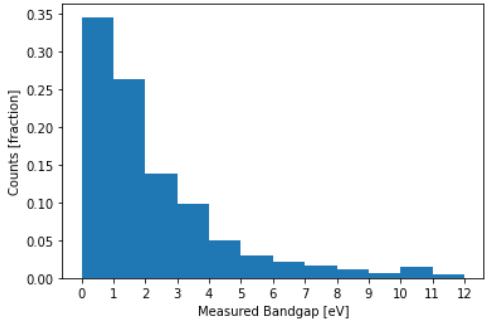
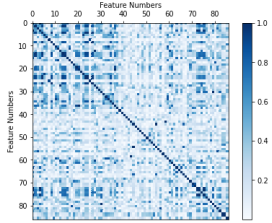
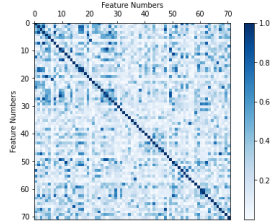
Assessment Figures



	AtomicNumber_composition_average	AtomicRadii_composition_average	AtomicVolume_composition_average	AtomicWeight_composition_average	BCCelfatcnt_composition_average	BCCenergy_pa_composition_average
0	6.000000	1.135000	9311.576313	12.966702	5.772386	-1.346741
1	10.000000	1.270000	9169.525548	21.197000	6.658641	-1.410040
2	19.000000	1.345000	32.035942	43.422500	6.919510	-1.432093
3	15.000000	1.560000	23.705899	35.645750	6.704252	-2.371930
4	28.000000	1.440000	32.101458	66.922735	7.343549	-1.459519
...	...	...	...	...	...	...
482	60.500000	1.422500	40.865008	147.423452	8.158525	-1.763974
483	83.000000	1.700000	35.463459	208.980400	7.821898	-3.895384
484	35.333333	1.096000	12405.753339	88.012287	5.956046	-4.062017
485	50.000000	1.057500	8336.473007	127.014155	5.880448	-6.738593
486	38.000000	0.948333	12401.714393	90.008237	5.551602	-5.312306





ML4ER - Basics of Machine Learning

Muhammad Zain Azeem,
Informatics Skunkworks (**non-credits),** Week 1
19/07/2024

Progress

Step 0: Libraires Installation

Installing all necessary libraries for the accomplishment of this lab activity.

	index	chemicalFormula Clean	Band gap values Clean	Band gap units	Band gap method	Reliability
0	0	Li1F1	13.60	eV	Reflection	1
1	1	Li1F1	12.61	eV	Reflection	1
2	2	Li1F1	12.60	eV	Estimated	2
3	3	Li1F1	12.10	eV	Absorption	2
4	4	Li1F1	12.00	eV	Absorption	2
...	...	...	...	...	...	...
1442	1454	Th1O2	3.30	eV	Reflection	2
1443	1455	UO	1.50	eV	Thermal activation	1
1444	1456	U1O2	2.18	eV	Absorption	1
1445	1457	UO	0.60	eV	Thermal activation	2
1446	1458	U1O2	1.30	eV	Thermal activation	2

Data Filtering

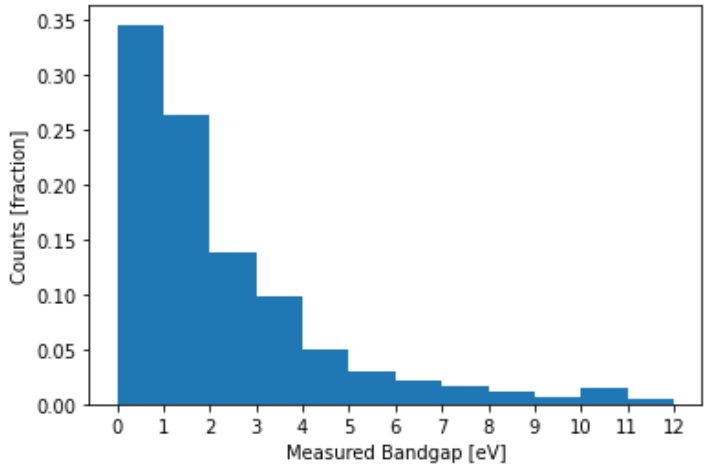
	chemicalFormula Clean	index	Band gap values Clean	Reliability
0	Ag1Br1	808.5	3.485	1.0
1	Ag1Cl1	793.5	4.190	1.0
2	Ag1N3	783.0	3.900	1.0
3	Ag1Te1	820.0	0.850	1.0
4	Ag2O1	785.0	1.200	1.0
...	...	...	...	...
462	Zr1Se2	739.0	2.000	1.0
463	Zr1Se3	740.0	1.250	1.0
464	Zr3O1	730.0	0.200	1.0
465	Zr6O1	729.0	0.180	1.0
466	ZrC	728.0	0.600	1.0

Data Analysis & Filtering

Step 1: Based on data filtering to work on specific datasets. First, import our CSV based on the semiconductor band gap to accomplish this task. Afterward, unreliable data points were removed (i.e., 2). As the given dataset consists of various kinds of solar cell material with different band gap values, we group all similar materials and create a new column with its mean value.

