# Machine Learning for Mineral Prospectivity Mapping

UCL Department of Civil, Environmental and Geomatic Engineering, Gower St, London, WC1E 6BT

**Lucille Ablett**
**MSc Spatio-Temporal Analytics and Big Data Mining**
**Supervisor: James Haworth**

## Background

Machine learning techniques are rarely used for geological data analysis, particularly outside of academia. The British Geological Survey has conducted initial studies into the use of machine learning for mapping the chemical composition of the UK near-surface using a range of geological and geophysical datasets, since chemical composition is a direct quantification of the soil or rock type present (Kirkwood et *al.*, 2016).

This is particularly useful for the mineral exploration industry as it allows companies to gain maximum knowledge from their datasets, aiding significantly in the discovery of potentially economic mineral deposits that would not otherwise have been identified using conventional methods.

This project expands on this area of research through investigation of the effectiveness of machine learning algorithms for mapping of hydrothermal mineralisation in SW England, via a comparison study of three machine learning algorithms.

## Aims and Objectives

The primary aim of this project was to investigate the effectiveness of machine learning algorithms for mapping of hydrothermal mineralisation in SW England.

This was achieved through completion of the following objectives:

1. Compare the accuracy of models created using the following machine learning methods, and the suitability for use in the mining and exploration industry:
   - ❖ Support vector regression (SVR)
   - ❖ Random forest (RF)
   - ❖ Extreme gradient boosted tree (XGB)
2. Simulate a range of differing sampling densities and investigate the effect on model accuracy.
3. Produce example maps of predicted mineralisation (Figure 3).

## Methodology

**Extract values of auxiliary variables at location of target variables**

**Divide data randomly into training and testing datasets:**
40%, 60%, and 80% of the full dataset was used to predict the remaining 60%, 40%, and 20% of the values respectively, simulating a differences in sampling density (future mentions refer to training %)

**Train models to predict the target variables using the auxiliary variables, using the following machine learning algorithms and R packages:**

1. Support Vector Regression (*caret*)
2. Random Forest (*randomForest*)
3. Extreme Gradient Boosted Tree (*xgboost*)

All models trained using 10-fold cross validation within the *caret* framework.

**Tune parameters until RMSE minimised**

**Select final model based on minimum RMSE**

**Validate models and analyse errors**

**Create predictive maps of mineralisation**

## Data

***Target Variables (point data):***

Concentrations of the elements

*Ag, Al, Br, Cu, Ni, K, Sn, Sc, Y, and Zr*

in mg/kg as measured from 1154 soil samples collected in SW England (Figure 1).

***Auxiliary variables (raster data):***

- ❖ *Regional geophysics:*
  - Land gravity
  - Radiometric and airborne magnetic surveys from TellusSW project
- ❖ *Remotely sensed data:*
  - Landsat 8 satellite imagery
  - NEXTMap aerial elevation survey

