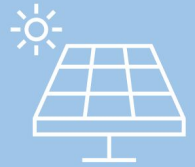
Predicting Perovskite Formability Using Machine Learning

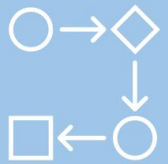
CHEM404 Project Presentation

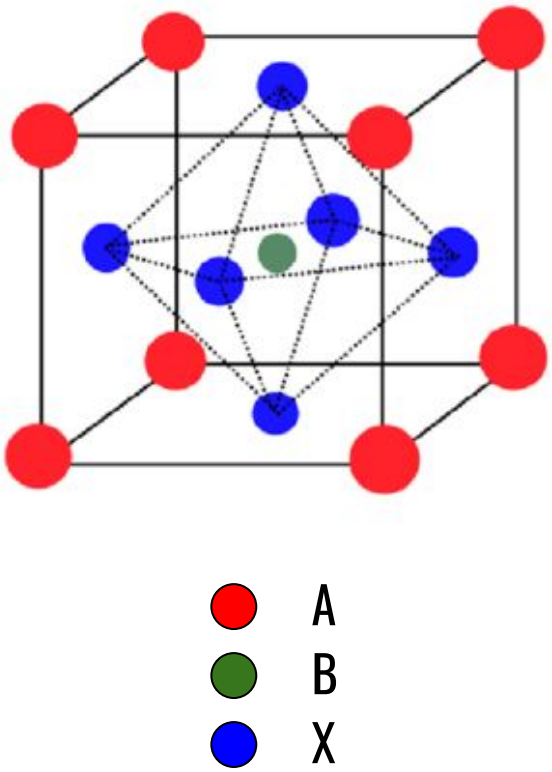
Tom Hou
Eleanor Liu
Leo Xie
Luke Yang

Outline

1. Introduction to Perovskites
2. Machine Learning
3. Methodology
4. Results
5. Conclusions







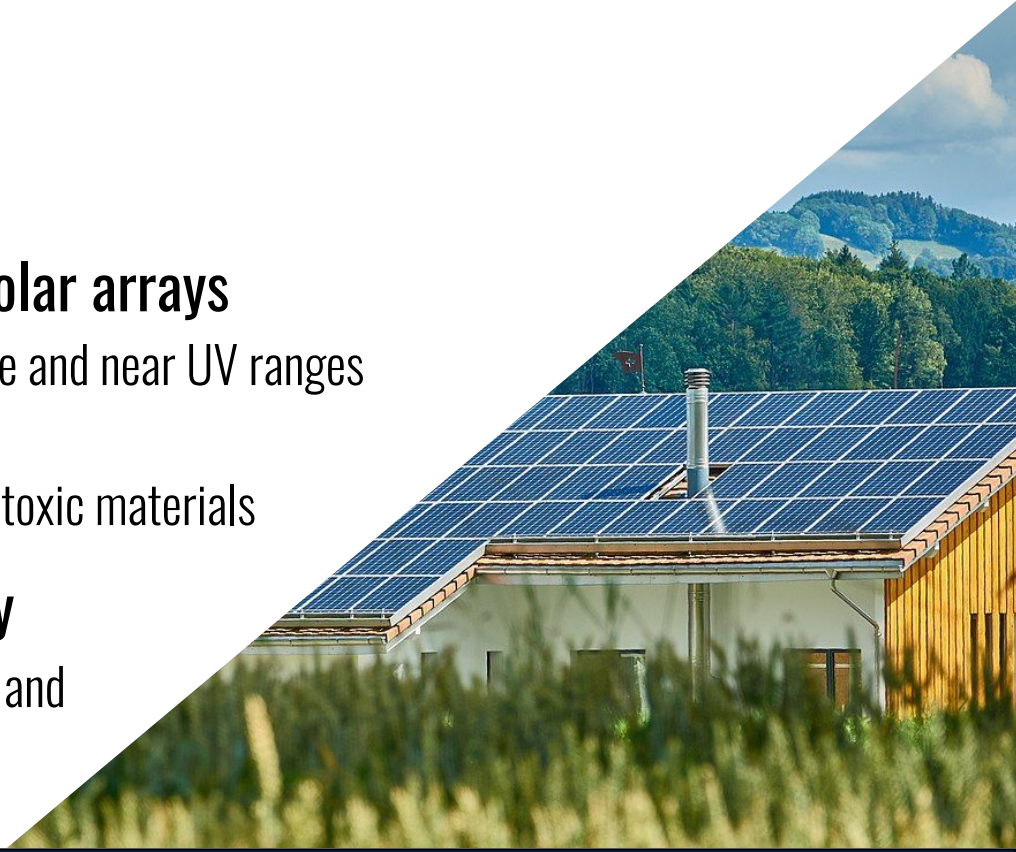
The Perovskite Structure

- Strictly a mineral of formula $CaTiO_3$
- Now used to refer to crystalline structures of ABX_3 of the same crystal structure
  - A = Large metal cation
  - B = Small metal cation
  - X = Electronegative anion (F, Cl, Br, I, O)

Tilley, R. J. D. *Perovskites: Structure-Property Relationships*; John Wiley & Sons: West Sussex, United Kingdom, **2016**.

Perovskite Applications

- Cheaper, safer, thinner, lighter solar arrays
  - Good absorbance coefficient in visible and near UV ranges
  - Long diffusion distances for excitons
  - Possibility to be made of cheap, non-toxic materials
- >20% solar conversion efficiency
  - GaAs is more efficient but expensive and potentially toxic



Chen, Y.; Zhang, L.; Zhang.; Gao, H.; Yan, H. *RSC Adv.*, **2018**, *8*, 10489-10508.

Stranks, S.D.; Eperon, G. E.; Grancini, G.; Menalaou, C.; Alcocer, M. J. P.; Leijtens, T.; Herz, L. M.; Petrozza, A.; Snaith, H. J. *Science*, **2013**, *342*, 341-344.

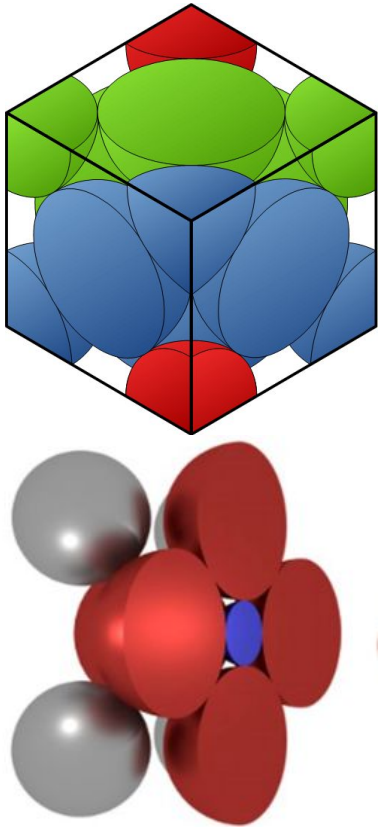
De Wolf, S.; Holovksy, J.; Moon, S. J.; Löper, P.; Niesen, B.; Ledinsky, M.; Haug, F. J.; Yum, J. H.; Ballif, C. *d*

Crystal Structure

- Not all structures of formula ABX_3 form perovskites
  - Dependent on spatial parameters: radii of ions
- Tolerance, τ, defined as:

$$\tau = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$

- τ = 1 defines ideal sphere packing of cubic perovskite
  - Formability probable in range $0.813 < \tau < 1.107$
- Octahedral, o_f, factor defined as: $o_f = \frac{r_B}{r_X}$



Assadi, M. K.; Bakhoda, S.; Saidur, R.; Hanaei, H.; *Renew. Sust. Energ. Rev.*, **2017**, *81*, 2812-2822.

