10.60	3560.15	0.123364	422.6			
	36	82	18.30	2022.15	0.177911	177.8			
	37	20	29.90		0.115412	153.6			
	38	73	10.90	5728.15	0.234581	758.0			

39	66	19.00	2840.15	מככמה מ	a	291.0
40	48	13.10		0.203336 3.150405		59.1
41	68	18.40		0.150407 3.212261		
41	46	8.90		0.212261 0.144028		317.0 372.4
43		28.90				
44	63 77			0.189935		176.0
	77	8.54		0.251060		604.0
45	12	14.00		0.121644 0.256016		131.8
46	78	9.10	4098.15	a.256916	9	470.0
	heat_of_formation	lattice_constant	melting_p	point s	specific_heat	\
0	426.70	2.51	1768	8.150	0.421	
1	232.20	3.54	1818	8.150	0.160	
2	424.70	3.65	1795	5.150	0.298	
3	774.00	2.76	3458	8.150	0.137	
4	430.10	3.52	1728	8.150	0.444	
5	300.60	3.58	1745	5.150	0.165	
6	368.20	4.08	1337	7.330	0.129	
7	377.80	3.31	1814	4.150	0.568	
8	556.00	3.80	2236	6.150	0.243	
9	851.00	3.16	3687	7.150	0.132	
10	397.50	3.64	1586	6.150	0.236	
11	388.70	3.60	1632	2.150	0.182	
12	618.40	3.20	2506	6.150	0.144	
13	155.60	5.49	1097	7.150	0.155	
14	76.50	6.05	301	1.650	0.242	
15	130.40	2.66	692	2.677	0.388	
16	179.10	5.02	1000	0.150	0.204	
17	283.30	8.89	1519	9.150	0.479	
18	415.50	2.87	1811	1.150	0.449	
19	658.98	3.15	2895	5.150	0.251	
20	107.50	4.23	376	0.944	1.228	
21	427.60	3.51	1936	5.150	0.154	
22	602.00	5.08	2023	3.150	0.118	
23	337.40	3.61	1357	7.770	0.385	
24	159.30	3.49	453	3.650	3.582	
25	182.20	3.46	577	7.150	0.129	
26	515.50	3.02	2183	3.150	0.489	
27	80.90	5.59	312	2.450	0.363	
28	610.00	3.23	2127	7.150	0.278	
29	397.48	2.88	2186	0.150	0.449	
30	733.00	3.30	2756	0.150	0.265	

31	2	84.90	4.09	1234.930	0.235	
32	3	24.00	2.29	1560.150	1.825	
33	6	50.60	2.70	2606.150	0.238	
34	3	30.90	4.05	933.473	0.897	
35	4	73.00	2.95	1943.150	0.523	
36	1	95.20	4.95	600.612	0.130	
37	1	77.80	5.58	1115.150	0.647	
38	7	82.00	3.31	3290.150	0.140	
39	2	90.40	3.59	1685.150	0.173	
40	1	11.80	2.98	594.219	0.232	
41	3	16.40	3.56	1802.150	0.168	
42	3	76.60	3.89	1827.950	0.246	
43	1	77.40	4.61	1095.150	0.182	
44	6	69.00	3.84	2719.150	0.131	
45	1	47.10	3.21	923.150	1.023	
46	5	65.70	3.92	2041.350	0.133	
	atomic_mass	atomic_radius	electrical	l_resistivity	molar_volume	\
0	58.933195	1.35		6.000000e-08	6.67	
1	168.934210	1.75		6.760000e-07	19.10	
2	88.905850	1.80		6.000000e-07	19.88	
3	186.207000	1.35		1.800000e-07	8.86	
4	58.693400	1.35		7.200000e-08	6.59	
5	164.930320	1.75		8.140000e-07	18.74	
6	196.966569	1.35		2.200000e-08	10.21	
7	44.955912	1.60		5.500000e-07	15.00	
8	102.905500	1.35		4.300000e-08	8.28	
9	183.840000	1.35		5.400000e-08	9.47	
10	157.250000	1.80		1.310000e-06	19.90	
11	158.925350	1.75		1.150000e-06	19.30	
12	178.490000	1.55		3.400000e-07	13.44	
13	173.040000	1.75		2.500000e-07	24.84	
14	132.905452	2.60		2.100000e-07	70.94	
15	65.409000	1.35		6.000000e-08	9.16	
16	137.327000	2.15		3.400000e-07	38.16	
17	54.938045	1.40		1.440000e-06	7.35	
18	55.845000	1.40		1.000000e-07	7.09	
19	95.940000	1.45		5.500000e-08	9.38	
20	22.989769	1.80		4.900000e-08	23.78	
21	174.967000	1.75		5.800000e-07	17.78	

