Global Rare Earth Elements

Josh Cyphers, Haissam Akhras, Kris Riedman

Dataset and Exploratory Data Analysis

About the Dataset

- 146 Rare-Earth Mining Projects split into two subsets
- Sourced by Shuang-Liang Liu and team from company annual reports and public presentations, government reports, and scientific journals.

Factories

- Company
- Project
- Location
- Status
- Capacity
- Yield
- Next factory upstream
- Next factory downstream

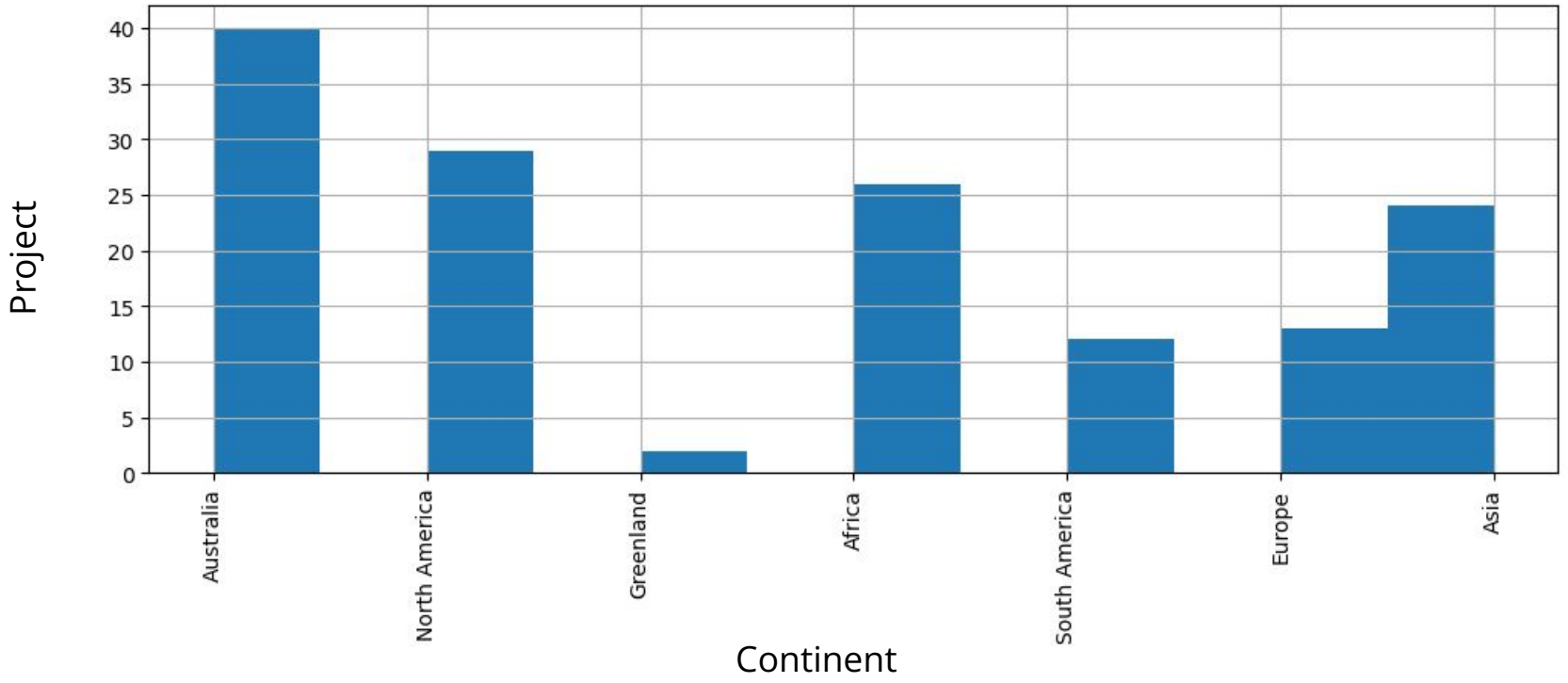
Projects

- Company
- Project
- Location
- Status
- Deposit Type
- Grade (wt. %)
- Element Composition
- % of total shale

EDA Findings

- Compiled from secondary sources which resulted in a messy dataset requiring lots of cleaning
- The amount of useful data for our purpose in the Projects dataset is small
  - Starting from only 146 entries, splitting into train/test sets makes this even smaller
  - Many projects are not yet complete so they do not have recorded HREE amounts
  - Grades, individual mineral amounts, and HREE percentage are large and specific numbers that are not easily categorized
  - Location entries are not uniformly formatted, continent will be better to work with
  - Some deposit types are specific to certain continents, some are found on all continents