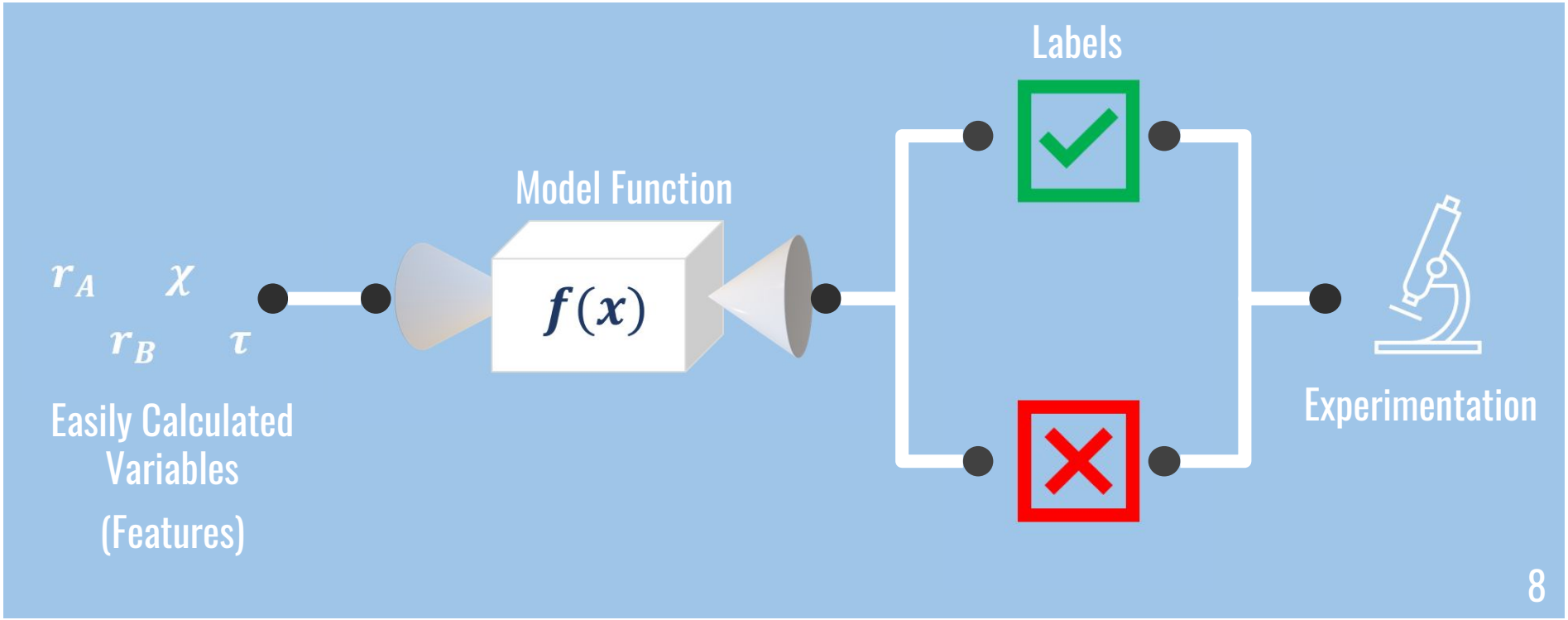
What combinations of A, B, X, will form a perovskite crystal structure?

Classification in Machine Learning

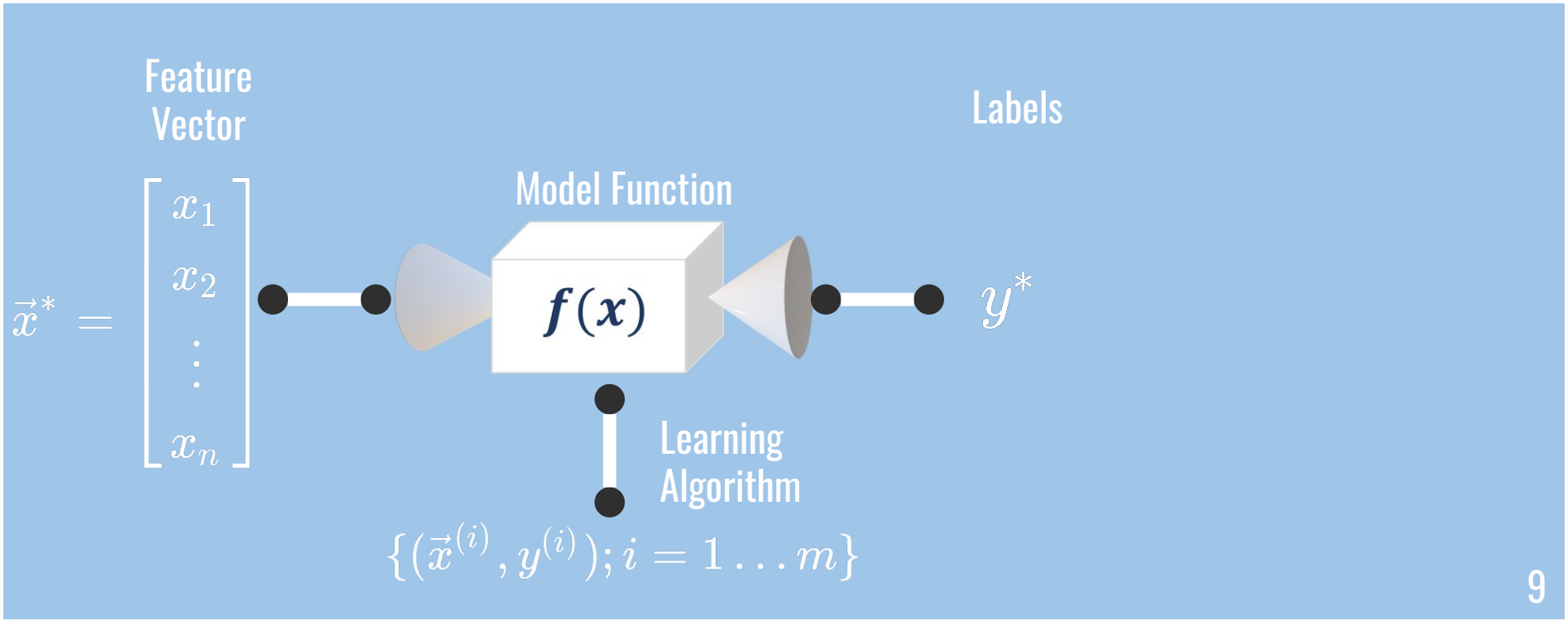
- In machine learning, classification refers to a predictive modeling problem of predicting the class of given data points.
- In mathematical language, the task is to approximate a model function f(**x**) that maps the input variables **x** (features) to discrete output variables y (labels).
- Model parameters of ML govern the position of decision boundaries, and optimal parameters are selected by evaluating the model performance (or prediction accuracy) on unseen data.

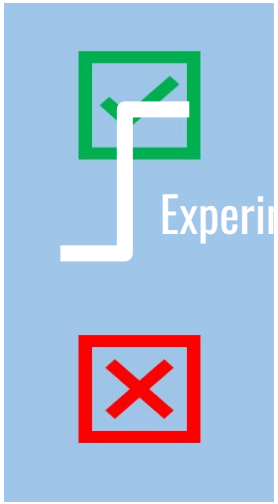
Vanderplas, J. T. *Python data science handbook: essential tools for working with data*; O'Reilly: Sebastopol, **2017**.

Workflow


Labels
Model Function
r_A χ
r_B τ
$f(x)$
Easily Calculated Variables
(Features)
Experimentation