1.80

19.80

1.500000e-07

22 232.038060

23	63.546000	1.35	1.720000e-08	7.11	
24	6.941000	1.45	9.500000e-08	13.02	
25	204.383300	1.90	1.500000e-07	17.22	
26	50.941500	1.35	2.000000e-07	8.32	
27	85.467800	2.35	1.330000e-07	55.76	
28	91.224000	1.55	4.330000e-07	14.02	
29	51.996100	1.40	1.270000e-07	7.23	
30	92.906380	1.45	1.520000e-07	10.83	
31	107.868200	1.60	1.630000e-08	10.27	
32	9.012182	1.05	3.800000e-08	4.85	
33	101.070000	1.30	7.100000e-08	8.17	
34	26.981539	1.25	2.700000e-08	10.00	
35	47.867000	1.40	4.000000e-07	10.64	
36	207.200000	1.80	2.100000e-07	18.26	
37	40.078000	1.80	3.400000e-08	26.20	
38	180.947880	1.45	1.350000e-07	10.85	
39	162.500000	1.75	9.260000e-07	19.01	
40	112.411000	1.55	7.000000e-08	13.00	
41	167.259000	1.75	8.600000e-07	18.46	
42	106.420000	1.40	1.080000e-07	8.56	
43	151.964000	1.85	9.000000e-07	28.97	
44	192.217000	1.35	4.700000e-08	8.52	
45	24.305000	1.50	4.400000e-08	14.00	
46	195.084000	1.35	1.060000e-07	9.09	
	_	youngs_modulus	average_ionic_radius	density_of_solid	\
0	180.0	209.0	0.768333	8900.0	
1	45.0	74.0	1.095000	9321.0	
2	41.0	64.0	1.040000	4472.0	
3	370.0	463.0	0.712500	21020.0	
4	180.0	200.0	0.740000	8908.0	
5	40.0	65.0	1.041000	8795.0	
6	220.0	78.0	1.070000	19300.0	
7	57.0	74.0	0.885000	2985.0	
8	380.0	275.0	0.745000	12450.0	
9	310.0	411.0	0.766667	19250.0	
10	38.0	55.0	1.075000	7901.0	
11	38.7	56.0	0.981500	8219.0	
12	110.0	78.0	0.850000	13310.0	
13	31.0	24.0	1.084000	6570.0	
14	1.6	1.7	1.810000	1879.0	

15	70.0	108.0	0.880000	7140.0
16	9.6	13.0	1.490000	3510.0
17	120.0	198.0	0.648333	7470.0
18	170.0	211.0	0.852500	7874.0
19	230.0	329.0	0.775000	10280.0
20	6.3	10.0	1.160000	968.0
21	48.0	69.0	1.001000	9841.0
22	54.0	79.0	1.080000	11724.0
23	140.0	130.0	0.820000	8920.0
24	11.0	4.9	0.900000	535.0
25	43.0	8.0	1.332500	11850.0
26	160.0	128.0	0.777500	6110.0
27	2.5	2.4	1.660000	1532.0
28	94.0	68.0	0.860000	6511.0
29	160.0	279.0	0.940000	7140.0
30	170.0	105.0	0.820000	8570.0
31	100.0	83.0	1.086667	10490.0
32	130.0	287.0	0.590000	1848.0
33	220.0	447.0	0.661000	12370.0
34	76.0	70.0	0.675000	2700.0
35	110.0	116.0	0.851667	4507.0
36	46.0	16.0	1.122500	11340.0
37	17.0	20.0	1.140000	1550.0
38	200.0	186.0	0.820000	16650.0
39	41.0	61.0	1.131000	8551.0
40	42.0	50.0	1.090000	8650.0
41	44.0	70.0	1.030000	9066.0
42	180.0	121.0	0.846250	12023.0
43	8.3	18.0	1.198500	5244.0
44	320.0	528.0	0.765000	22650.0
45	45.0	45.0	0.860000	1738.0
46	230.0	168.0	0.805000	21090.0

coefficient_of_linear_thermal_expansion

0	0.000013
1	0.000013
2	0.000011
3	0.000006
4	0.000013
5	0.000011
6	0.000014

7	0.000010	
8	0.000008	
9	0.000005	
10	0.000009	
11	0.000010	
12	0.000006	
13	0.000026	
14	0.000097	
15	0.000030	
16	0.000021	
17	0.000022	
18	0.000012	
19	0.000005	
20	0.000071	
21	0.000010	
22	0.000011	
23	0.000017	
24	0.000046	
25	0.000030	
26	0.000008	
27	0.000090	
28	0.000006	
29	0.000005	
30	0.000007	
31	0.000019	
32	0.000011	
33	0.000006	
34	0.000023	
35	0.000009	
36	0.000029	
37	0.000022	
38	0.000006	
39	0.000010	
40	0.000031	
41	0.000012	
42	0.000012	
43	0.000035	
44	0.000006	
45	0.000008	
46	0.000009	>

2. Processing and Organizing Data

We again normalize the data and organize it into training and testing sets as before.

SETS

We have 47 elements for which the crystal structure is known and we will use 40 of these as a training set and the remaining 7 as testing set.

NORMALIZATION

We will again use the Standard Score Normalization, which subtracts the mean of the feature and divide by its standard deviation.

$$\frac{X-\mu}{\sigma}$$

While our model might converge without feature normalization, the resultant model would be difficult to train and would be dependent on the choice of units used in the input.

```
In [ ]: # SETS

all_values = [list(df.iloc[x]) for x in range(len(all_values))]

# List of lists are turned into Numpy arrays to facilitate calculations in steps to follow (Normalization).
all_values = np.array(all_values, dtype = float)
print("Shape of Values:", all_values.shape)
all_labels = np.array(all_labels, dtype = int)
print("Shape of Labels:", all_labels.shape)

# Training Set
train_values = all_values[:40, :]
train_labels = all_labels[:40, :]

# Testing Set
test_values = all_values[-7:, :]
test_labels = all_labels[-7:, :]
# NORMALIZATION
```

```
mean = np.nanmean(train_values, axis = 0) # mean, np.nanmean is a version of mean function in numpy that throws errors if all std = np.nanstd(train_values, axis = 0) # standard deviation

train_values = (train_values - mean) / std # input scaling

test_values = (test_values - mean) / std # input scaling

print(train_values[0]) # print a sample entry from the training set

#print(train_labels[0])

Shape of Values: (47, 18)

Shape of Labels: (47, 3)

[-0.80084167 -0.75983551  0.02340813 -0.40732945  0.15599373  0.16654528
-1.09549525  0.09167774 -0.03493069 -0.82400017 -0.80570946 -0.67799461
-0.75661221  0.70972845  0.6516648 -0.77257498  0.11409173 -0.3075323 ]
```

3. Creating the Model

For this classification, we will use a simple sequential neural network with one densely connected hidden layer. We will try many optimizers.