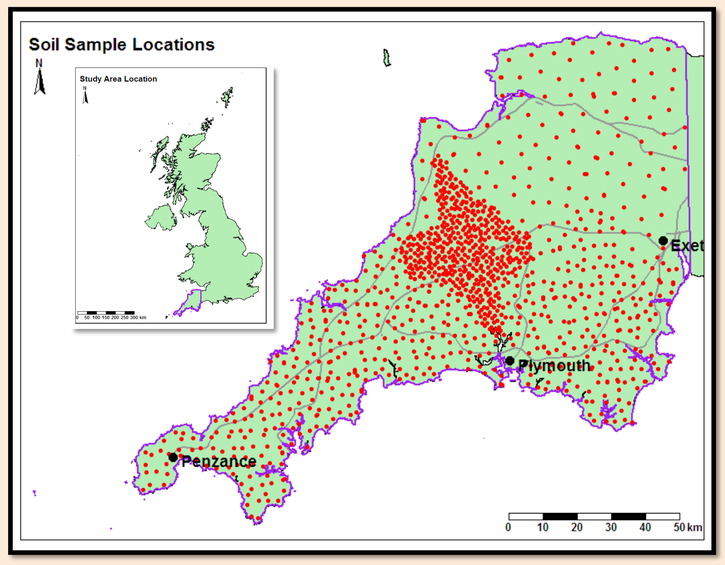


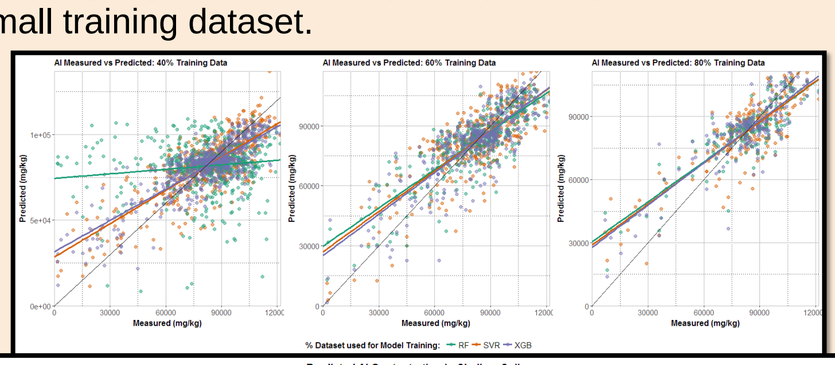

***Figure 1: Study area showing location of soil samples used as target variables***

## Results

- ❖ The best model was interpreted to be that with the highest cross-validated $R^2$ value and the lowest RMSE, as calculated by *caret* during training.
- ❖ $R^2$ varies considerably across the models, with a highest value of 0.720, and lowest of 0.006 for Al (RF) and Ag (XGB) respectively (Table 1).
- ❖ Although 10-fold cross validation was used, $R^2$ values calculated using the testing data were often significantly lower than those calculated during the tuning process. This is particularly apparent for the RF models created using only 40% of the data for training (Figures 4→6).
- ❖ Of all the methods, RF was most biased towards the mean, most frequently over-estimating low values and under-estimating higher values (Figure 2), and performed worst with a small training dataset.

***Table 1: highest and lowest $R^2$ values for all models***

| *Element* | ***Highest*** *Al* | | ***Lowest*** *Ag* | |
|---|---|---|---|---|
| ***%Train*** | 80% | 40% | 80% | 40% |
| ***SVR*** | 0.701 | 0.661 | 0.053 | 0.040 |
| ***RF*** | 0.717 | 0.720 | 0.072 | 0.044 |
| ***XGB*** | 0.695 | 0.616 | 0.021 | 0.006 |



