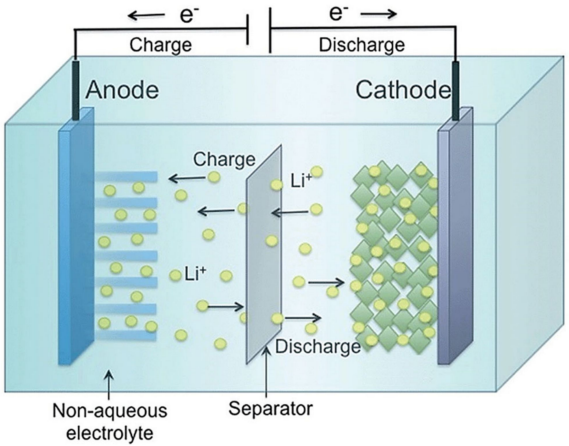
Predicting Crystal System of Cathodes in Lithium-Ion Batteries Using Machine Learning

Tyler Fu
PRISMS

Introduction

Lithium ion batteries

- Lithium ion batteries are widely used around the world
- Lithium ions travel from the anode to the cathode through electrolyte, current flows from cathode to anode through wire

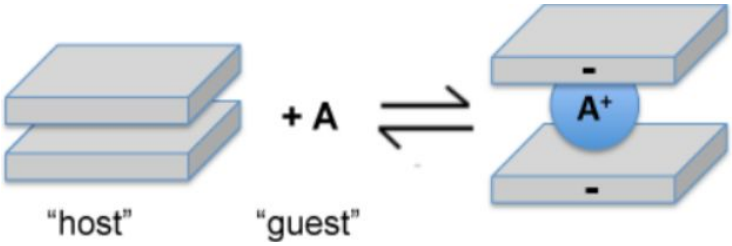


Source: https://doi.org/10.3390/en13195117

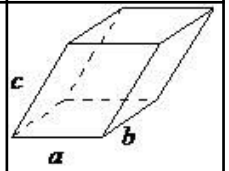
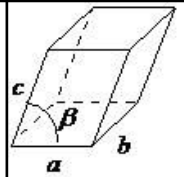
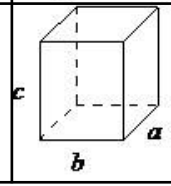
Background information

Cathodes

- Cathodes are crystal lattices that intercalate lithium ions
- Many cathode crystals are used
- Three main crystal systems: monoclinic, triclinic, orthorhombic



Source: Chem libreTexts

System	Lengths and Angles	Unit Cell Shape
Triclinic	a≠b≠c and ∠bc≠∠ca≠∠ab	
monoclinic	a≠b≠c, ∠bc=∠ab = 90°, and ∠ca ≠ 90°	
orthorhombic	a≠b≠c and ∠bc=∠ca=∠ab = 90°	

Purpose

- Knowing the crystal structure of a cathode material will help predict its properties
- The purpose is to predict the crystal system of a cathode material

Data

- Dataset from the Materials Project
- 339 data points, 80% for training, 20% for testing

Example data

Material Id	Formula	Space group	Formation Energy (eV)	E Above Hull (eV)	Band Gap (eV)	Nsites	Density (gm/cc)	Volume	Has Bandstructure	Crystal System
mp-849394	Li2MnSiO4	Pc	-2.699	0.006	3.462	16	2.993	178.513	TRUE	monoclinic
mp-762762	LiFe2(SiO4)2	P1	-2.426	0.114	0	39	2.753	547.911	FALSE	triclinic
mp-762828	LiMnSiO4	Pna21	-2.623	0.054	0.11	84	3.55	864.216	FALSE	orthorhombic
mp-566680	Li2MnSiO4	P21nm	-2.705	0	3.052	16	3.039	175.842	TRUE	orthorhombic
mp-767709	Li2Mn3Si3O10	C2/c	-2.747	0.016	2.578	36	3.334	421.286	TRUE	monoclinic

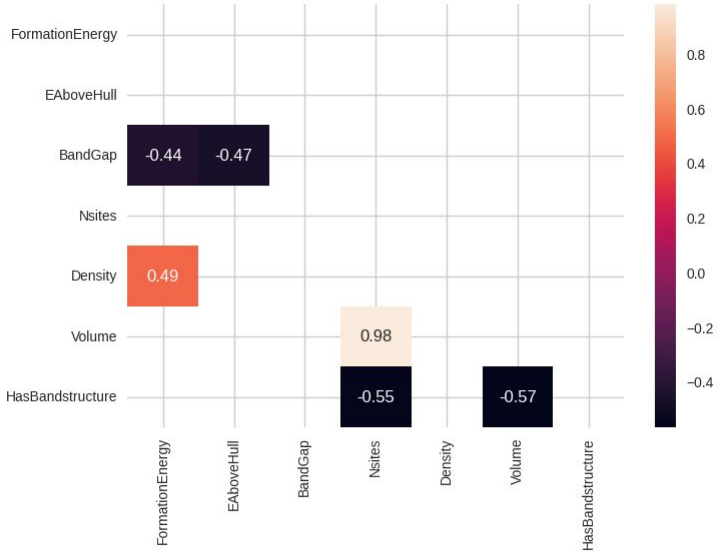
Exploratory data analysis

- Analysis of numerical data

Statistically significant values

	count	mean	std	min	25%	50%	75%	max
Formation Energy (eV)	339.0	-2.616950	0.183809	-2.985	-2.7575	-2.605	-2.5255	-2.012
E Above Hull (eV)	339.0	0.058215	0.030363	0.000	0.0355	0.062	0.0815	0.190
Band Gap (eV)	339.0	2.079740	1.087968	0.000	1.2655	2.499	2.9680	3.823
Nsites	339.0	38.837758	23.133142	10.000	26.0000	31.000	52.0000	132.000
Density (gm/cc)	339.0	2.984003	0.353968	2.200	2.7605	2.947	3.1060	4.201
Volume	339.0	467.765619	292.674559	122.581	286.3815	358.537	601.6965	1518.850
Has Bandstructure	339.0	0.808260	0.394252	0.000	1.0000	1.000	1.0000	1.000

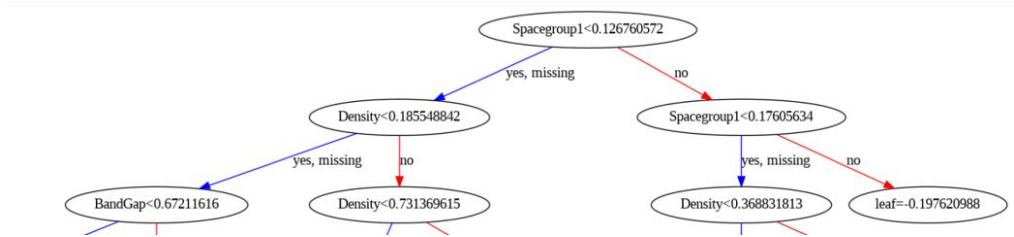
Correlation Heatmap



Machine Learning model

XGBoost classifier model

- Decision tree model with gradient boosting
- Known to be very accurate
- Good at handling smaller datasets



First three rows of decision tree

Preprocessing

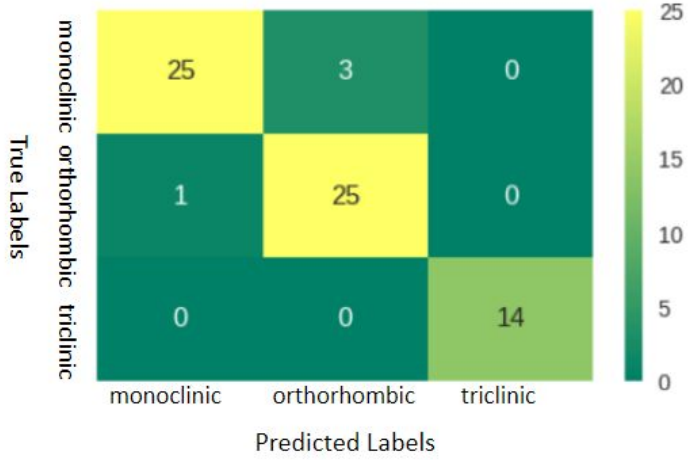
- Space group was encoded with frequency encoding
- chemical formula was encoded with Catboost encoder
- Data was normalized