To learn more about Root Mean Squared Propagation, click here.

The key difference between the regression model and the classification model is our metric to measure network performance. While we used mean squared error (between the true outputs and the network's predicted output) for the regression task, we use categorical cross entropy (click here to learn more about it), using classification accuracy as a metric where higher accuracy implies a better network.

```
In []: from tensorflow.keras import optimizers
# DEFINITION OF THE MODEL

# The weights of our neural network will be initialized in a random manner, using a seed allows for reproducibility
kernel_init = initializers.RandomNormal(seed=14)

model = Sequential()
model.add(Dense(32, activation='relu', input_shape=(train_values.shape[1],), kernel_initializer=kernel_init))
#model.add(Dense(16, activation='relu', kernel_initializer=kernel_init))
model.add(Dense(3, activation=tf.nn.softmax)) # Output Layer

# DEFINITION OF THE OPTIMIZER
```

```
#optimizer = optimizers.RMSprop(0.002) # AdaM Optimizer. 0.002 is the learning rate.
#optimizer = optimizers.Adam(0.002) # AdaM Optimizer. 0.002 is the learning rate.
# optimizer = optimizers.SGD(0.002) # AdaM Optimizer. 0.002 is the learning rate.

# This line matches the optimizer to the model and states which metrics will evaluate the model's accuracy model.compile(loss='categorical_crossentropy', optimizer=optimizer, metrics=['accuracy'])
model.summary()
```

Model: "sequential 15"

Layer (type)	Output Shape	Param #
dense_30 (Dense)	(None, 32)	608
dense_31 (Dense)	(None, 3)	99

Total params: 707 (2.76 KB)
Trainable params: 707 (2.76 KB)
Non-trainable params: 0 (0.00 Byte)

Layer (type)	Output Shape	Param #
dense_30 (Dense)	(None, 32)	608
dense_31 (Dense)	(None, 3)	99

Total params: 707 (2.76 KB)
Trainable params: 707 (2.76 KB)
Non-trainable params: 0 (0.00 Byte)

In []: import numpy as np

import matplotlib.pyplot as plt

 $\textbf{from} \ \texttt{tensorflow}. \texttt{keras}. \texttt{models} \ \textbf{import} \ \texttt{Sequential}$

from tensorflow.keras.layers import Dense

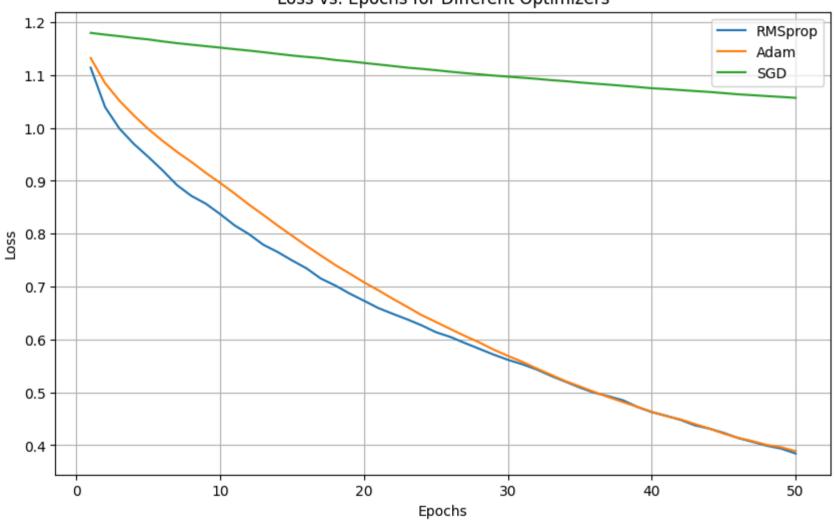
from tensorflow.keras import initializers

from tensorflow.keras import optimizers

```
# Assuming you have training data in 'train values' and corresponding labels in 'train labels'
# Create a Sequential model
kernel init = initializers.RandomNormal(seed=14)
model = Sequential()
model.add(Dense(32, activation='relu', input shape=(train values.shape[1],), kernel initializer=kernel init))
model.add(Dense(3, activation='softmax')) # Output Layer
# Define the optimizers
optimizers list = {
    'RMSprop': optimizers.RMSprop(0.002),
    'Adam': optimizers.Adam(0.002),
    'SGD': optimizers.SGD(0.002)
# Compile the model
model.compile(loss='categorical crossentropy', optimizer='adam', metrics=['accuracy'])
# Plot loss vs epochs for each optimizer on the same graph
epochs = 50 # Define the number of epochs
plt.figure(figsize=(10, 6))
for optimizer name, optimizer in optimizers list.items():
    # Reinitialize the model
    model = Sequential()
    model.add(Dense(32, activation='relu', input shape=(train values.shape[1],), kernel initializer=kernel init))
    model.add(Dense(3, activation='softmax')) # Output Layer
    model.compile(loss='categorical crossentropy', optimizer=optimizer, metrics=['accuracy'])
    # Train the model
    history = model.fit(train values, train labels, epochs=epochs, verbose=0)
    # Plot loss vs epochs
    plt.plot(np.arange(1, epochs + 1), history.history['loss'], label=optimizer name)
    # Reinitialize optimizer
    if optimizer name != list(optimizers list.keys())[-1]:
        optimizer = optimizers.get(optimizer name)
        optimizer.learning rate.assign(0.002)
plt.title('Loss vs. Epochs for Different Optimizers')
plt.xlabel('Epochs')
```

```
plt.ylabel('Loss')
plt.legend()
plt.grid(True)
plt.show()
```





This model is trained for 500 epochs, and we record the training accuracy in the history object. This way, by plotting "history" we can see the evolution of the "learning" of the model, that is the decrease of the Mean Absolute Error. Models in Keras are fitted to the training set using the **fit** method.