Progress

Step 3 (a): Data Visualization – Band Gap Distribution



The given histogram shows the range of band gap in our dataset which is about up to 12 eV

Step 3 (b): Data Visualization – Composition Analysis

	Element	Count
8	O	240
10	Se	196
9	S	191
11	Te	187
15	As	141
...	...	...
66	Ta	3
39	Y	3
70	Ir	2
62	Tm	1
64	Lu	1

The table indicates several elements available in our dataset and in what quantity.

Progress

Step 4: Composition Average Elemental Properties to featurize a dataset (MASTML CONFIGURATION)

	chemicalFormula Clean	Band gap values Clean	Band gap units	index	Reliability
0	Li1F1	13.105	eV	0	1
1	Li1Cl1	9.33	eV	6	1
2	Li1Br1	7.95	eV	7	1
3	Li3Sb1	1	eV	9	1
4	Li1I1	6	eV	10	1
...	...	...	...	...	...
462	Bi1I3	1.9625	eV	1437	1
463	Bi	0.015	eV	1445	1
464	Th1O2	4.625	eV	1448	1
465	UO	1.5	eV	1455	1
466	U1O2	2.18	eV	1456	1

	AtomicNumber_composition_average	AtomicRadii_composition_average	AtomicVolume_composition_average	AtomicWeight_composition_average	BCCefflatcnt_composition_average
0	6.000000	1.135000	9311.576313	12.969702	5.772386
1	10.000000	1.270000	9169.525548	21.197000	6.658641
2	19.000000	1.345000	32.035942	43.422500	6.919518
3	15.000000	1.560000	23.705899	35.645750	6.704252
4	28.000000	1.440000	32.101458	66.922735	7.343549
...	...	...	...	...	...
462	60.500000	1.422500	40.865008	147.423452	8.158525
463	83.000000	1.700000	35.483459	208.980400	7.821898
464	35.333333	1.086000	12405.753339	88.012287	5.956046
465	50.000000	1.057500	9306.473007	127.014155	5.880448
466	36.000000	0.948333	12401.714393	90.009237	5.551922

467 rows × 87 columns

Progress

Step 5: Correlation Between Features

Pearson Correlation

	AtomicNumber_composition_average	AtomicRadii_composition_average	AtomicVolume_composition_average	AtomicWeight_composition_average	BCCefflatcnt_composition_average
AtomicNumber_composition_average	1.000000	0.585580	0.405061	0.998598	0.628049
AtomicRadii_composition_average	0.585580	1.000000	0.609457	0.571820	0.918506
AtomicVolume_composition_average	0.405061	0.609457	1.000000	0.382296	0.449378
AtomicWeight_composition_average	0.998598	0.571820	0.382296	1.000000	0.615523
BCCefflatcnt_composition_average	0.628049	0.918506	0.449378	0.615523	1.000000

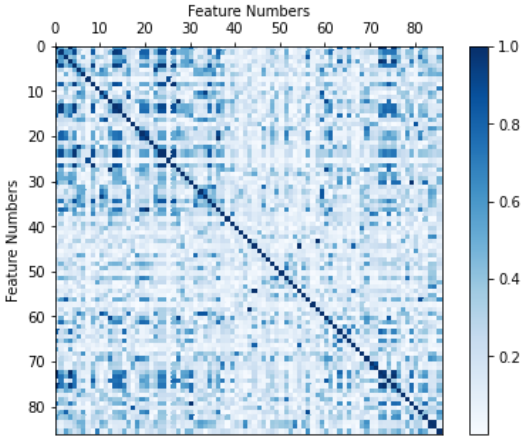
- Atomic Number is strongly correlated to atomic weight composition same for Atomic Radii Vs BCC; atomic weight Vs atomic number; and BCC vs atomic radii
- Atomic Volume vs. Atomic radii showed a weak correlation

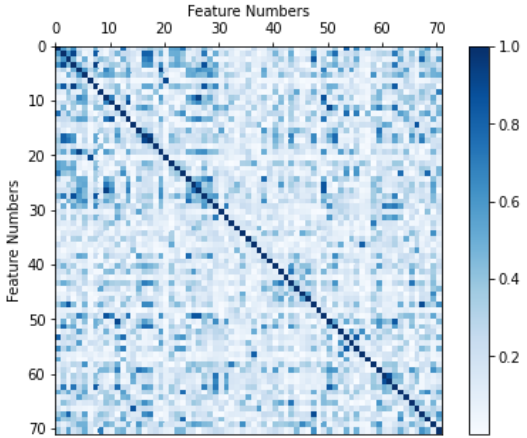
Progress

Pearson Correlation

	AtomicNumber_composition_average	AtomicRadii_composition_average	AtomicVolume_composition_average	AtomicWeight_composition_average	BCCefflatcnt_composition_average
AtomicNumber_composition_average	1.000000	0.585580	0.405061	0.998598	0.628049
AtomicRadii_composition_average	0.585580	1.000000	0.609457	0.571820	0.918506
AtomicVolume_composition_average	0.405061	0.609457	1.000000	0.382296	0.449378
AtomicWeight_composition_average	0.998598	0.571820	0.382296	1.000000	0.615523
BCCefflatcnt_composition_average	0.628049	0.918506	0.449378	0.615523	1.000000

- Before removing the features
- Total features: 80+





- After removing the features
- Total features: 70