Results

Classification Report

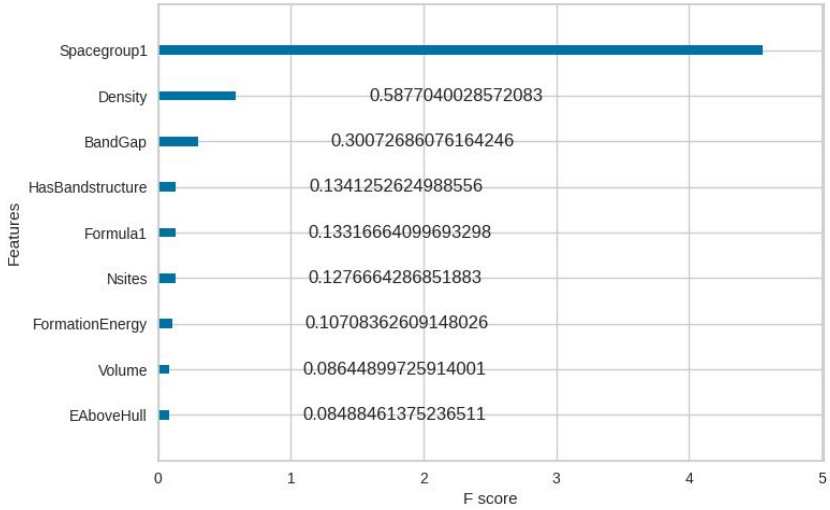
	precision	recall	f1-score	support
monoclinic (0)	0.96	0.89	0.93	28
orthorhombic (1)	0.89	0.96	0.93	26
triclinic (2)	1.00	1.00	1.00	14
accuracy			0.94	68
macro average/ unweighted average	0.95	0.95	0.95	68
weighted average	0.94	0.94	0.94	68

Confusion Matrix

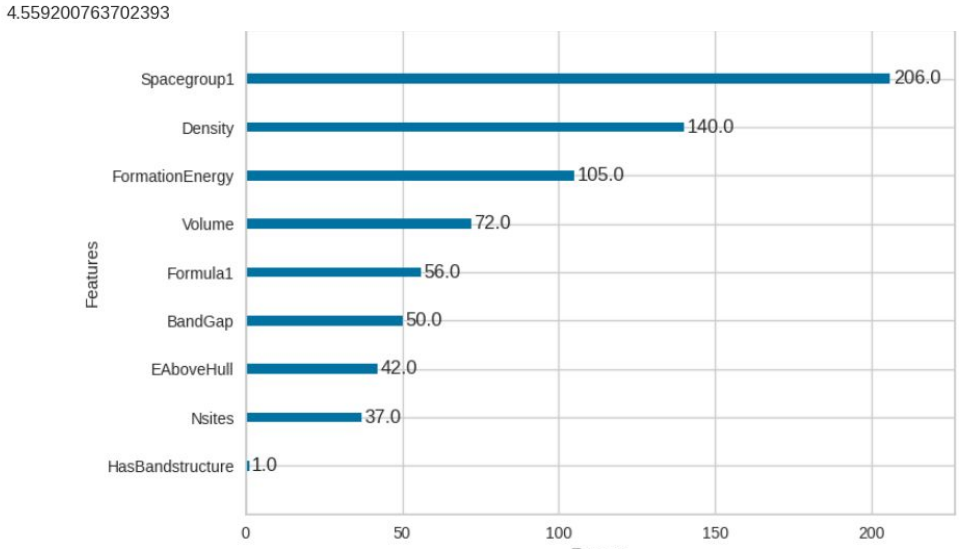


Results

Feature importance based on gain



Feature importance based on count



Conclusion

- The model was very effective with 94% accuracy
- 100% accuracy for triclinic
- Lower accuracy for other crystal systems

Limitations

- Not enough data to do cross-validation tests to prevent overfitting

Future goals

- Use more comprehensive data
- Investigate how much crystal system affects electrochemical properties of a material