One **Epoch** occurs when you pass the entire dataset through the model. One **Batch** contains a subset of the dataset that can be fed to the model at the same time. A more detailed explanation of these concepts can be found in this blog. As we have a really small dataset compared to the ones that are usually considered to be modeled by these neural networks, we are feeding all entries at the same time, so our batch is the entire dataset, and an epoch occurs when the batch is processed.

```
In [ ]: class PrintEpNum(keras.callbacks.Callback): # This is a function for the Epoch Counter
            def on_epoch_end(self, epoch, logs):
                sys.stdout.flush()
                sys.stdout.write("Current Epoch: " + str(epoch+1) + '\r') # Updates current Epoch Number
        EPOCHS = 500 # Number of EPOCHS
        # HISTORY Object which contains how the model Learned
        history = model.fit(train values, train labels, batch size=train values.shape[0], \
                            epochs=EPOCHS, validation split=0.1, verbose = True, callbacks=[PrintEpNum()])
        # PLOTTING HISTORY USING MATPLOTLIB
        plt.figure()
        plt.xlabel('Epoch')
        plt.ylabel('Accuracy')
        #plt.ylim(0.98,1.05)
        plt.plot(history.epoch, np.array(history.history['accuracy']),label='Training Accuracy')
        plt.plot(history.epoch, np.array(history.history['val_accuracy']),label = 'Validation Accuracy')
        plt.legend()
        plt.show()
```

```
a
Epoch 2/500
Epoch 3/500
Epoch 4/500
Epoch 5/500
Epoch 6/500
Epoch 7/500
Epoch 8/500
Epoch 9/500
Epoch 10/500
Epoch 11/500
Epoch 12/500
Epoch 13/500
Epoch 14/500
Epoch 15/500
Epoch 16/500
Epoch 17/500
1/1 [============ ] - 0s 31ms/step - loss: 1.0490 - accuracy: 0.5556 - val loss: 0.9731 - val accuracy: 1.0000
Epoch 18/500
Epoch 19/500
Epoch 20/500
1/1 [============ ] - 0s 63ms/step - loss: 1.0464 - accuracy: 0.5556 - val loss: 0.9714 - val accuracy: 1.0000
Epoch 21/500
```

```
Epoch 22/500
Epoch 23/500
Epoch 24/500
Epoch 25/500
Epoch 26/500
1/1 [============= ] - 0s 47ms/step - loss: 1.0412 - accuracy: 0.5833 - val loss: 0.9682 - val accuracy: 1.0000
Epoch 27/500
Epoch 28/500
Epoch 29/500
Epoch 30/500
Epoch 31/500
Epoch 32/500
Epoch 33/500
Epoch 34/500
Epoch 35/500
Epoch 36/500
Epoch 37/500
Epoch 38/500
Epoch 39/500
Epoch 40/500
Epoch 41/500
```

```
Epoch 42/500
Epoch 43/500
1/1 [============= ] - 0s 32ms/step - loss: 1.0275 - accuracy: 0.6389 - val loss: 0.9587 - val accuracy: 1.0000
Epoch 44/500
1/1 [=========== ] - 0s 47ms/step - loss: 1.0267 - accuracy: 0.6389 - val loss: 0.9582 - val accuracy: 1.0000
Epoch 45/500
1/1 [============= ] - 0s 31ms/step - loss: 1.0259 - accuracy: 0.6389 - val loss: 0.9577 - val accuracy: 1.0000
Epoch 46/500
Epoch 47/500
1/1 [============ ] - 0s 31ms/step - loss: 1.0244 - accuracy: 0.6389 - val loss: 0.9566 - val accuracy: 1.0000
Epoch 48/500
1/1 [============ ] - 0s 47ms/step - loss: 1.0236 - accuracy: 0.6389 - val loss: 0.9560 - val accuracy: 1.0000
Epoch 49/500
Epoch 50/500
1/1 [============= ] - 0s 47ms/step - loss: 1.0221 - accuracy: 0.6389 - val loss: 0.9549 - val accuracy: 1.0000
Epoch 51/500
Epoch 52/500
Epoch 53/500
Epoch 54/500
Epoch 55/500
1/1 [============ ] - 0s 47ms/step - loss: 1.0185 - accuracy: 0.6389 - val loss: 0.9523 - val accuracy: 1.0000
Epoch 56/500
Epoch 57/500
1/1 [============ ] - 0s 47ms/step - loss: 1.0170 - accuracy: 0.6389 - val loss: 0.9512 - val accuracy: 1.0000
Epoch 58/500
1/1 [============== ] - 0s 62ms/step - loss: 1.0163 - accuracy: 0.6389 - val loss: 0.9507 - val accuracy: 1.0000
Epoch 59/500
Epoch 60/500
Epoch 61/500
1/1 [============ ] - 0s 47ms/step - loss: 1.0141 - accuracy: 0.6389 - val loss: 0.9492 - val accuracy: 1.0000
Epoch 62/500
```

```
Epoch 63/500
Epoch 64/500
Epoch 65/500
Epoch 66/500
Epoch 67/500
Epoch 68/500
Epoch 69/500
Epoch 70/500
Epoch 71/500
Epoch 72/500
Epoch 73/500
Epoch 74/500
Epoch 75/500
Epoch 76/500
Epoch 77/500
Epoch 78/500
Epoch 79/500
Epoch 80/500
Epoch 81/500
Epoch 82/500
```

```
Epoch 83/500
Epoch 84/500
Epoch 85/500
Epoch 86/500
Epoch 87/500
Epoch 88/500
Epoch 89/500
Epoch 90/500
Epoch 91/500
Epoch 92/500
Epoch 93/500
Epoch 94/500
Epoch 95/500
Epoch 96/500
Epoch 97/500
Epoch 98/500
Epoch 99/500
1/1 [============= ] - 0s 47ms/step - loss: 0.9886 - accuracy: 0.6389 - val loss: 0.9313 - val accuracy: 0.7500
Epoch 100/500
Epoch 101/500
Epoch 102/500
1/1 [============= ] - 0s 32ms/step - loss: 0.9867 - accuracy: 0.6389 - val loss: 0.9299 - val accuracy: 0.7500
Epoch 103/500
```

```
Epoch 104/500
Epoch 105/500
Epoch 106/500
Epoch 107/500
Epoch 108/500
Epoch 109/500
Epoch 110/500
Epoch 111/500
Epoch 112/500
Epoch 113/500
Epoch 114/500
Epoch 115/500
Epoch 116/500
Epoch 117/500
Epoch 118/500
Epoch 119/500
Epoch 120/500
Epoch 121/500
Epoch 122/500
Epoch 123/500
```

```
Epoch 124/500
Epoch 125/500
Epoch 126/500
Epoch 127/500
Epoch 128/500
Epoch 129/500
Epoch 130/500
Epoch 131/500
Epoch 132/500
Epoch 133/500
Epoch 134/500
Epoch 135/500
Epoch 136/500
Epoch 137/500
Epoch 138/500
Epoch 139/500
Epoch 140/500
Epoch 141/500
Epoch 142/500
Epoch 143/500
Epoch 144/500
```

```
Epoch 145/500
Epoch 146/500
Epoch 147/500
Epoch 148/500
Epoch 149/500
Epoch 150/500
Epoch 151/500
Epoch 152/500
Epoch 153/500
Epoch 154/500
Epoch 155/500
Epoch 156/500
Epoch 157/500
Epoch 158/500
Epoch 159/500
Epoch 160/500
Epoch 161/500
Epoch 162/500