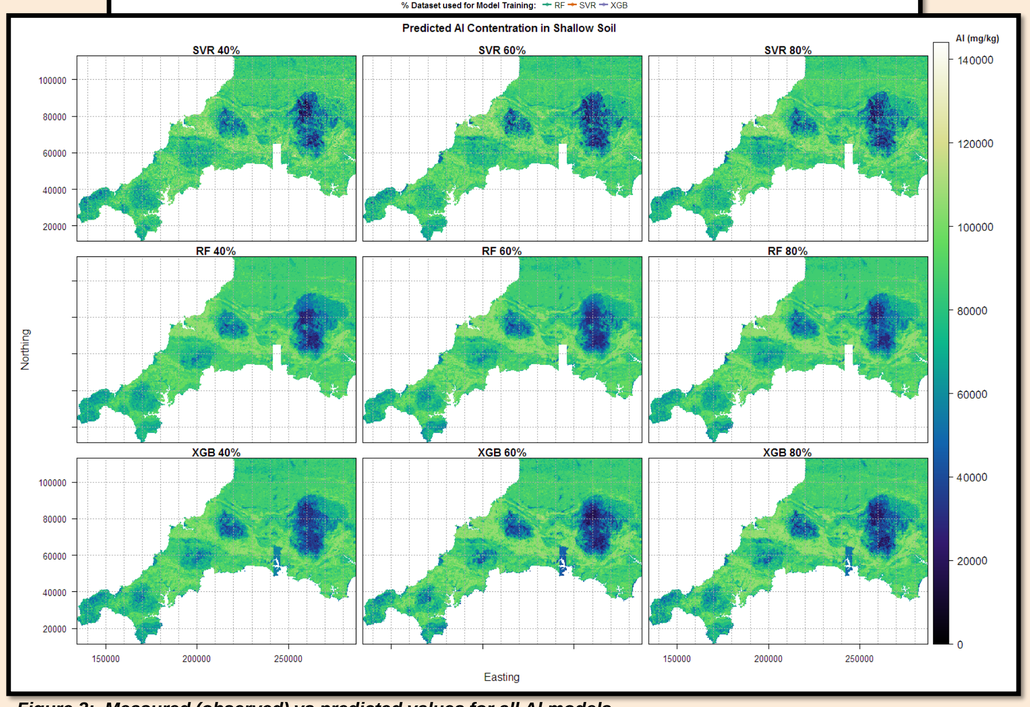
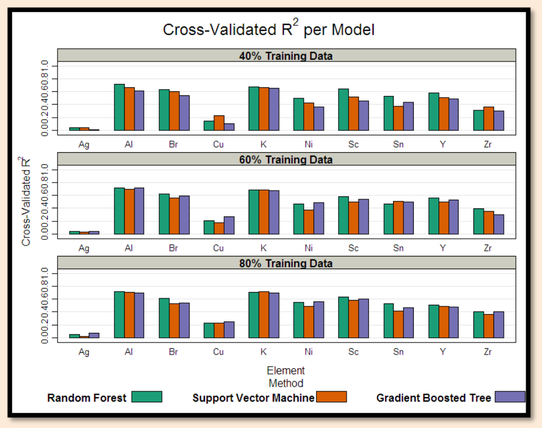
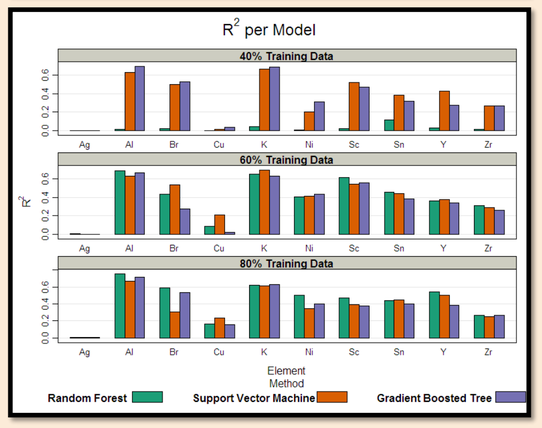
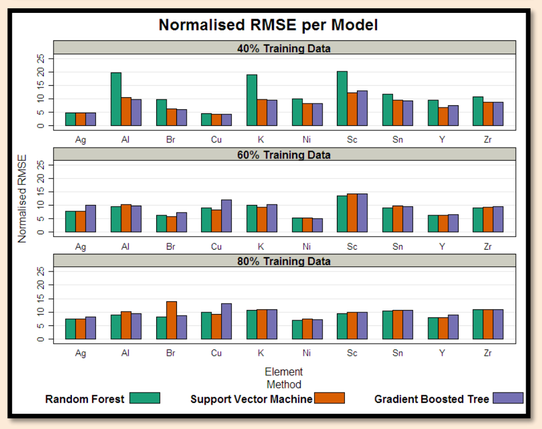




***Figure 2: Measured (observed) vs predicted values for all Al models***
***Figure 3: Prospectivity maps of predicted Al mineralisation for all methods and % training data (Derived from BGS Digital Data under License 2017/044GC ED).***




***Figures 4, 5, 6 (top → bottom):***
***4) X-validated $R^2$ values as calculated during training***
***5) $R^2$ as calculated from testing dataset***
***6) Normalised RMSE as calculated from training dataset***

## Conclusions

- ❖ Machine learning algorithms can be used to predict individual elements to a reasonable degree of accuracy
- ❖ RF is least suitable for small training datasets, XGB is most suitable for small datasets, SVR most successfully predicts extreme high and low values.
- ❖ RF performs best with larger training datasets, but is less capable of predicting the extreme high and low values
- ❖ XGB easier to train: significantly faster and less input required in the training process than SVR and RF, but creation of prospectivity maps more time consuming.

## Further Work

- ❖ Compare maps to those produced using kriging
- ❖ Investigate the mapping of confidence levels
- ❖ Explore prediction of whole rock composition, rather than individual elements
- ❖ Asses suitability of artificial neural networks
- ❖ Examine contribution of spatial and covariate information.
- ❖ Consider feature engineering: reducing dimensionality of predictor variables or creating new ones to improve prediction accuracy.

British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

***References***
Kirkwood C, Cave M, Beamish D, Grebby S and Ferreira A 2016 A machine learning approach to geochemical mapping *J. Geochemical Explor.* 167 49–61 Online: http://dx.doi.org/10.1016/j.gexplo.2016.05.003