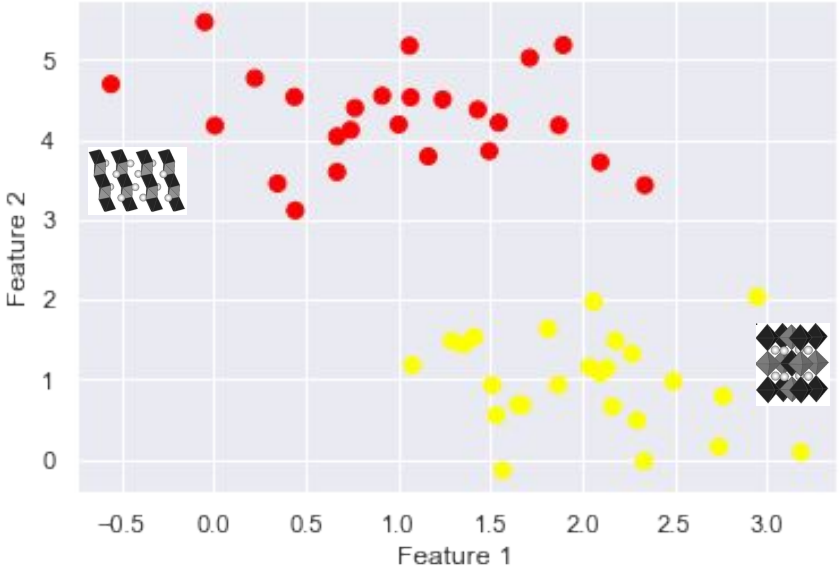
Workflow

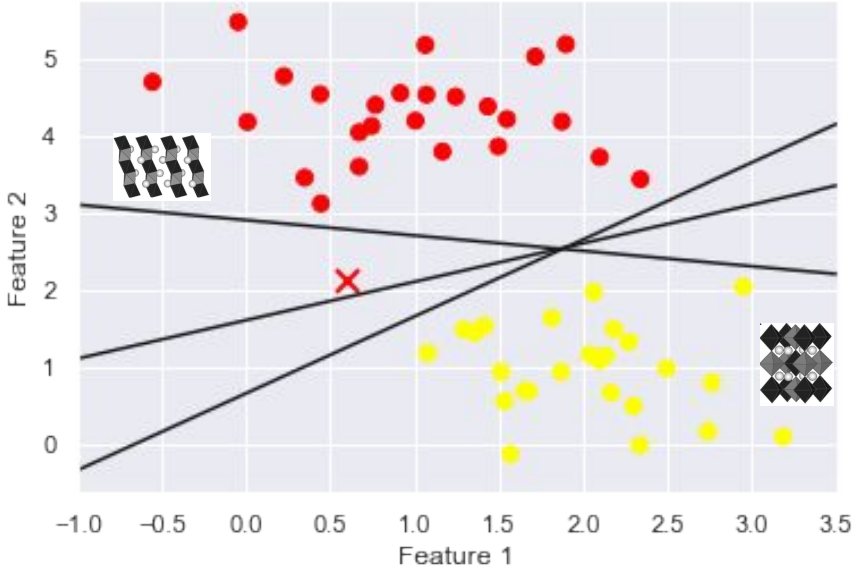

Feature Vector
Labels
Model Function
$\vec{x}^* = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$
$f(x)$
y^*
Learning Algorithm
$\{(\vec{x}^{(i)}, y^{(i)}); i = 1 \ldots m\}$



Vanderplas, J. T. *Python data science handbook: essential tools for working with data*; O'Reilly: Sebastopol, **2017**.

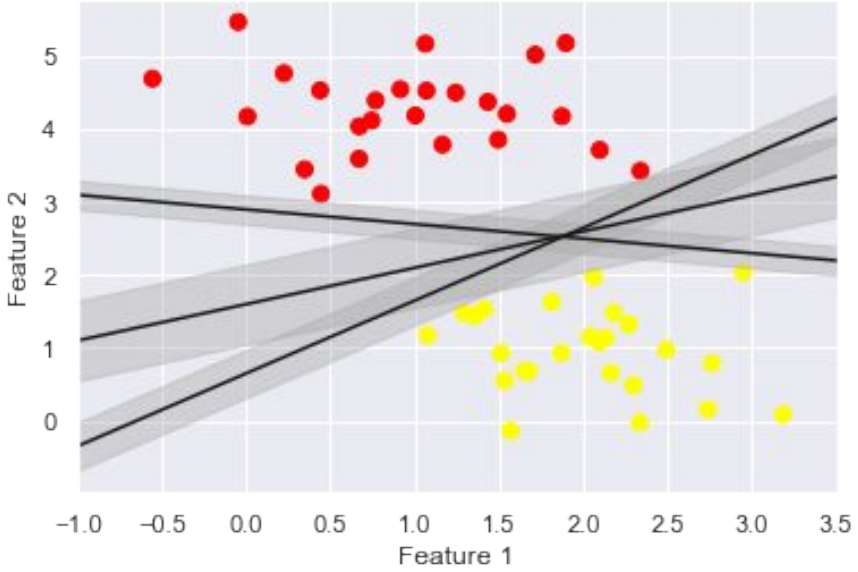
Support Vector Machine (SVM)

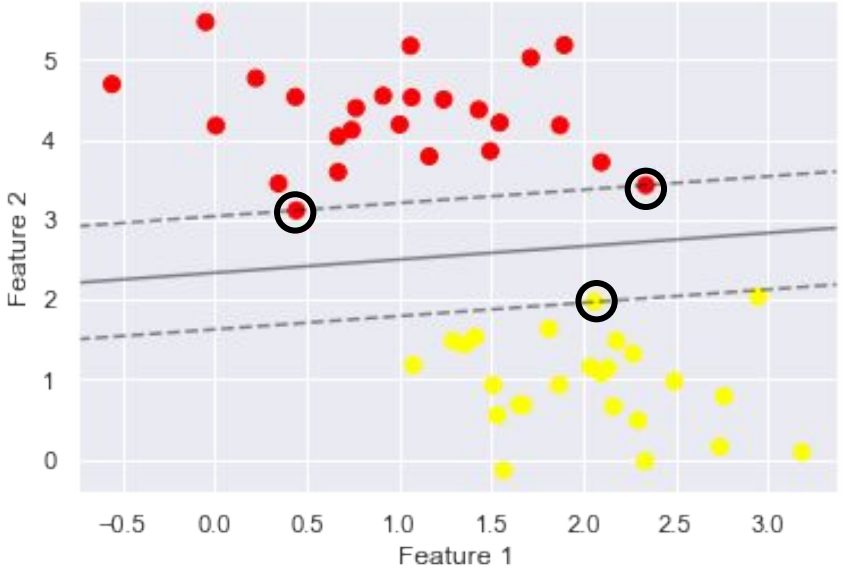




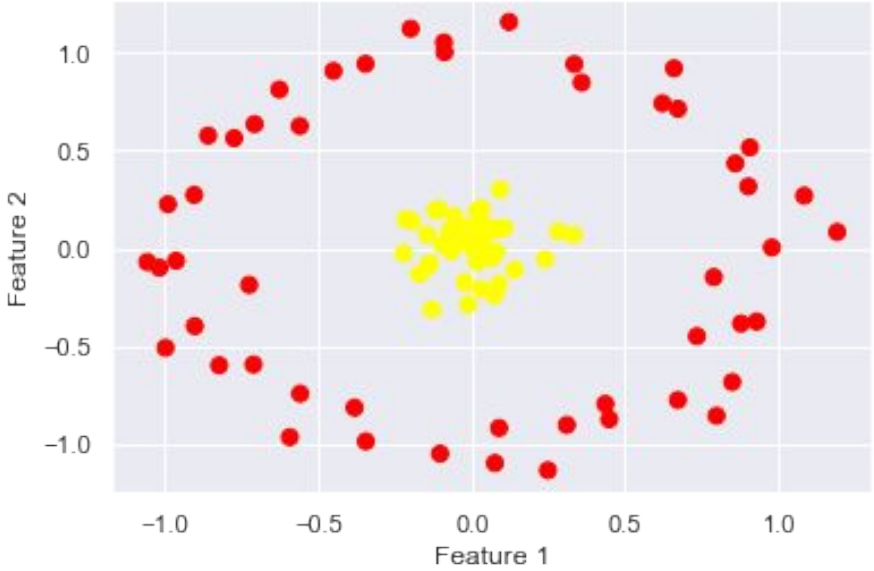
Vanderplas, J. T. *Python data science handbook: essential tools for working with data*; O'Reilly: Sebastopol, **2017**.