Epoch 163/500
Epoch 164/500
```

```
Epoch 165/500
Epoch 166/500
Epoch 167/500
Epoch 168/500
Epoch 169/500
Epoch 170/500
Epoch 171/500
Epoch 172/500
Epoch 173/500
Epoch 174/500
Epoch 175/500
Epoch 176/500
Epoch 177/500
Epoch 178/500
Epoch 179/500
Epoch 180/500
Epoch 181/500
Epoch 182/500
Epoch 183/500
Epoch 184/500
Epoch 185/500
```

```
Epoch 186/500
Epoch 187/500
Epoch 188/500
Epoch 189/500
Epoch 190/500
Epoch 191/500
Epoch 192/500
Epoch 193/500
Epoch 194/500
Epoch 195/500
Epoch 196/500
Epoch 197/500
Epoch 198/500
Epoch 199/500
Epoch 200/500
Epoch 201/500
Epoch 202/500
Epoch 203/500
Epoch 204/500
Epoch 205/500
```

```
Epoch 206/500
Epoch 207/500
Epoch 208/500
Epoch 209/500
Epoch 210/500
Epoch 211/500
Epoch 212/500
Epoch 213/500
Epoch 214/500
Epoch 215/500
Epoch 216/500
Epoch 217/500
Epoch 218/500
Epoch 219/500
Epoch 220/500
Epoch 221/500
Epoch 222/500
Epoch 223/500
1/1 [============= ] - 0s 55ms/step - loss: 0.9234 - accuracy: 0.6944 - val loss: 0.8840 - val accuracy: 0.7500
Epoch 224/500
Epoch 225/500
```

```
Epoch 226/500
Epoch 227/500
Epoch 228/500
Epoch 229/500
Epoch 230/500
Epoch 231/500
Epoch 232/500
Epoch 233/500
Epoch 234/500
Epoch 235/500
Epoch 236/500
Epoch 237/500
Epoch 238/500
Epoch 239/500
1/1 [============= ] - 0s 52ms/step - loss: 0.9160 - accuracy: 0.6944 - val loss: 0.8787 - val accuracy: 0.7500
Epoch 240/500
Epoch 241/500
1/1 [============= ] - 0s 58ms/step - loss: 0.9151 - accuracy: 0.6944 - val loss: 0.8780 - val accuracy: 0.7500
Epoch 242/500
1/1 [============ ] - 0s 33ms/step - loss: 0.9146 - accuracy: 0.6944 - val loss: 0.8777 - val accuracy: 0.7500
Epoch 243/500
Epoch 244/500
Epoch 245/500
1/1 [============= ] - 0s 25ms/step - loss: 0.9132 - accuracy: 0.6944 - val loss: 0.8767 - val accuracy: 0.7500
Epoch 246/500
```

```
Epoch 247/500
Epoch 248/500
Epoch 249/500
Epoch 250/500
Epoch 251/500
Epoch 252/500
Epoch 253/500
Epoch 254/500
Epoch 255/500
Epoch 256/500
Epoch 257/500
Epoch 258/500
Epoch 259/500
Epoch 260/500
Epoch 261/500
Epoch 262/500
Epoch 263/500
Epoch 264/500
Epoch 265/500
```

```
Epoch 266/500
Epoch 267/500
Epoch 268/500
Epoch 269/500
Epoch 270/500
Epoch 271/500
Epoch 272/500
Epoch 273/500
Epoch 274/500
Epoch 275/500
Epoch 276/500
Epoch 277/500
Epoch 278/500
Epoch 279/500
Epoch 280/500
Epoch 281/500
1/1 [============ ] - 0s 41ms/step - loss: 0.8971 - accuracy: 0.6944 - val loss: 0.8647 - val accuracy: 0.7500
Epoch 282/500
1/1 [============ ] - 0s 79ms/step - loss: 0.8967 - accuracy: 0.6944 - val loss: 0.8643 - val accuracy: 0.7500
Epoch 283/500
Epoch 284/500
Epoch 285/500
1/1 [============ ] - 0s 83ms/step - loss: 0.8954 - accuracy: 0.6944 - val loss: 0.8633 - val accuracy: 0.7500
Epoch 286/500
```

```
Epoch 287/500
Epoch 288/500
Epoch 289/500
Epoch 290/500
Epoch 291/500
Epoch 292/500
Epoch 293/500
Epoch 294/500
Epoch 295/500
Epoch 296/500
Epoch 297/500
Epoch 298/500
Epoch 299/500
Epoch 300/500
Epoch 301/500
Epoch 302/500
Epoch 303/500
Epoch 304/500
Epoch 305/500
Epoch 306/500
```

```
Epoch 307/500
Epoch 308/500
Epoch 309/500
Epoch 310/500
Epoch 311/500
Epoch 312/500
Epoch 313/500
Epoch 314/500
Epoch 315/500
Epoch 316/500
Epoch 317/500
Epoch 318/500
Epoch 319/500
Epoch 320/500
Epoch 321/500
Epoch 322/500
Epoch 323/500
Epoch 324/500
Epoch 325/500
Epoch 326/500
```

```
Epoch 327/500
Epoch 328/500
Epoch 329/500
Epoch 330/500
Epoch 331/500
Epoch 332/500
Epoch 333/500
Epoch 334/500
Epoch 335/500
Epoch 336/500
Epoch 337/500
Epoch 338/500
Epoch 339/500
Epoch 340/500
Epoch 341/500
Epoch 342/500
Epoch 343/500
Epoch 344/500
1/1 [============ ] - 0s 63ms/step - loss: 0.8718 - accuracy: 0.6944 - val loss: 0.8444 - val accuracy: 0.7500
Epoch 345/500
Epoch 346/500
1/1 [=========================== ] - 0s 56ms/step - loss: 0.8711 - accuracy: 0.6944 - val loss: 0.8437 - val accuracy: 0.7500
```

```
Epoch 347/500
Epoch 348/500
Epoch 349/500
Epoch 350/500
Epoch 351/500
Epoch 352/500
Epoch 353/500
Epoch 354/500
Epoch 355/500
Epoch 356/500
Epoch 357/500
Epoch 358/500
Epoch 359/500
Epoch 360/500
Epoch 361/500
Epoch 362/500
1/1 [============= ] - 0s 78ms/step - loss: 0.8650 - accuracy: 0.6944 - val loss: 0.8386 - val accuracy: 0.7500
Epoch 363/500
1/1 [============== ] - 0s 31ms/step - loss: 0.8646 - accuracy: 0.6944 - val loss: 0.8383 - val accuracy: 0.7500
Epoch 364/500
Epoch 365/500
Epoch 366/500
1/1 [============= ] - 0s 47ms/step - loss: 0.8635 - accuracy: 0.6944 - val loss: 0.8373 - val accuracy: 0.7500
Epoch 367/500
```

```
Epoch 368/500
Epoch 369/500
Epoch 370/500
Epoch 371/500
Epoch 372/500
Epoch 373/500
Epoch 374/500
Epoch 375/500
Epoch 376/500
Epoch 377/500
Epoch 378/500
Epoch 379/500
Epoch 380/500
Epoch 381/500
Epoch 382/500
Epoch 383/500
Epoch 384/500
Epoch 385/500
Epoch 386/500
Epoch 387/500
```

```
Epoch 388/500
Epoch 389/500
Epoch 390/500
Epoch 391/500
Epoch 392/500
Epoch 393/500
Epoch 394/500
Epoch 395/500
Epoch 396/500
Epoch 397/500
Epoch 398/500
Epoch 399/500
Epoch 400/500
Epoch 401/500
Epoch 402/500
Epoch 403/500
1/1 [============= ] - 0s 62ms/step - loss: 0.8499 - accuracy: 0.7222 - val loss: 0.8265 - val accuracy: 0.7500
Epoch 404/500
1/1 [============= ] - 0s 62ms/step - loss: 0.8495 - accuracy: 0.7222 - val loss: 0.8262 - val accuracy: 0.7500
Epoch 405/500
Epoch 406/500
Epoch 407/500
1/1 [============= ] - 0s 78ms/step - loss: 0.8485 - accuracy: 0.7222 - val loss: 0.8255 - val accuracy: 0.7500
Epoch 408/500
```