Projects by Continent



HREE Percentage and Deposit Type by Continent

Continent	Deposit type	HREE percentage
Africa	Carbonatite	2.415983e+09
	Carbonatite (Tailings)	5.871301e+08
	Hydrothermal/IOCG (Monazite veins)	7.914286e+09
	Ionic Clay	1.156500e+03
	Placer	2.940000e+02
Asia	Alkaline	1.460000e+02
	Alkaline granite	9.728000e+03
	Carbonatite	8.885331e+08
	Clays	5.109000e+03
	Placer	2.898551e+09
Australia	Alkaline rock	3.168400e+09
	Carbonatite	4.855534e+09
	Hydrothermal/IOCG	5.654860e+09
	Hydrothermal/IOCG (Tailings)	1.475410e+09
	Ionic Clay	3.130000e+02
	Placer	1.008021e+09
Europe	Alkaline rock	2.495770e+09
	Carbonatite	3.016723e+09
	Hydrothermal/IOCG	3.371336e+09
Greenland	Alkaline rock	2.139951e+09
North America	Alkaline granite	3.986928e+09
	Alkaline rock	1.017145e+09
	Alkaline/Alkaline granite	3.731844e+09
	Carbonatite	1.103486e+09
	Hydrothermal/IOCG	1.582015e+09
	Metamorphic rocks (quartz feldspar gneiss), sedimentary layers, basic rock layers	1.666667e+09
	Paleo-sedimentary formation	7.291667e+09
	Rhyolite	7.419355e+09
South America	Alkaline granite	7.486000e+03
	Carbonatite	7.762995e+08
	Ionic Clay	1.163522e+09

Approach and Model

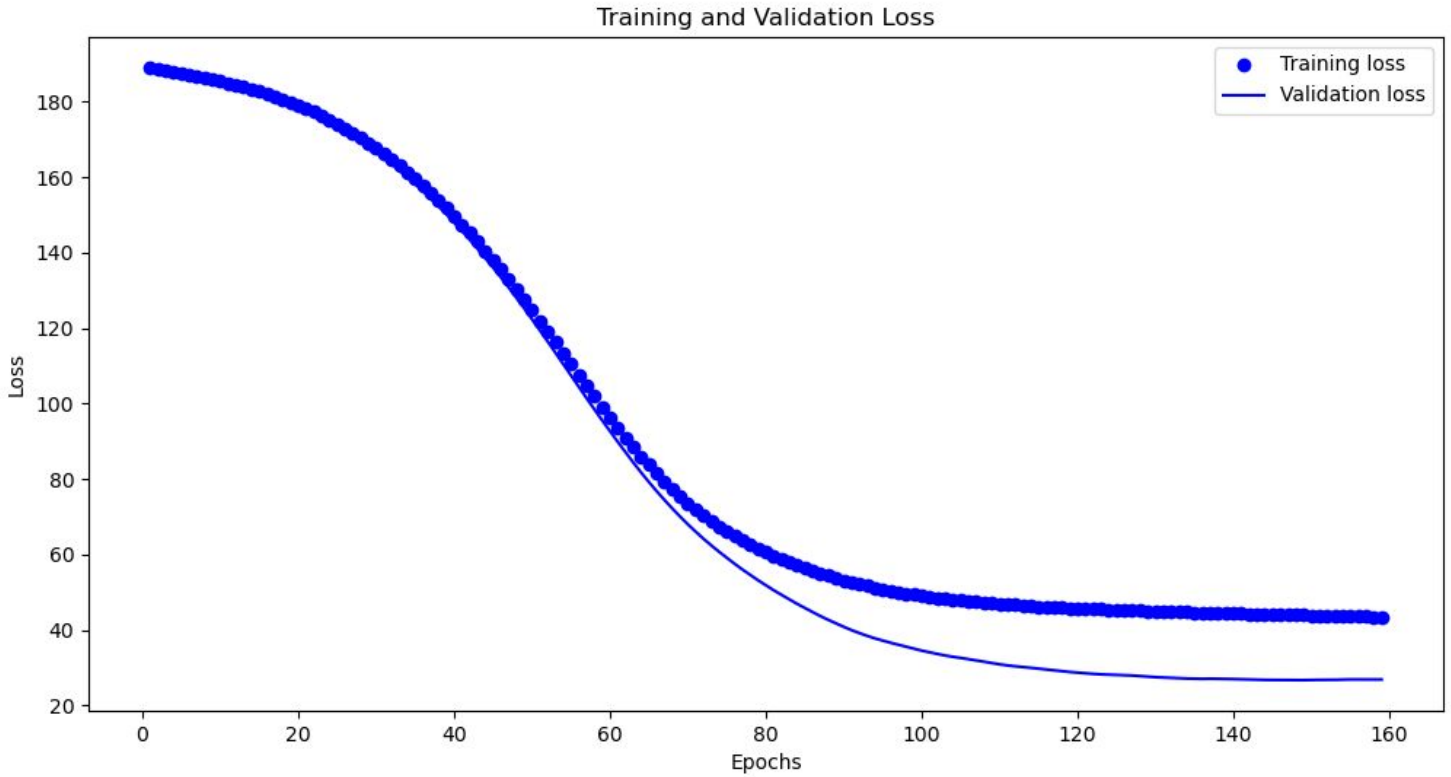
Model Iterations

- Predict mine location using element yields: 30% effective ~twice as effective as baseline
- Predict mine location based on HREE percentage and deposit type: ~20$ effective
  - Complicated prediction based on small dataset and specific HREE percentage values
- **Final Model: Predict HREE percentage based on Continent and Deposit type**

Model parameters Overview

- **Model Architecture:**
  - Input Layer: 64 neurons, ReLU activation
  - Hidden Layer: 32 neurons, ReLU activation
  - Output Layer: 15 neurons (corresponding to the elements list)
- **Loss Function:** Root Mean Squared Error (RMSE)
- **Optimizer:** Adam
  - Learning Rate: Default (0.001)
- **Training Configuration:**
  - Epochs: 100 - 160
  - Validation Split: 20% of training data

Training Process



Key Findings

Element	RMSE	Element	RMSE
Lu2O3	0.06	La2O3	0.23
Yb2O3	0.12	Y2O3	0.25
Nd2O3	0.14	Ce2O3	0.25
Tm2O3	0.16	Sm2O3	0.26
Eu2O3	0.19	Tb4O7	0.31
Pr6O11	0.20	Dy2O3	0.32
Ho2O3	0.22	Gd2O3	0.34

Overall Test RMSE

0.23

Tuning and Improvements

Hyperparameter Tuning

Our starting model's hyperparameters were set based on empirical choices:

- Hidden layers: 2
- Neurons in the first hidden layer: 64
- Neurons in the second hidden layer: 32

To optimize performance, a **grid search methodology** was adopted to fine-tune hyperparameters. The hyperparameters that were explored include:

- Neuron count in the hidden layers.
- Adam optimizer's learning rate.
- Training batch size.

Ensemble/Boosting Approaches

In addition to the single neural network model, ensemble techniques were also evaluated to bolster results:

- **Model Averaging**
- **Bootstrap Aggregating (Bagging)**

Through employing ensemble strategies, the prediction variance was decreased, leading to an improved RMSE across all target elements.

Conclusion

NeurIPS Checklist

1) For all authors...

   a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? **Yes**
   b) Have you read the **ethics review guidelines** and ensured that your paper conforms to them? **Yes**
   c) Did you discuss any potential **negative societal impacts** of your work? **Yes**
   d) Did you describe the **limitations** of your work? **Yes**

2) If you are including theoretical results...

   a) Did you state the full set of **assumptions** of all theoretical results? **n/a**
   b) Did you include complete **proofs** of all theoretical results? **n/a**

3) If you ran experiments...

   a) Did you include the code, data, and instructions needed to **reproduce** the main experimental results (either in the supplemental material or as a URL)? **Yes**
   b) Did you specify all the **training details** (e.g., data splits, hyperparameters, how they were chosen)? **Yes**
   c) Did you report **error bars** (e.g., with respect to the random seed after running experiments multiple times)? **Yes**
   d) Did you include the amount of **compute** and the type of **resources** used (e.g., type of GPUs, internal cluster, or cloud provider)? **Yes**

NeurIPS Checklist Continued

4) If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

   a) If your work uses existing assets, did you **cite** the creators? **Yes**
   b) Did you mention the **license** of the assets? **Yes**
   c) Did you include any **new assets** either in the supplemental material or as a URL? **n/a**
   d) Did you discuss whether and how **consent** was obtained from people whose data you're using/curating? **Yes**
   e) Did you discuss whether the data you are using/curating contains **personally identifiable information** or **offensive content**? **Yes**

5) If you used crowdsourcing or conducted research with human subjects...

   a) Did you include the full text of instructions given to participants and screenshots, if applicable? **n/a**
   b) Did you describe any potential participant **risks**, with links to Institutional Review Board (IRB) approvals, if applicable? **n/a**
   c) Did you include the estimated hourly wage paid to participants and the **total amount spent** on participant compensation? **n/a**

Future Work

- Expand dataset by looking for further references
- Add Latitude and Longitude for each mine location
- Research additional datasets that are derived from a single authority
- Analyze how demand for REE may drive future mine locations