Group 1: Predicting Crystal Systems in Lithium-ion Battery Cathode Materials

Group Members:

Abstract: This project aims to predict crystal systems in lithium-ion battery cathode materials by leveraging machine learning techniques. We explore key physical and chemical properties, including crystal system, formation energy, and density, using a dataset obtained from Kaggle. Our analysis involves 2D and 3D regressions, histograms, and a K-nearest neighbors (KNN) classification model. The extended heat map and scatter plots provide insights into correlations, while regression models offer predictive capabilities. The KNN model achieves a 5-fold cross-validation error of 0.366, showing promise for crystal structure predictions.

Background: Lithium-ion batteries play a crucial role in various applications, and understanding the properties of cathode materials is vital for battery performance. The crystal system of cathode materials influences battery characteristics, making it essential to predict and analyze these systems.

Dataset: The dataset, sourced from Kaggle, comprises detailed information on cathode materials for lithium-ion batteries. It includes physical and chemical properties such as crystal system, formation energy, density, space group, and chemical formula.

Methods:

2. 2D Regression:

a. Band Gap vs. E above Hull: (I) Linear Regression: A linear regression model is fitted, showing a correlation between band gap and energy above hull. (II) Parity Plot: A parity plot is used to evaluate the linear regression model, indicating discrepancies between predicted and actual values. (III) Polynomial Regression: A polynomial regression model is explored, with degree selection based on maximizing R-squared values.

b. Formation Energy vs. Density: (I) 2D Linear Regression: A linear regression model is applied to study the correlation between formation energy and density. (II) Parity Plot: The parity plot and residual plot reveal areas for model improvement. (III) Polynomial Regression: Polynomial regression is employed, and the best fit is determined based on R-squared values.

3. 3D Regression:

(a) 3D Plot: A 3D plot is created to visualize the correlation between formation energy, energy above hull, and band gap. (b) Polynomial Fitting: Polynomial regression is applied for a more accurate fit.

4. Histograms:

Histograms are generated for key features like space group, crystal structure, number of sites, chemical formula, and band gap. These visualizations provide insights into the distribution of data within each category.

5. Classification (KNN):

a. Extended Heat Map: The dataset is extended with numerical representations of space group and chemical formula. The extended heatmap demonstrates additional correlations.

b. Scatter Plot: A scatter plot illustrates the relationship between the number of sites and space group, colored by crystal system.

c. KNN Training: A K-nearest neighbors model is trained to predict crystal structures. The model's accuracy is evaluated on the test set.

d. Comparison of Train and Predicted Labels: Train and predicted labels are compared through scatter plots, visually demonstrating the model's performance.

Results and Discussion:

Correlation analysis revealed significant relationships between certain properties.

Linear regression models provided initial insights, while polynomial models improved accuracy.

3D plots visualized complex relationships between formation energy, energy above hull, and band gap.

Histograms illustrated the distribution of various features.

The KNN classification model showed promise with a 5-fold cross-validation error of 0.366.

Conclusion:

This comprehensive analysis of lithium-ion battery cathode materials provides valuable insights into the relationships between various physical and chemical properties. The correlations identified through linear and polynomial regressions, 3D plots, and classification models contribute to the understanding of material behavior. The extended heat map and histograms offer a detailed overview of the dataset distribution. The KNN classification model shows promise in predicting crystal structures, with a 5-fold cross-validation error of 0.366.

Future Work:

Future research could focus on refining regression models for more accurate predictions and exploring additional classification algorithms. Further feature engineering and dimensionality reduction techniques might enhance the predictive capabilities of the models. Additionally, expanding the dataset with more diverse cathode materials could lead to more robust and generalized models. The insights gained from this analysis lay the groundwork for advancing lithium-ion battery research and development.

In conclusion, this project not only contributes to the understanding of lithium-ion battery cathode materials but also provides a framework for similar analyses in materials science. The combination of regression analysis, visualization techniques, and machine learning models offers a holistic approach to studying complex datasets and extracting meaningful information for future applications.