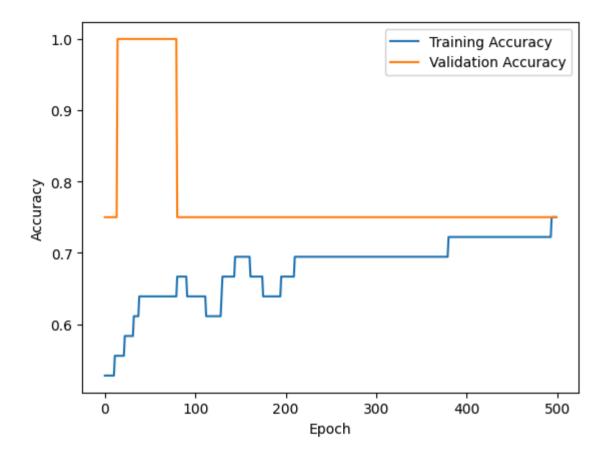
```
Epoch 409/500
Epoch 410/500
Epoch 411/500
Epoch 412/500
Epoch 413/500
Epoch 414/500
Epoch 415/500
Epoch 416/500
Epoch 417/500
Epoch 418/500
Epoch 419/500
Epoch 420/500
Epoch 421/500
Epoch 422/500
Epoch 423/500
Epoch 424/500
Epoch 425/500
Epoch 426/500
Epoch 427/500
Epoch 428/500
```

```
Epoch 429/500
Epoch 430/500
Epoch 431/500
Epoch 432/500
Epoch 433/500
Epoch 434/500
Epoch 435/500
Epoch 436/500
Epoch 437/500
Epoch 438/500
Epoch 439/500
Epoch 440/500
Epoch 441/500
Epoch 442/500
Epoch 443/500
Epoch 444/500
Epoch 445/500
1/1 [============ ] - 0s 39ms/step - loss: 0.8354 - accuracy: 0.7222 - val loss: 0.8151 - val accuracy: 0.7500
Epoch 446/500
Epoch 447/500
Epoch 448/500
1/1 [============= ] - 0s 67ms/step - loss: 0.8344 - accuracy: 0.7222 - val loss: 0.8143 - val accuracy: 0.7500
Epoch 449/500
```

```
Epoch 450/500
Epoch 451/500
Epoch 452/500
Epoch 453/500
Epoch 454/500
Epoch 455/500
Epoch 456/500
Epoch 457/500
Epoch 458/500
Epoch 459/500
Epoch 460/500
Epoch 461/500
Epoch 462/500
Epoch 463/500
Epoch 464/500
Epoch 465/500
Epoch 466/500
Epoch 467/500
Epoch 468/500
Epoch 469/500
```

```
Epoch 470/500
Epoch 471/500
Epoch 472/500
Epoch 473/500
Epoch 474/500
Epoch 475/500
Epoch 476/500
Epoch 477/500
Epoch 478/500
Epoch 479/500
Epoch 480/500
Epoch 481/500
Epoch 482/500
Epoch 483/500
Epoch 484/500
Epoch 485/500
Epoch 486/500
Epoch 487/500
Epoch 488/500
Epoch 489/500
```

```
Epoch 490/500
Epoch 491/500
Epoch 492/500
Epoch 493/500
Epoch 494/500
Epoch 495/500
Epoch 496/500
Epoch 497/500
Epoch 498/500
Epoch 499/500
Epoch 500/500
```