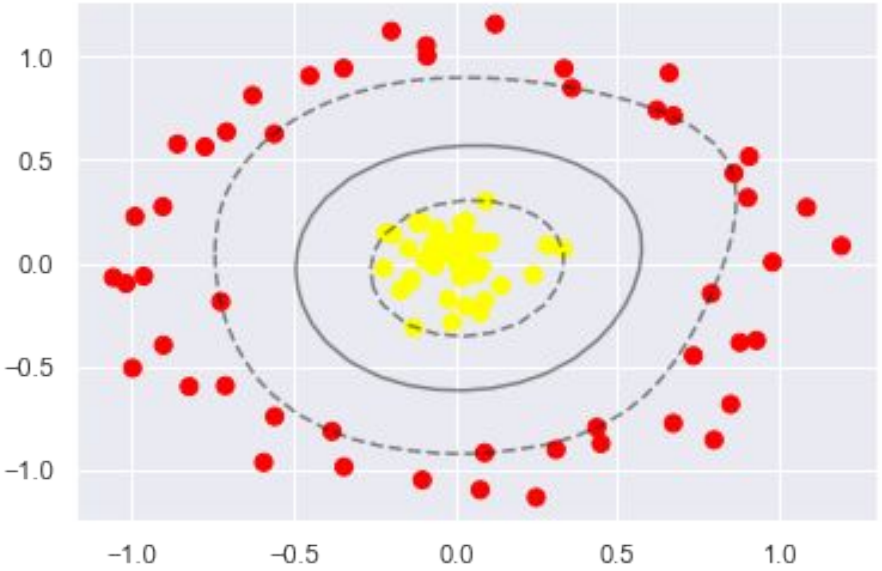
Support Vector Machine (SVM) - Linear



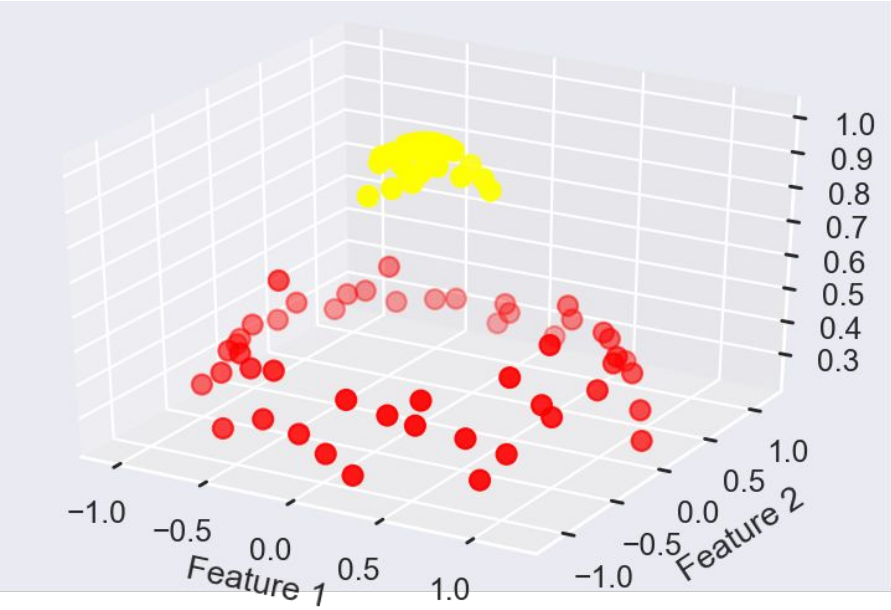


Support Vector Machine (SVM) - Non-Linear





Support Vector Machine (SVM) - Kernel Trick



Radial Basis Function (RBF):

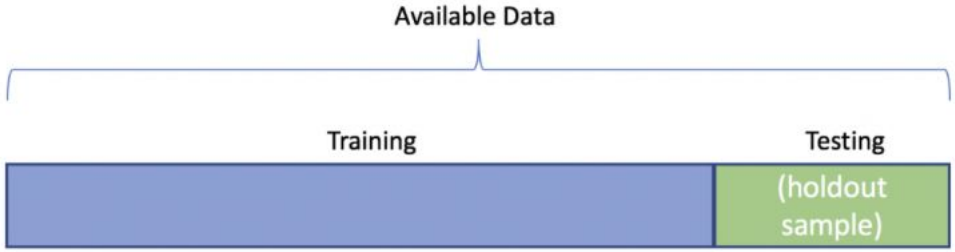
$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma|\mathbf{x}_i - \mathbf{x}_j|^2\right)$$

$\mathbf{x}_i$ --- Feature vector

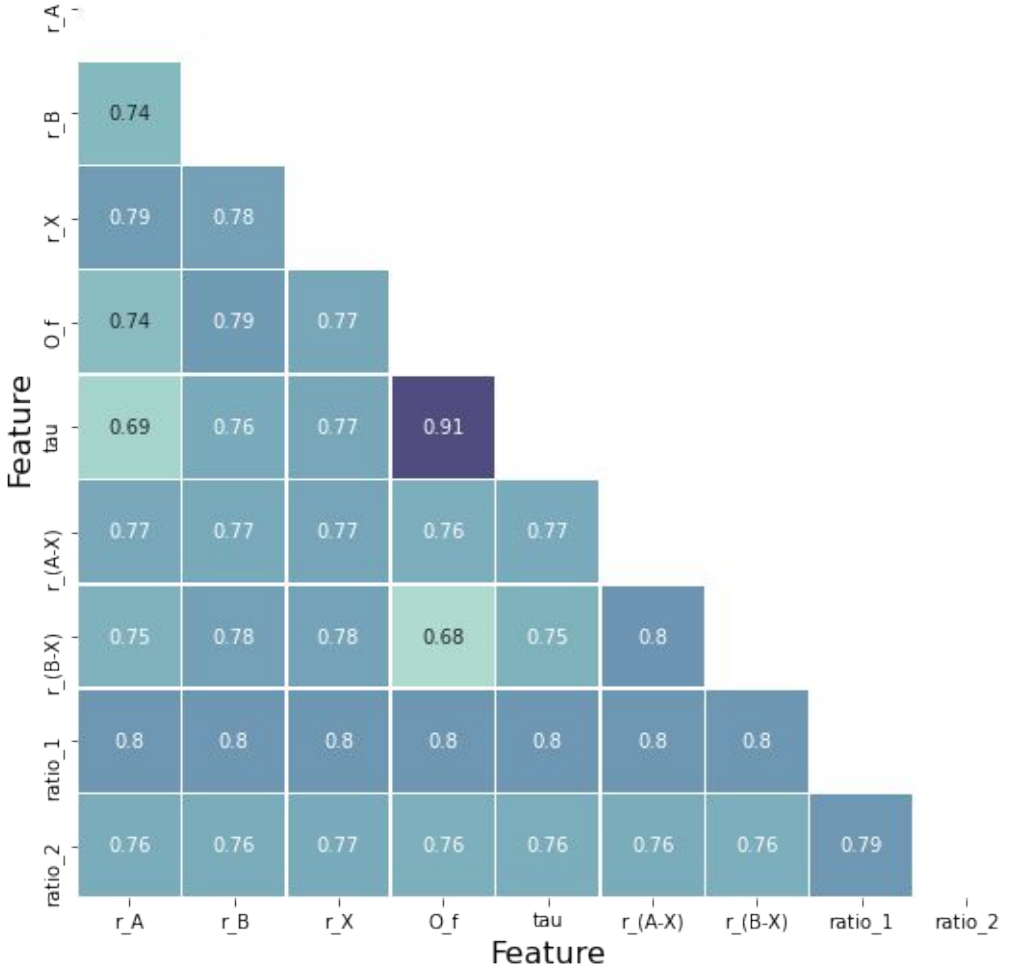
$$\gamma = \frac{1}{2\sigma^2}$$

Training Sets vs Test Sets

- Data
  - Contain the parameters and results given by experiments
- Training Set
  - Used to construct and optimize the prediction model
- Test Set
  - Used to evaluate the accuracy of the model



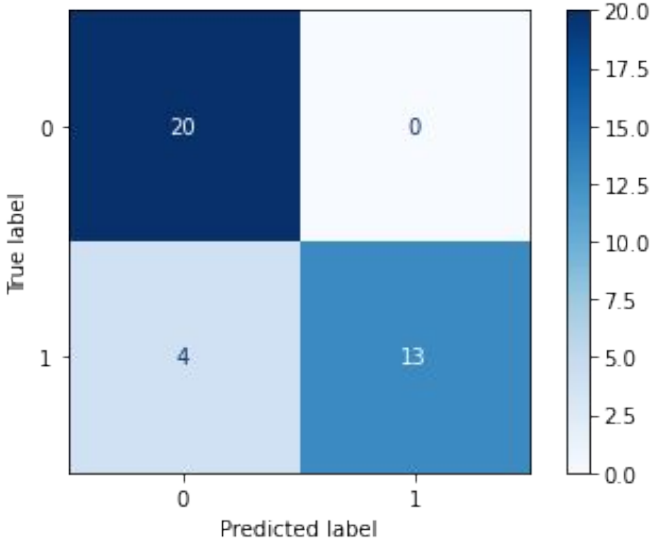
https://algotrading101.com/learn/train-test-split/



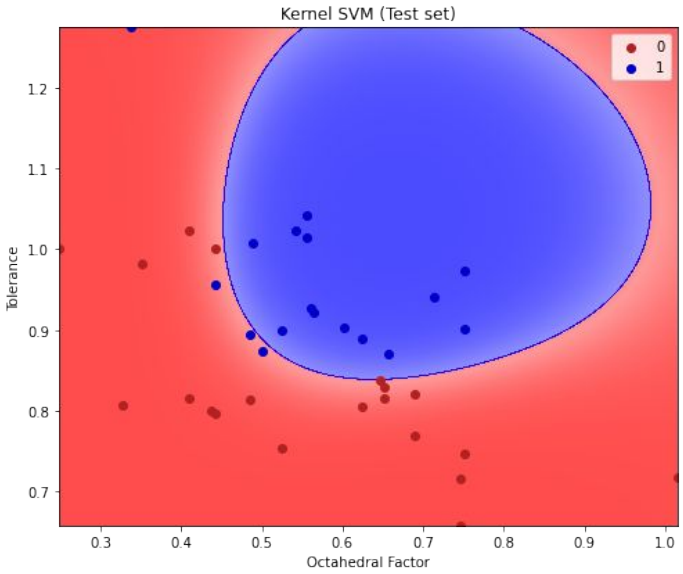
Feature Combination & Accuracy

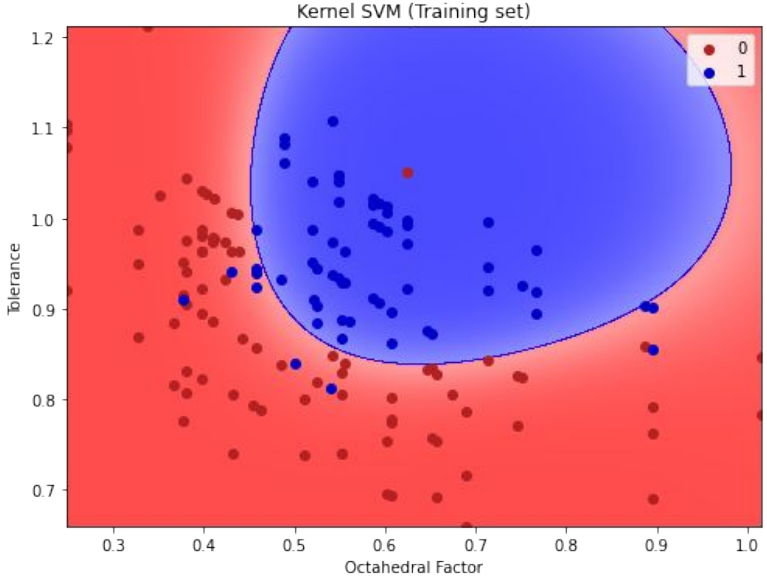
Confusion Matrix of

- Diagonal quadrants: number of correctly classified data
- Off-diagonal quadrants: number of incorrectly classified data
- Classifies all non-Perovskite compounds correctly
- Classified 13/17 Perovskite compounds correctly

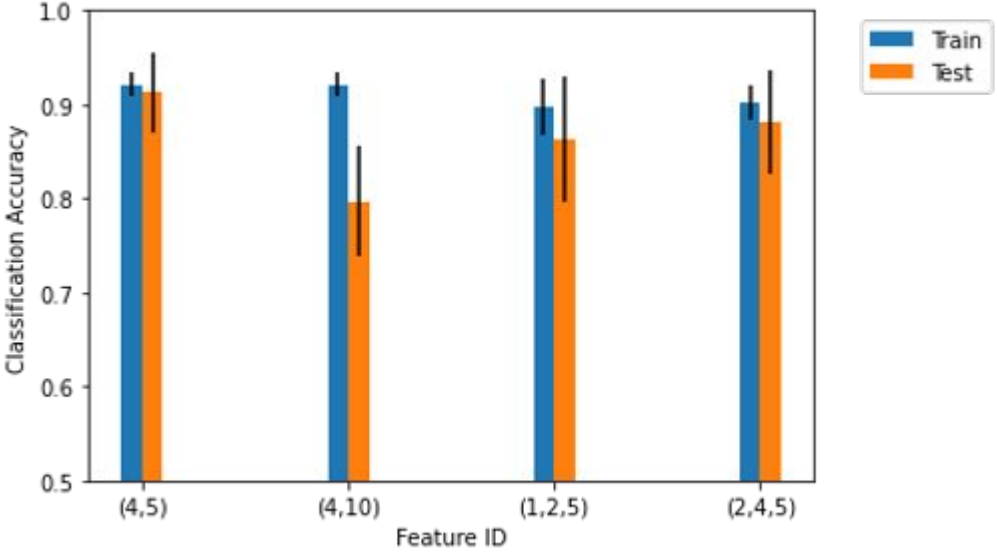


Contour of Tolerance & Octahedral Factor





Best Performing Features



1: Ionic Radii of Large Metal (r_A)
2: Ionic Radii of Small Metal (r_B)
4: Tolerance (τ)
5: Octahedral Factor ($O_h = r_B/r_X$)
10: Ratio of the sum of the s and p orbital radii r^{s+p}_A / r^{s+p}_X

Misclassifications

Most of the misclassifications are perovskite classified as non-perovskite

- Occur at the boundary
  - Predicted probabilities close to 50%
  - Non-perovskite can be synthesized in a long-lived metastable perovskite phase through non-equilibrium high pressure synthesis route.

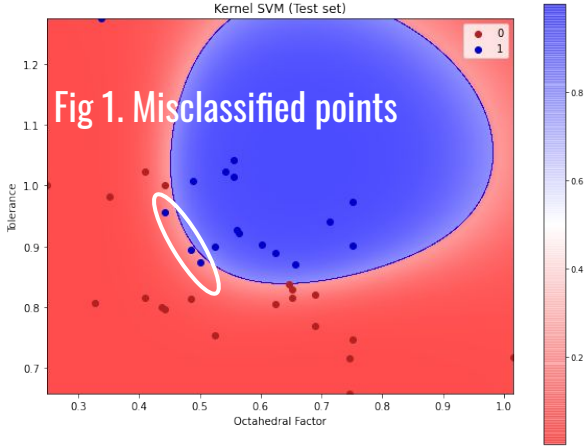
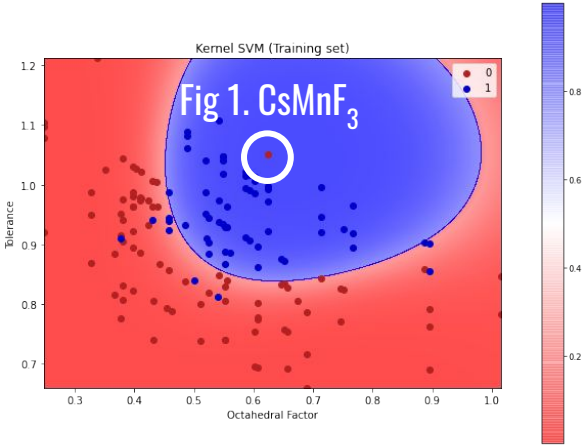


Fig 1. Misclassified points

Pilania G., Balachandran P., Kim C. and Lookman T. *Frontiers in material*, **2016**, Finding new perovskite halides via machine learning.3,19

Misclassifications

- Non-boundary misclassifications
  - $CsMnF_3$ non-perovskite as perovskite
  - Hexagonal antiferromagnetic structure of $CsMnF_3$ can be easily transformed to cubic perovskite at high pressure



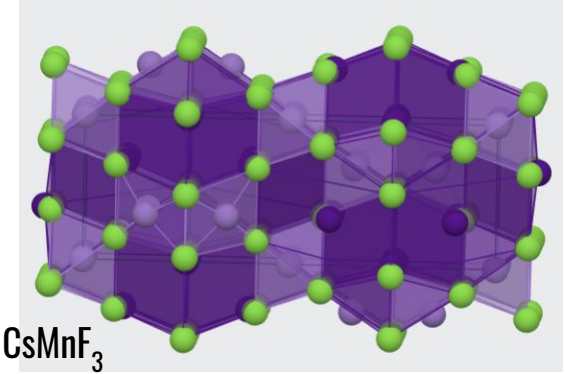


Fig.2 $CsMnF_3$

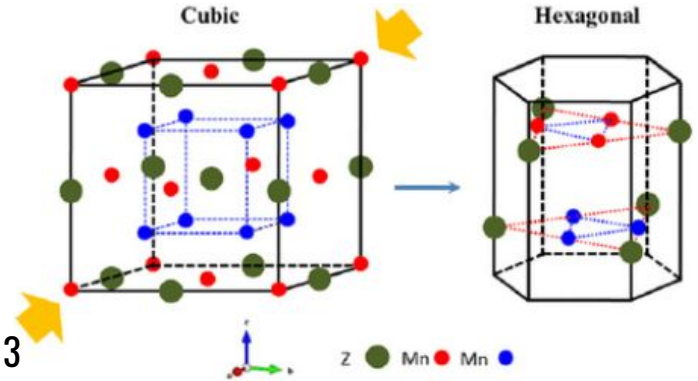


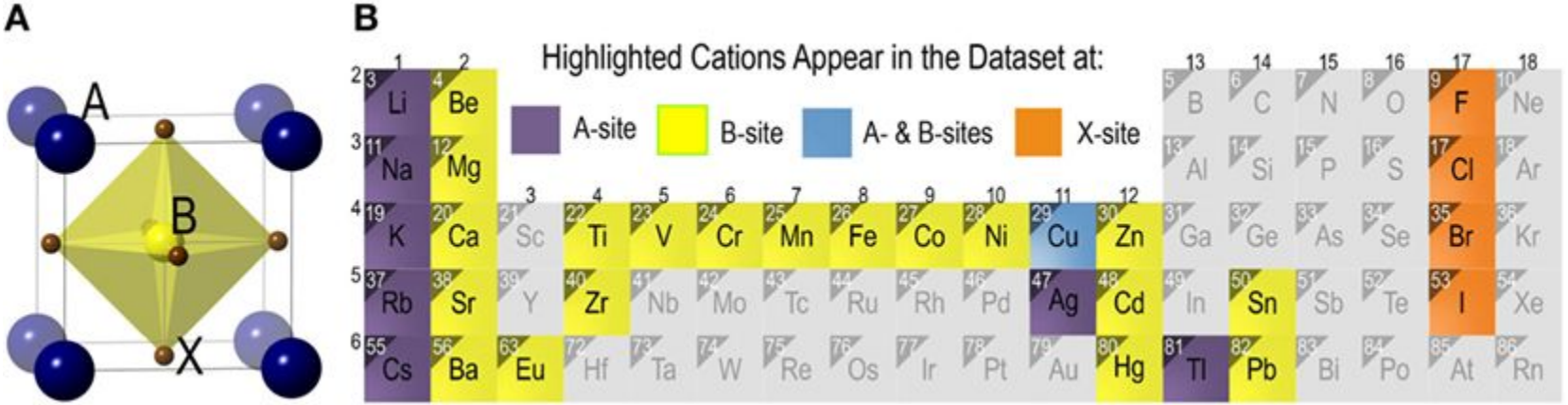
Fig. 3

Kafalas, J. and Longo J. *Solid state Chem* **1972** High pressure synthesis of (ABX3) (AX) n compounds. 4, 55

Zhang D., Yan B., Wu S., Kubler J., Kreiner G.,Parkin S and Felser C. *Journal of physics: Condensed matter*, **2013** First-principles study of the structural stability of cubic, tetragonal and hexagonal phases in Mn3z (Z = Ga, Sn and Ge) Heusler compounds, 25,206006

Implication

- Small dataset
  - 640 possible compositions
  - 185 know data
    - 8 A-site, 20 B-site, 4 X-site



Pilania G., Balachandran P., Kim C. and Lookman T. *Frontiers in material*, **2016**, Finding new perovskite halides via machine learning.3,19

New tolerance, τ, factor

$$\tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$

n_A : oxidation state of A
R_A : ionic radius of ion A
R_B : ionic radius of ion B
- $r_A > r_B$ by definition

Bartel C., Sutton C., et al. *Materials science* **2019** New tolerance factor to predict the stability of perovskite oxides and halides. 5,2, eaav0693

New tolerance, τ, factor

$$\tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$

Why it's better?

- Better sum of ionic radii estimates the interatomic bond distances for the structure
- Refined input radii and increased the dimensionality of the descriptor
- Monotonic dependency

Bartel C., Sutton C., et al. *Materials science* **2019** New tolerance factor to predict the stability of perovskite oxides and halides. 5,2, eaav0693

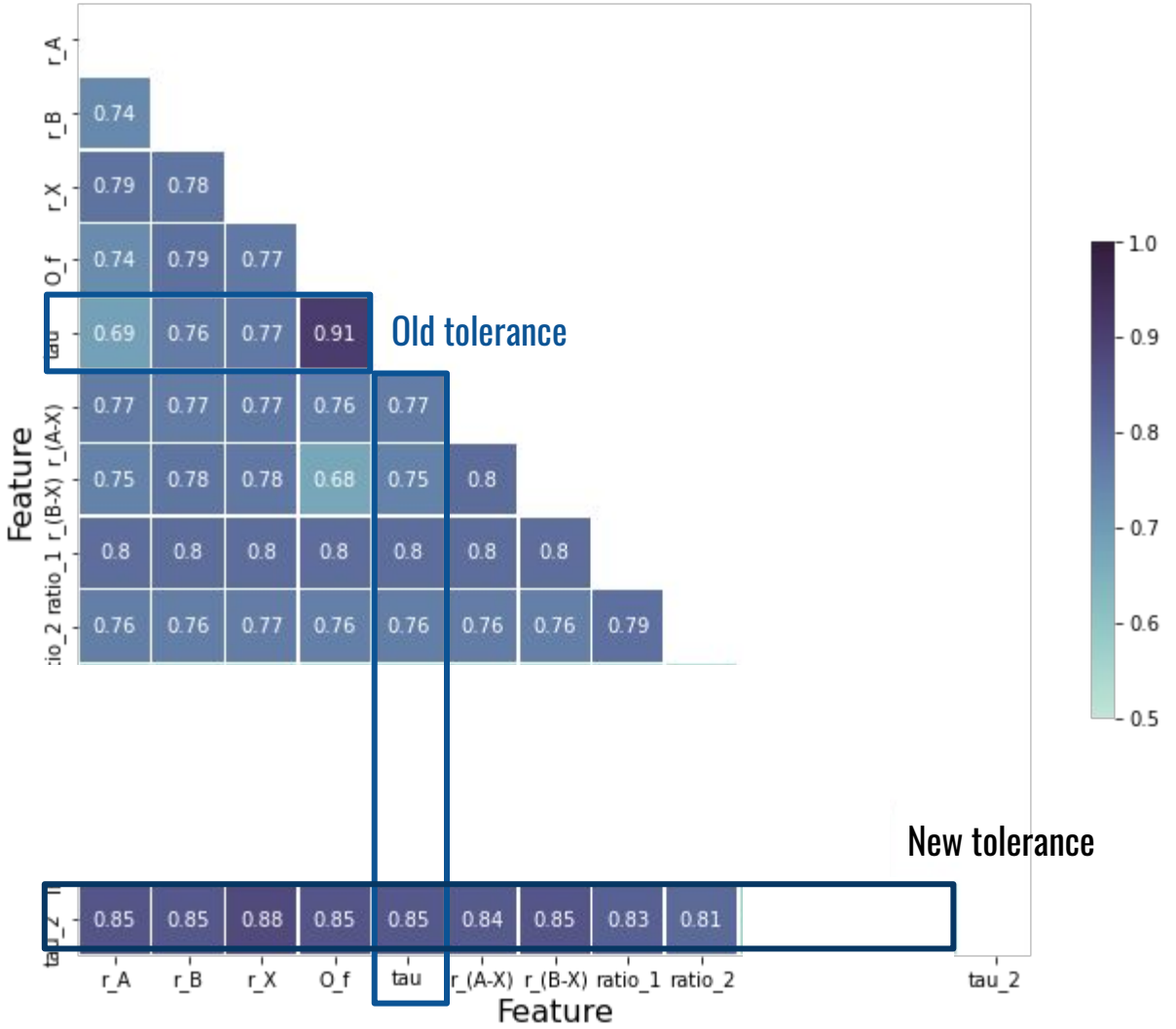
New tolerance, τ, factor

$$\tau = \frac{r_X}{r_B} - n_A\left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)}\right)$$

How it's better?

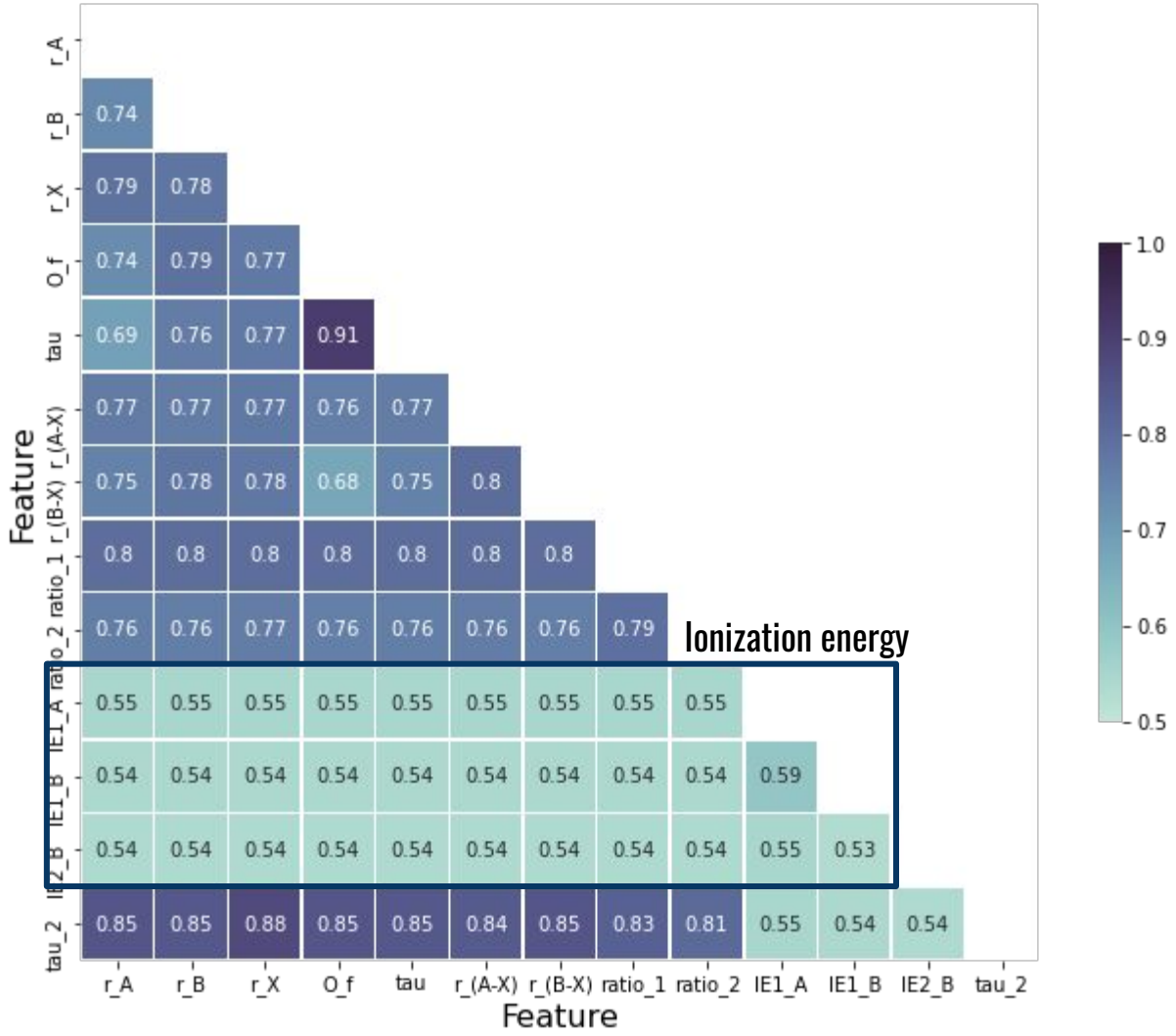
- Reduce false-positive prediction occurrence
- Increase the overall classification accuracy

Bartel C., Sutton C., et al. *Materials science* **2019** New tolerance factor to predict the stability of perovskite oxides and halides. 5,2, eaav0693



New Feature Combination & Accuracy

New tolerance



New Feature Combination & Accuracy

Ionization energy

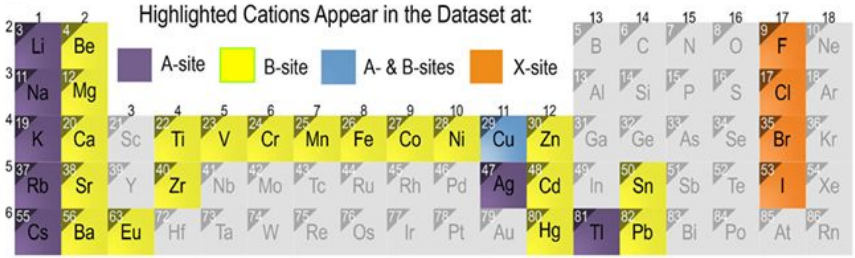
Limitations

- Data are highly repetitive
  - Only 4 different X-sites and a few A-sites and B-sites
- Data are similar
  - Close in periodic table, has similar structure and properties

No	System	Formabili	Assignme	1. Ra	2. Rb	3. Rx	4. t	5. of	miu	Ax	Bx	Xx	E(A-X)	8.(Ra
1	LiF–MgF2	No	0	1.13	0.72	1.33	0.849	0.54135	0.541	0.9	1.31	3.78	-2.88	-2.0
2	LiF–ZnF2	No	0	1.13	0.74	1.33	0.84	0.55639	0.556	0.9	1.44	3.78	-2.88	-1.9
3	LiF–MnF2	No	0	1.13	0.83	1.33	0.805	0.62406	0.624	0.9	2.04	3.78	-2.88	-1.4
4	LiF–CaF2	No	0	1.13	1	1.33	0.747	0.75188	0.752	0.9	1.17	3.78	-2.88	-2.2
5	LiF–PbF2	No	0	1.13	1.19	1.33	0.69	0.89474	0.895	0.9	1.92	3.78	-2.88	-1.5
6	NaF–NiF2	Yes	1	1.39	0.69	1.33	0.952	0.5188	0.519	0.89	1.76	3.78	-2.89	-2.1
7	NaF–MgF2	Yes	1	1.39	0.72	1.33	0.938	0.54135	0.541	0.89	1.31	3.78	-2.89	-2.5
8	NaF–CuF2	Yes	1	1.39	0.73	1.33	0.934	0.54887	0.549	0.89	1.08	3.78	-2.89	-2.8
9	NaF–ZnF2	Yes	1	1.39	0.74	1.33	0.929	0.55639	0.556	0.89	1.44	3.78	-2.89	-2.4
10	NaF–CoF2	Yes	1	1.39	0.745	1.33	0.927	0.56015	0.56	0.89	1.72	3.78	-2.89	-2.1
11	NaF–FeF2	Yes	1	1.39	0.78	1.33	0.912	0.58647	0.586	0.89	1.67	3.78	-2.89	-2.2
12	NaF–VF2	Yes	1	1.39	0.79	1.33	0.907	0.59398	0.594	0.89	2.22	3.78	-2.89	-1.6
13	NaF–CrF2	Yes	1	1.39	0.8	1.33	0.903	0.6015	0.602	0.89	2	3.78	-2.89	-1.8
14	NaF–MnF2	Yes	1	1.39	0.83	1.33	0.89	0.62406	0.624	0.89	2.04	3.78	-2.89	-1.8
15	NaF–CdF2	No	0	1.39	0.95	1.33	0.844	0.71429	0.714	0.89	1.4	3.78	-2.89	-2.4
16	NaF–CaF2	No	0	1.39	1	1.33	0.825	0.75188	0.752	0.89	1.17	3.78	-2.89	-2.7
17	NaF–PbF2	No	0	1.39	1.19	1.33	0.763	0.89474	0.895	0.89	1.92	3.78	-2.89	-1.9
18	NaF–BaF2	No	0	1.39	1.35	1.33	0.718	1.01504	1.01	0.89	1.08	3.78	-2.89	-2.8
19	AgF–CoF2	Yes	1	1.49	0.65	1.33	1.007	0.48872	0.489	1.07	1.72	3.78	-2.71	-2.3
20	AgF–NiF2	Yes	1	1.49	0.69	1.33	0.987	0.5188	0.519	1.07	1.76	3.78	-2.71	-2
21	AgF–MgF2	Yes	1	1.49	0.72	1.33	0.973	0.54135	0.541	1.07	1.31	3.78	-2.71	-2.7
22	AgF–ZnF2	Yes	1	1.49	0.74	1.33	0.963	0.55639	0.556	1.07	1.44	3.78	-2.71	-2.6
23	AgF–MnF2	Yes	1	1.49	0.83	1.33	0.923	0.62406	0.624	1.07	2.04	3.78	-2.71	-1.9
24	AgF–PbF2	No	0	1.49	1.19	1.33	0.791	0.89474	0.895	1.07	1.92	3.78	-2.71	-2.0
25	KF–CoF2	Yes	1	1.64	0.65	1.33	1.061	0.48872	0.489	0.8	1.72	3.78	-2.98	-2.5
26	KF–NiF2	Yes	1	1.64	0.69	1.33	1.04	0.5188	0.519	0.8	1.76	3.78	-2.98	-2.4
27	KF–MgF2	Yes	1	1.64	0.72	1.33	1.024	0.54135	0.541	0.8	1.31	3.78	-2.98	-3.0
28	KF–CuF2	Yes	1	1.64	0.73	1.33	1.019	0.54887	0.549	0.8	1.08	3.78	-2.98	-3.3
29	KF–ZnF2	Yes	1	1.64	0.74	1.33	1.015	0.55639	0.556	0.8	1.44	3.78	-2.98	-2.8
30	KF–FeF2	Yes	1	1.64	0.78	1.33	0.995	0.58647	0.586	0.8	1.67	3.78	-2.98	-2.6
31	KF–VF2	Yes	1	1.64	0.79	1.33	0.991	0.59398	0.594	0.8	2.22	3.78	-2.98	-1.9
32	KF–CrF2	Yes	1	1.64	0.8	1.33	0.986	0.6015	0.602	0.8	2	3.78	-2.98	-2.1
33	KF–MnF2	Yes	1	1.64	0.83	1.33	0.972	0.62406	0.624	0.8	2.04	3.78	-2.98	-2.1
34	KF–CdF2	Yes	1	1.64	0.95	1.33	0.921	0.71429	0.714	0.8	1.4	3.78	-2.98	-2.9
35	KF–CaF2	Yes	1	1.64	1	1.33	0.901	0.75188	0.752	0.8	1.17	3.78	-2.98	-3.2
36	KF–HgF2	Yes	1	1.64	1.02	1.33	0.894	0.76692	0.767	0.8	1.49	3.78	-2.98	-2.8
37	KF–BaF2	No	0	1.64	1.35	1.33	0.784	1.01504	1.015	0.8	1.08	3.78	-2.98	-3.3

Limitations

- Data are highly repetitive
  - Only 4 different X-sites and a few A-sites and B-sites
- Data are similar
  - Close in periodic table, has similar structure and properties

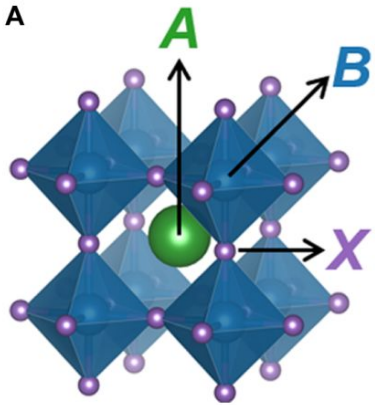


No	System	Formabili	Assignme	1. Ra	2. Rb	3. Rx	4. t	5. of	miu	Ax	Bx	Xx	E(A-X)	8.(Ra
1	LiF–MgF2	No	0	1.13	0.72	1.33	0.849	0.54135	0.541	0.9	1.31	3.78	-2.88	-2.0
2	LiF–ZnF2	No	0	1.13	0.74	1.33	0.84	0.55639	0.556	0.9	1.44	3.78	-2.88	-1.9
3	LiF–MnF2	No	0	1.13	0.83	1.33	0.805	0.62406	0.624	0.9	2.04	3.78	-2.88	-1.4
4	LiF–CaF2	No	0	1.13	1	1.33	0.747	0.75188	0.752	0.9	1.17	3.78	-2.88	-2.2
5	LiF–PbF2	No	0	1.13	1.19	1.33	0.69	0.89474	0.895	0.9	1.92	3.78	-2.88	-1.5
6	NaF–NiF2	Yes	1	1.39	0.69	1.33	0.952	0.5188	0.519	0.89	1.76	3.78	-2.89	-2.1
7	NaF–MgF2	Yes	1	1.39	0.72	1.33	0.938	0.54135	0.541	0.89	1.31	3.78	-2.89	-2.5
8	NaF–CuF2	Yes	1	1.39	0.73	1.33	0.934	0.54887	0.549	0.89	1.08	3.78	-2.89	-2.8
9	NaF–ZnF2	Yes	1	1.39	0.74	1.33	0.929	0.55639	0.556	0.89	1.44	3.78	-2.89	-2.4
10	NaF–CoF2	Yes	1	1.39	0.745	1.33	0.927	0.56015	0.56	0.89	1.72	3.78	-2.89	-2.1
11	NaF–FeF2	Yes	1	1.39	0.78	1.33	0.912	0.58647	0.586	0.89	1.67	3.78	-2.89	-2.2
12	NaF–VF2	Yes	1	1.39	0.79	1.33	0.907	0.59398	0.594	0.89	2.22	3.78	-2.89	-1.6
13	NaF–CrF2	Yes	1	1.39	0.8	1.33	0.903	0.6015	0.602	0.89	2	3.78	-2.89	-1.8
14	NaF–MnF2	Yes	1	1.39	0.83	1.33	0.89	0.62406	0.624	0.89	2.04	3.78	-2.89	-1.8
15	NaF–CdF2	No	0	1.39	0.95	1.33	0.844	0.71429	0.714	0.89	1.4	3.78	-2.89	-2.4
16	NaF–CaF2	No	0	1.39	1	1.33	0.825	0.75188	0.752	0.89	1.17	3.78	-2.89	-2.7
17	NaF–PbF2	No	0	1.39	1.19	1.33	0.763	0.89474	0.895	0.89	1.92	3.78	-2.89	-1.9
18	NaF–BaF2	No	0	1.39	1.35	1.33	0.718	1.01504	1.01	0.89	1.08	3.78	-2.89	-2.8
19	AgF–CoF2	Yes	1	1.49	0.65	1.33	1.007	0.48872	0.489	1.07	1.72	3.78	-2.71	-2.3
20	AgF–NiF2	Yes	1	1.49	0.69	1.33	0.987	0.5188	0.519	1.07	1.76	3.78	-2.71	-2
21	AgF–MgF2	Yes	1	1.49	0.72	1.33	0.973	0.54135	0.541	1.07	1.31	3.78	-2.71	-2.7
22	AgF–ZnF2	Yes	1	1.49	0.74	1.33	0.963	0.55639	0.556	1.07	1.44	3.78	-2.71	-2.6
23	AgF–MnF2	Yes	1	1.49	0.83	1.33	0.923	0.62406	0.624	1.07	2.04	3.78	-2.71	-1.9
24	AgF–PbF2	No	0	1.49	1.19	1.33	0.791	0.89474	0.895	1.07	1.92	3.78	-2.71	-2.0
25	KF–CoF2	Yes	1	1.64	0.65	1.33	1.061	0.48872	0.489	0.8	1.72	3.78	-2.98	-2.5
26	KF–NiF2	Yes	1	1.64	0.69	1.33	1.04	0.5188	0.519	0.8	1.76	3.78	-2.98	-2.4
27	KF–MgF2	Yes	1	1.64	0.72	1.33	1.024	0.54135	0.541	0.8	1.31	3.78	-2.98	-3.0
28	KF–CuF2	Yes	1	1.64	0.73	1.33	1.019	0.54887	0.549	0.8	1.08	3.78	-2.98	-3.3
29	KF–ZnF2	Yes	1	1.64	0.74	1.33	1.015	0.55639	0.556	0.8	1.44	3.78	-2.98	-2.8
30	KF–FeF2	Yes	1	1.64	0.78	1.33	0.995	0.58647	0.586	0.8	1.67	3.78	-2.98	-2.6
31	KF–VF2	Yes	1	1.64	0.79	1.33	0.991	0.59398	0.594	0.8	2.22	3.78	-2.98	-1.9
32	KF–CrF2	Yes	1	1.64	0.8	1.33	0.986	0.6015	0.602	0.8	2	3.78	-2.98	-2.1
33	KF–MnF2	Yes	1	1.64	0.83	1.33	0.972	0.62406	0.624	0.8	2.04	3.78	-2.98	-2.1
34	KF–CdF2	Yes	1	1.64	0.95	1.33	0.921	0.71429	0.714	0.8	1.4	3.78	-2.98	-2.9
35	KF–CaF2	Yes	1	1.64	1	1.33	0.901	0.75188	0.752	0.8	1.17	3.78	-2.98	-3.2
36	KF–HgF2	Yes	1	1.64	1.02	1.33	0.894	0.76692	0.767	0.8	1.49	3.78	-2.98	-2.8
37	KF–BaF2	No	0	1.64	1.35	1.33	0.784	1.01504	1.015	0.8	1.08	3.78	-2.98	-3.3

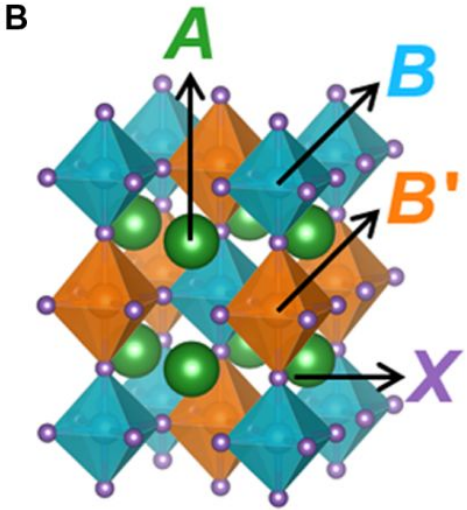
Conclusion

- Best performed features
  - Tolerance
  - Octahedral factor

- Future study
  - Double perovskite



perovskite



Double perovskite

Bartel C., Sutton C., et al. *Materials science* **2019** New tolerance factor to predict the stability of perovskite oxides and halides. 5,2, eaav0693

Thank you.

Questions?