

Ryan Carlin, Andrew Copsey, Corey Johnson, Michelle Katz

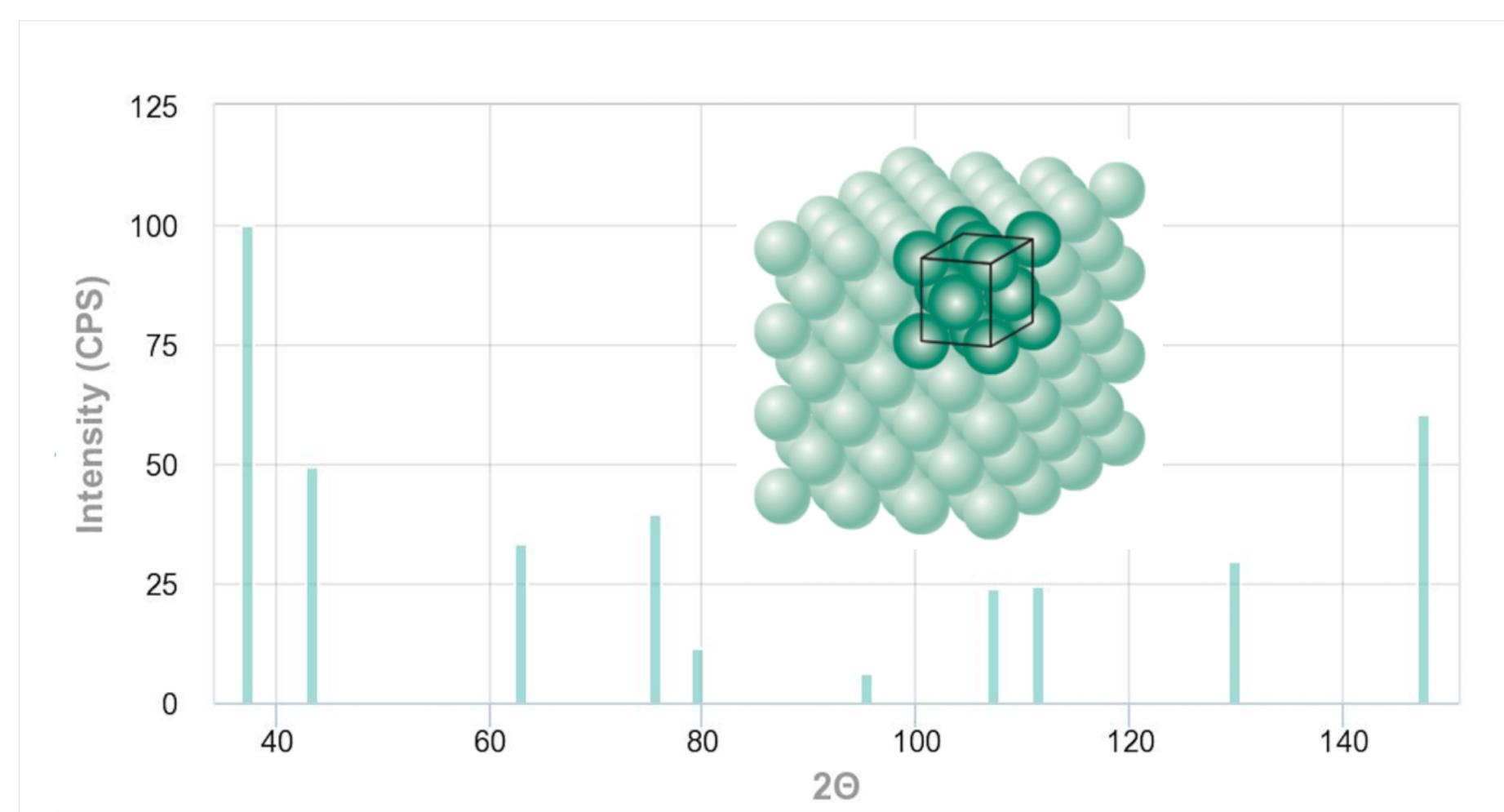
Materials Science and Engineering Department

## Introduction

### Motivation

X-ray diffraction (XRD) is a simple and commonly used method to determine a range of crystallographic properties including structure, defects, and grain sizes. Researchers are also often interested in a materials electronic or functional properties, however gathering enough measurements to calculate these properties can be arduous. xrados is a Python package developed with the intention of overcoming this challenge using machine learning to understand the relationship between XRD measurements and electronic properties.

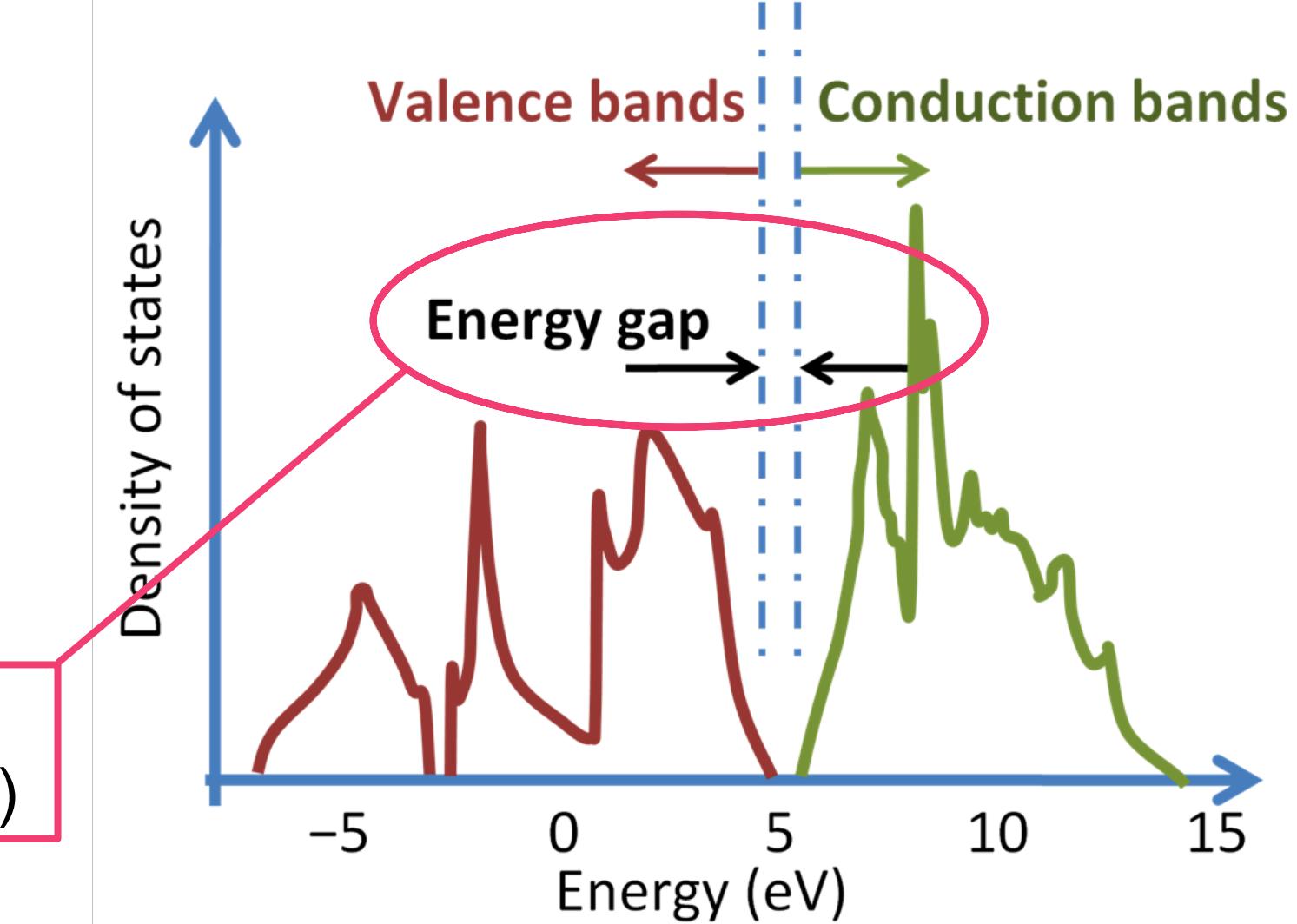
**INPUT**  
Crystal Structure & XRD Data



**OUTPUT**

Density of State

Initial Goal: Predict Bandgap Energy ( $E_g$ )



### Use Cases

Import data from MPD or use unprocessed XRD data

Train and optimize machine learning model

Quantify and visualize Density of State

## Methods

### Data Mining and Organization

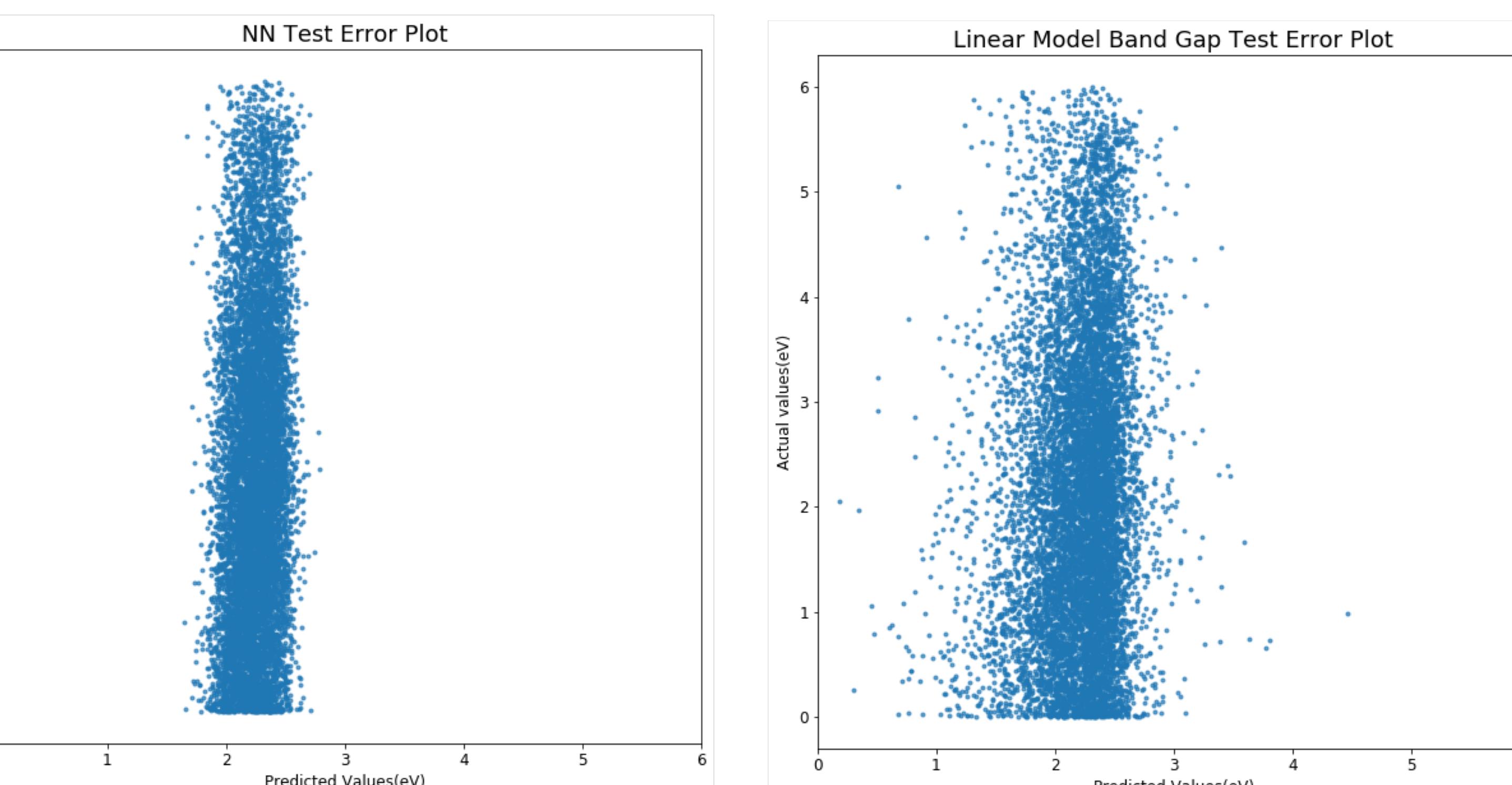
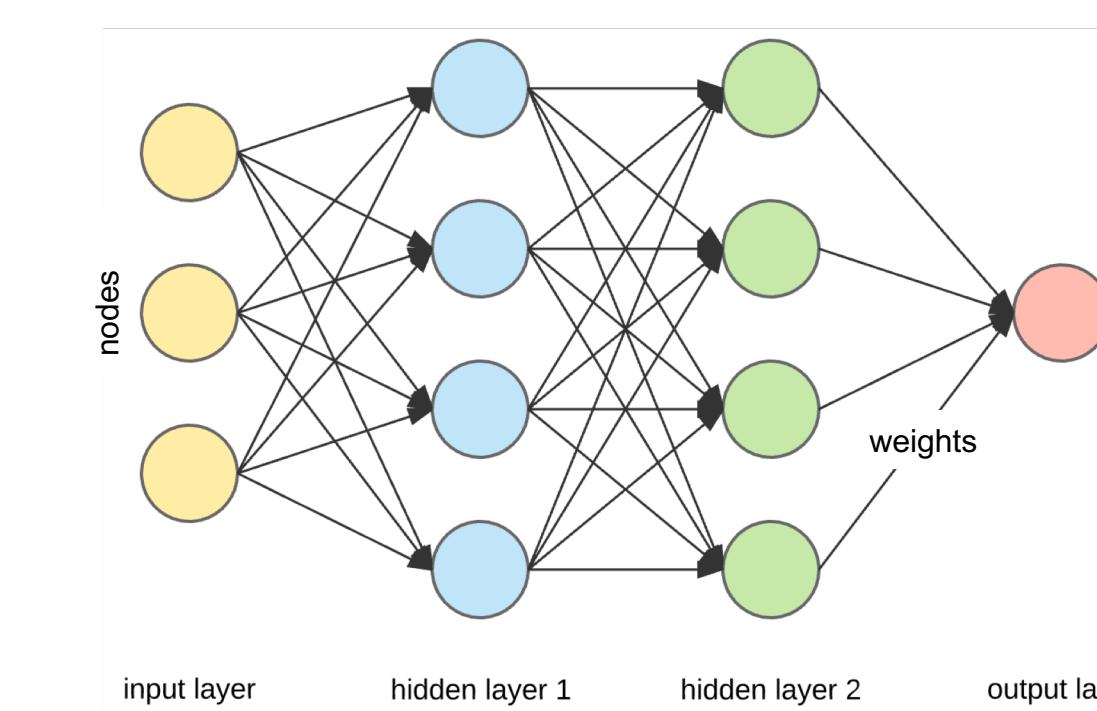
- MatMiner was used to pull data off the Materials Project Database (MPD)
- Modeled semiconductors with  $E_g$  between 0-6 eV
- Data organization after cleaning:  
 $N [ 10 \text{ highest intensity peaks} | \text{corresponding } 20 | E_g ]$

### Data Processing

- Added polynomial interaction terms in attempt to capture complexity of crystal structure (Degree 2)
- Data organization after addition of polynomials
- Data organization after interaction terms:  
 $N [ 10 \text{ intensities} | 20 \text{ values} | \text{interaction terms} | E_g ]$

### Models Used

- Neural Network
  - 4 hidden layers
  - Ridge regression regularizer
- Multiple Linear Regression
  - Coefficients for every interaction term



Both neural network (left) and linear regression (right) demonstrate poor ability to predict DOS using only XRD data

## Results

- Root Mean Squared Error =  $1.46 \text{ eV}^{1/2}$  (NN and MLR)
- $R^2 = 0.013$  (MLR)
- P-value of Coefficients Average to zero (MLR)
- Models were not able to accurately describe a relationship between crystallography and electronic parameters
- Electronic properties are likely dependent on many other parameters including composition, coordination chemistry and bonding to be modeled from XRD

## Future Work

The following adjustments could be made to potentially improve the performance of this package:

### Data

- Narrow dataset to similar material types (i.e. oxides)
- Filter redundancies found in Materials Project Database
- Create function to clean/process raw XRD JSON files

### Model

- Explore other machine learning models (kernel ridge regression, random forest, etc.)
- Incorporate calculations emphasizing physical properties

$$g_{\alpha\beta}(r) = \frac{1}{N_\alpha V_r} \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\beta} \theta(d_{\alpha_i\beta_j} - r)\theta(r + dr - d_{\alpha_i\beta_j}) \quad [1]$$

### Material

- Incorporate parameters that capture electronic properties: elemental composition, valency, bonding
- Train model using entire density of state instead of only bandgap

[1] Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K. R.; Gross, E. K. U. How to Represent Crystal Structures for Machine Learning: Towards Fast Prediction of Electronic Properties. *Phys. Rev. B* 2014, 89 (20). <https://doi.org/10.1103/PhysRevB.89.205118>.

### Acknowledgements

Thank you Dave Beck, Chad Curtis, Kelly Thornton, Jim Pfaendtner  
Crystallographic and electronic data is publicly available from the Materials Project Database  
All packages used and developed in this work are open source and can be found at the following web address: <https://github.com/michkatz/xrados>

### Group Dynamics

In-person communication was prioritized over digital correspondence. Slack and Trello tools were used for task management. Github was used to host and organize this collaboration.