TESTING

Models in Keras are tested using the method evaluate. This method returns the classification accuracy on the training and the testing sets.

```
In [ ]: loss, acc = model.evaluate(train_values, train_labels, verbose=1)
    print("Training Set Accuracy: %f" %(acc))
    loss, acc = model.evaluate(test_values, test_labels, verbose=1)
    print("Testing Set Accuracy: %f" %(acc))
```

MAKING PREDICTIONS

The last step in a Regression Model is to make predictions for values not in the training set, which are determined by the method **predict**. In the following cell we print the Elements in the testing set, the real values for their Young's Moduli and the predictions generated by the Machine Learning model.

```
In [ ]: train predictions = model.predict(train values)
        test predictions = model.predict(test values)
        all labels = np.vstack((train labels, test labels))
        all predictions = np.vstack((train predictions, test predictions))
        predicted labels = []
        true labels = []
        for i in range(all predictions.shape[0]):
            if (np.argmax(all predictions[i]) == 0): # np.argmax returns the index of maximum value along an axis.
                # Here we are looking for the value
                predicted labels.append("FCC")
            if (np.argmax(all labels[i]) == 0):
                true labels.append("FCC")
            if (np.argmax(all predictions[i]) == 1):
                predicted labels.append("BCC")
            if (np.argmax(all labels[i]) == 1):
                true labels.append("BCC")
            if (np.argmax(all predictions[i]) == 2):
                predicted labels.append("HCP")
            if (np.argmax(all labels[i]) == 2):
                true labels.append("HCP")
        predicted labels = np.array(predicted labels).reshape((-1, 1))
        true labels = np.array(true labels).reshape((-1, 1))
        headings = ["Atomic number", "True crystal structure", "Predicted crystal structure"]
```

Out[]:		Atomic number	True crystal structure	Predicted crystal structure
	0	27	НСР	ВСС
	1	69	НСР	НСР
	2	39	НСР	НСР
	3	75	НСР	НСР
	4	28	FCC	ВСС
	5	67	НСР	НСР
	6	79	FCC	FCC
	7	21	НСР	НСР
	8	45	FCC	ВСС
	9	74	ВСС	НСР
	10	64	НСР	НСР
	11	65	НСР	НСР
	12	72	НСР	НСР
	13	70	FCC	НСР
	14	55	ВСС	ВСС
	15	30	НСР	НСР
	16	56	ВСС	ВСС
	17	25	ВСС	ВСС
	18	26	ВСС	ВСС
	19	42	ВСС	ВСС
	20	11	ВСС	ВСС
	21	71	НСР	НСР

	Atomic number	True crystal structure	Predicted crystal structure
22	90	FCC	FCC
23	29	FCC	НСР
24	3	ВСС	ВСС
25	81	НСР	НСР
26	23	ВСС	ВСС
27	37	ВСС	ВСС
28	40	НСР	НСР
29	24	ВСС	ВСС
30	41	ВСС	ВСС
31	47	FCC	НСР
32	4	НСР	НСР
33	44	НСР	ВСС
34	13	FCC	ВСС
35	22	НСР	НСР
36	82	FCC	FCC
37	20	ВСС	ВСС
38	73	ВСС	НСР
39	66	НСР	НСР
40	48	НСР	FCC
41	68	НСР	НСР
42	46	FCC	FCC
43	63	ВСС	НСР

	Atomic number	irue crystai structure	Predicted crystal structure
44	77	FCC	НСР
45	12	НСР	ВСС
46	78	FCC	НСР

```
In [ ]: crystal structures = ["FCC", "BCC", "HCP"]
        FCC prediction = []
        BCC prediction = []
        HCP prediction = []
        for item in range(len(all predictions)):
            FCC prediction.append(all predictions[item].tolist()[0])
            BCC prediction.append(all predictions[item].tolist()[1])
            HCP prediction.append(all predictions[item].tolist()[2])
        # This block will be used to sort the elements by their atomic number
        atomic number = list(df.iloc[:, 0]) # From the Pandas Dataset
        order = np.argsort(atomic number) # Sorting Indexes
        # Sorting the lists by the indexes
        # elements = [elements[x] for x in order]
        # FCC prediction = [FCC prediction[x] for x in order]
        # BCC prediction = [BCC prediction[x] for x in order]
        # HCP prediction =[HCP prediction[x] for x in order]
```

```
import plotly as py
import plotly.graph_objs as go
from plotly.subplots import make_subplots
from plotly.offline import iplot

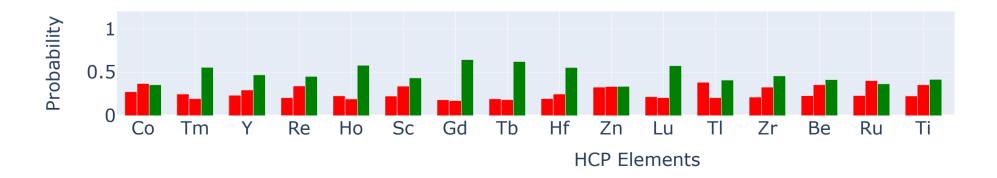
py.offline.init_notebook_mode(connected=True)
```

```
fig = make subplots(rows=3, cols=1, vertical spacing=0.2)
fig.append trace(go.Bar(x=[ for in elements if in fcc elements], y=[FCC prediction[] for in range(len(FCC prediction))
                       text=['*' if in elements[-7:] else None for in [ for in elements if in fcc elements]]), row=
fig.append trace(go.Bar(x=[ for in elements if in fcc elements], y=[BCC prediction[] for in range(len(BCC prediction))
fig.append trace(go.Bar(x=[ for in elements if in fcc elements], y=[HCP prediction[] for in range(len(HCP prediction))
# -----
fig.append trace(go.Bar(x=[ for in elements if in bcc elements], y=[FCC prediction[] for in range(len(FCC prediction))
fig.append trace(go.Bar(x=[ for in elements if in bcc elements], y=[BCC prediction[] for in range(len(BCC prediction))
                       text=['*' if in elements[-7:] else None for in [ for in elements if in bcc elements]]), row=
fig.append trace(go.Bar(x=[ for in elements if in bcc elements], y=[HCP prediction[] for in range(len(HCP prediction))
fig.append trace(go.Bar(x=[ for in elements if in hcp elements], y=[FCC prediction[] for in range(len(FCC prediction))
fig.append trace(go.Bar(x=[ for in elements if in hcp elements], y=[BCC prediction[] for in range(len(BCC prediction))
fig.append trace(go.Bar(x=[ for in elements if in hcp elements], y=[HCP prediction[] for in range(len(HCP prediction))
                       text=['*' if in elements[-7:] else None for in [ for in elements if in hcp elements]]), row=
# -----
fig.update xaxes(title=go.layout.xaxis.Title(text="FCC Elements", font=dict(size=18)), showgrid=True, tickfont=dict(size=18), r
fig.update xaxes(title=go.layout.xaxis.Title(text="BCC Elements", font=dict(size=18)), showgrid=True, tickfont=dict(size=18), r
fig.update xaxes(title=go.layout.xaxis.Title(text="HCP Elements", font=dict(size=18)), showgrid=True, tickfont=dict(size=18), r
fig.update yaxes(title=go.layout.yaxis.Title(text="Probability", font=dict(size=18)),showgrid=True, tickfont=dict(size=18),ran
fig.update yaxes(title=go.layout.yaxis.Title(text="Probability", font=dict(size=18)),showgrid=True, tickfont=dict(size=18),ran
fig.update vaxes(title=go.layout.vaxis.Title(text="Probability", font=dict(size=18)),showgrid=True, tickfont=dict(size=18),ran
fig.update layout(height=700, width=1200, barmode='group', bargap=0.3)
```

fig.show()







In []: