Mineral Exploration Using Gaussian Process Regression & Boosting Algorithms

Team Sigma
Alexander Pak, Bailey Lei, Birinder Singh, Chao Wang, Harjyot Kaur
Date: May 30, 2019

Executive Summary

In this proposal, we researched the geological approaches in mineral detection. With our understanding of the drill hole database, external hydrochemical data, soil type data gathered from SARIG, and other internal data sources, we used boosting algorithms to predict the top 10 gold and copper deposits in the Mount Woods Area. Furthermore, used Gaussian Process Regression to validate the top sites we predicted. Lastly, we conducted a risk analysis and a feasibility study to validate our selections.

Background

According to Infographic: The Mineral Exploration Roadmap[1], the odds of detecting a profitable greenfield mineral target mine is 1 in 1000. The proportion of gold deposits on the map with enough gold to justify any further development is only 10%. Due to this low success rate, our team believes that it is essential to incorporate diverse strategies from fields such as data science and geophysics to increase the likelihood of finding a profitable mine.

Based on the map, there is a significant decrease in both the number of major gold deposits discovered (>2.5 Moz Au) and the amount of gold contained in these deposits. Consequently, the resulting loss of profit would lead to increased costs, as well as higher pressure for annual reserve and resource replacement[2].

Gold: Native gold generally contains impurities such as silver (Ag), copper (Cu) and iron (Fe) with traces of bismuth (Bi), lead (Pb), zinc (Zn), antimony (Sb), tellurium (Te), mercury (Hg), tin (Sn) and platinum (Pt).

Copper: Copper is produced from the sulphide ores.The main ore mineral of copper in Australia (and worldwide) is chalcopyrite (CuFeS2).

Prominent Hill is a copper-gold mining operation in northern South Australia that is solely owned by OZ Minerals.

The asset comprises:
- Open pit mine (Malu) - which concluded operations in Q1 2018
- Underground mine (Ankata and Malu mine areas)
- Conventional crushing, grinding and flotation Processing Plant

[1] (2018, March 5). Infographic: The Mineral Exploration Roadmap - Visual Capitalist. Retrieved May 30, 2019, from https://www.visualcapitalist.com/mineral-exploration-roadmap/
[2] (n.d.). Models and Exploration Methods for Major Gold Deposit Types.pdf. Retrieved May 30, 2019, from https://www.911metallurgist.com/blog/wp-content/uploads/2015/10/Models-and-Exploration-Methods-for-Major-Gold-Deposit-Types.pdf

Located in one of the world's most mining-friendly locations, Prominent Hill produces one of the highest grades of copper concentrate traded on the open market. The company possesses highly competitive fundamentals and contains great opportunities for organic growth[3].

Exploration Strategy

Traditional mineral exploration techniques include some of the following steps:

1. Prospecting, which consists of mapping outcrops or analysis of the minerals indicating presence of valuable minerals.

   The exposed outcrops contain information about prior geological activities that might have resulted in the exposure.

2. Early stage exploration using existing maps, historical data, as well as physical methods like gravity, magnetic, and electromagnetic surveys.

   The information about the fault lines indicate prospects of having associated minerals. For instance, small-scale fault systems in the Earth's crust have a strong correlation with the location of gold as found in a study regarding the St. Ives Goldfields in Western Australia[3.1]. Researchers proposed that the relationship between fault systems and gold traces is essential to understanding the genesis of gold and could be used to assist in locating any commodity. The information obtained allows for the focus to be directed onto one region of interest.

3. Surface mapping and sampling in the field of the region is conducted and the rock samples collected are analysed for any anomalous behaviour.

   Significant information is obtained by analysing the hydro chemistry of the groundwater. The inorganic constituents resulting from the chemical interactions with geological materials and to a lesser extent contributions from the atmosphere help to evaluate hydrogeochemical processes responsible for temporal and spatial changes in the chemistry of groundwater in the region. For instance SO4CISW index helps in differentiating the SO4-rich groundwaters from the poor ones. Major areas of high-U groundwaters can act as an indicator for mineral systems such as iron oxide copper gold (IOCG).

4. Important observations are acquired during the process of drilling cores including valuable insights like rock type, structure, and alteration.

---

[3] (n.d.). Prominent Hill | OZ Minerals. Retrieved May 30, 2019, from https://www.ozminerals.com/operations/prominent-hill/
[3.1] (n.d.). Fault lines lead to gold | . Retrieved May 30, 2019, from http://www.mining.com/web/fault-lines-lead-to-gold/

Magnetic susceptibility and specific gravity data is useful in mineral identification.

5. Once the discovery has been made, the drill holes are extended to evaluate the extent of the ore body. Consequently, the information gathered is used to construct a lithology model.

   Hundreds of drill holes may be required to define the boundaries and evaluate the quality of an orebody. Often, directional drilling significantly reduces the number of drill holes required to discover a resource in the ground. In case of Mounts Wood Domain, the available data from the vertical wells provides sufficient information about the cross section of the ore bodies. The latitudes and longitudes of the location along with the its depth combined with other observable properties like magnetic susceptibility and specific gravity can act as an indicator for the presence of other minerals in the vicinity.

6. Cost analysis is conducted to understand the feasibility of exploring the mineral ore.

   The process of extracting ores is highly costly. In addition, the drilling process can have a long lasting environmental impact. Therefore, the quality of ore needs to be considered before making a decision for mining.

Since the change in geology is not abrupt and any geological property at one location is likely to be similar to the nearest point than the ones far apart, we can deploy the kriging technique to create a mesh grid for generating a continuous spatial distribution of the property. Theoretically, kriging is an advanced geostatistical procedure that generates an estimated surface from a scattered set of points (denoted by latitude and longitude) with z-values(measured depth in our case). Unlike other interpolation methods in the Interpolation toolset, the use of Kriging tool effectively involves an interactive investigation of the spatial behavior of the phenomenon represented by the z-values before selecting the best estimation method for generating the output surface. Exploratory Data Analysis helps choose relevant features for inputting in the kriging model. Kriging is most appropriate and has always been used in soil science and geology. The image below shows the pairing of one point with all other measured locations. This process continues for each measured point until we obtain a continuous grid of measurements[4].

---

[4] (n.d.). How Kriging works—Help | ArcGIS Desktop - ArcGIS Pro - ArcGIS Online. Retrieved May 30, 2019, from https://pro.arcgis.com/en/pro-app/tool-reference/3d-analyst/how-kriging-works.htm

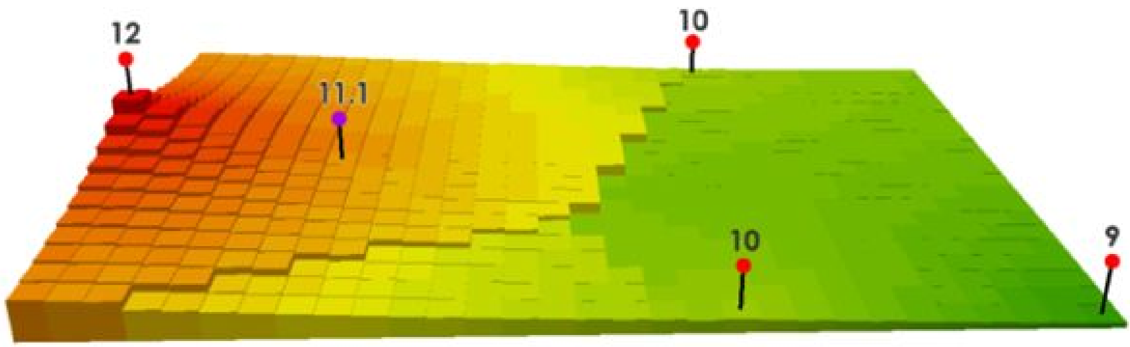

Figure 1: Examples of Usage of Kriging Interpolation
Figure source: https://gisgeography.com/kriging-interpolation-prediction/

High IOCG prospectivity of Mount Woods Inlier was established by:

- High magnetic relief – Fe-oxides
- Presence of known mineralisation and alteration systems; anomalism of REE, F, Ba, Cu and Au
- Na, K and Si alteration halos
- Evidence of high structural levels (i.e. GRV)
- Various mafic to felsic intrusive suites & major NE and NW structures
- Presence of mantle derived mafic rocks and related Hiltaba aged granites.

High-grade copper-gold mineralization at Prominent Hill was discovered in October 2001 through drill testing of a high frequency, non-magnetic portion of gravity anomaly. Results of URAN1 included; 20 metres @ 3.0g/t Au, 107 metres @ 1.9% Cu and 0.65g/t Au and, deeper, 152 metres @ 1.1% Cu and 0.6g/t Au[5]. With the available information from various resources, we conducted our analysis as detailed in the sections below.

Data Description and Usage:

Datasets were provided to study and gauge the geological features of every drill hole/site.

1. **Assays:** Each row in this dataset describes the mineral concentration found across 54 elements at different depths in a sample taken at a particular drill hole/site. The data was processed and cleaned for further analysis. The following describes the pre-processing done:

[5] (2011, April 18). prominent hill - OZ Minerals. Retrieved May 30, 2019, from https://www.ozminerals.com/uploads/media/ASX_20110418_Analysts_Exploration_Overview-2eaa3e67-f970-46b4-8ee1-2fd320bda72a-0.pdf

- Since there was a disparity in the units of measurement of the elements across different samples, all the quantities were converted to PPM (parts per million) respectively to achieve uniformity in the data.
- Some concentrations were less than 0 and thus, replaced by a zero value
- Null values are taken out from the dataset.

2. **Collars:** Each row in this dataset describes the location, status, type, and dip of each drill hole/site.

3. **Lithology:** Each row in this table describes the type of rock of a certain depth found at a particular drill hole/site. Below you can see the lithology mapping for the top 10 sites by aggregated copper concentration.

4. **Magnetic Susceptibility:** Each row in this describes the magnetic susceptibility measured at certain depth at a particular site.

5. **Specific Gravity:** Each row in this dataset describes the specific gravity, mass in air, mass in water measured at certain depth at a particular site.

Data Merge

The following steps showcase how all the datasets were merged together to conduct further analysis.

Step 1: A new feature, Mid_Depth, was created to obtain the midpoint between the depth interval for every sample in assays. Further, a new mapping, Site_Depth, was created to define the midpoints for every site/drill hole.

Step 2: Since Magnetic Susceptibility is available at various depth intervals for each site, these depth intervals were filtered using Site_Depth. Similarly, the same filtering was done over Specific Gravity.

Step 3: Assays table was joined with Collars table using SiteId to retrieve latitude and longitude for each drill hole/site.

Step 4: The table from Step 3 was used to join the filtered rows of the Magnetic Susceptibility and Specific Gravity table to retrieve the magnetic susceptibility, specific gravity, mass in air, and mass in water for each drill hole/site at varying depths.

Step 5: The merged table from Step 4 was used to get element- concentrations that were correlated to Copper and Gold. The elements that crossed the threshold for correlation were further used to select the final set of features for Copper and Gold.

Step 6: Multiple Imputation method was used to impute missing values in magnetic susceptibility, specific gravity, mass in air, and mass in water.

The steps outlined above provided the datasets used for predicting the concentration of Copper and Gold.

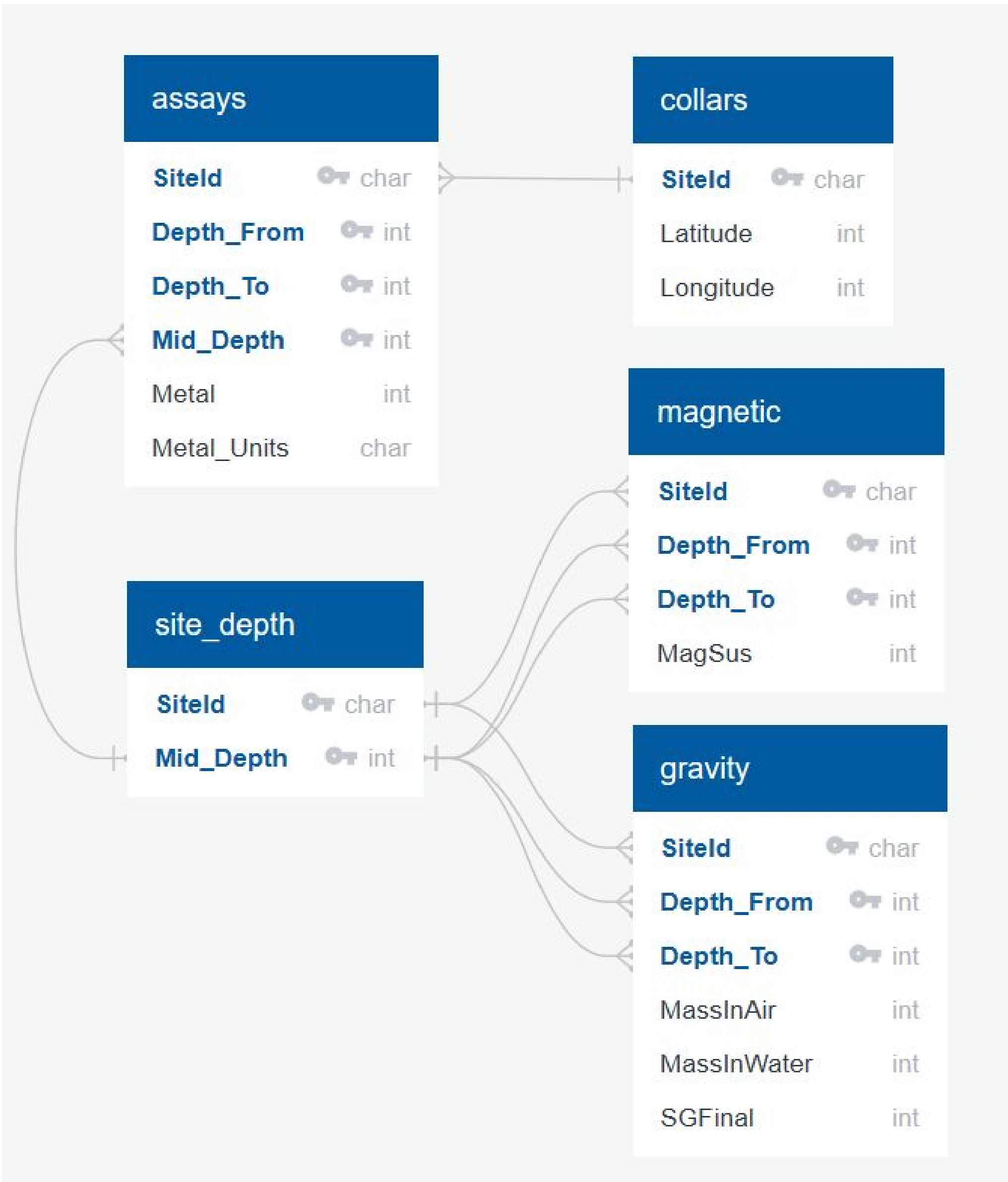

Figure 2: Entity Relation Diagram of Drill Hole Database

Multivariate Imputation of Missing Values in Python

After following the above steps to create the final merged dataset, the next step was to handle the missing values. Although some machine learning models have built-in methods of handling missing values, further imputation of missing values would minimize error and lead to more accurate predictions.

To fill in the missing values, our team used an iterative imputation approach, a type of multivariate imputation. Using the 'fancyimpute' package in Python, the missing values in the merged dataset were imputed using other features in the dataset in a round-robin fashion. This process provided a rich dataset containing complete values of all features for model training.

Predictive Modelling: Light GBM and Gaussian Process Regression

To predict areas of high copper and gold deposit concentration, we combined two different approaches: a Light Gradient Boosting Model (LGBM), and Gaussian Process Regression (Kriging). This was done to combine data science methods with current geology teachings to reach the most accurate predictions possible.

Model 1: *Light GBM*

An LGB regression model offers several advantages over other types of regression models. One main advantage is that training and validating LGBM requires only a few seconds on relatively light hardware (no GPU) and hence, allows for efficient hyperparameter tuning. Our team created several iterations of the original LGBM to design our final predictive model. During the training process, we ensured that our model was not overfitting to the data through 5-fold cross-validation. This process was separately done for both gold and copper.

The second main advantage of using this model is that LGBM is a random forest model at its core. This means we are able to view and quantify exactly how important each feature is in the final prediction. Therefore, in addition to predicting sites of high mineral concentration, we can also gain insight into our predictions to improve speculation for various locations.

To create predictions on the prospective area, we outlined a 1 degree latitude by 1 degree longitude map roughly encompassing the area that OZ Minerals can mine. This square was then divided into 1,000,000 equal points (latitude and longitude were incremented by 0.001, creating a 1000 x 1000 grid). This process was repeated 7 times to simulate depths between 0 and 600m at 100m increments, inclusively. In total, there are seven grids with 1,000,000 points.

For each of the 7 grids, every one of the 1,000,000 points had copper and gold concentration values predicted using the previously validated models. This information was used in two different ways:

- The predictions for each depth were joined together and sorted based on highest concentration to determine the Top 10 sites by latitude, longitude, and depth. These results will be submitted as the final predictions.
- The predictions were vertically averaged (ie. the depth axis was removed) and the resulting concentrations were sorted to find the Top 10 sites by latitude and longitude. These results can be plotted on top of a 2D map, to show the 10 sites with the highest potential.

Although these are the two methods of aggregation we have decided upon, these are not the only ones possible. Because we have a final generative model, we can create maps with as much granularity or as much aggregation as the situation demands.

Model 2: *Gaussian Process Regression (Kriging)*

We used the spatial prediction with Gaussian process regression to help validate the results produced by LGBM. We then selected the top 10 sites that corresponded to the mineral prediction from the LGBM model as well as the spatial prediction from the drill/hole site for each mineral. We performed the ordinary Kriging on both gold and copper drill/hole site. Using the data on concentration, we were able to map out a smooth gradient of various levels of gold and copper deposit. The copper gradient suggests that there is a fairly even amount of copper throughout the landscape. However, the gold gradient suggests that there is a clear gradient change between the area of overlapping high and low gold concentrations (Figure 3). This observation is consistent with our result predicted from the LGBM model.

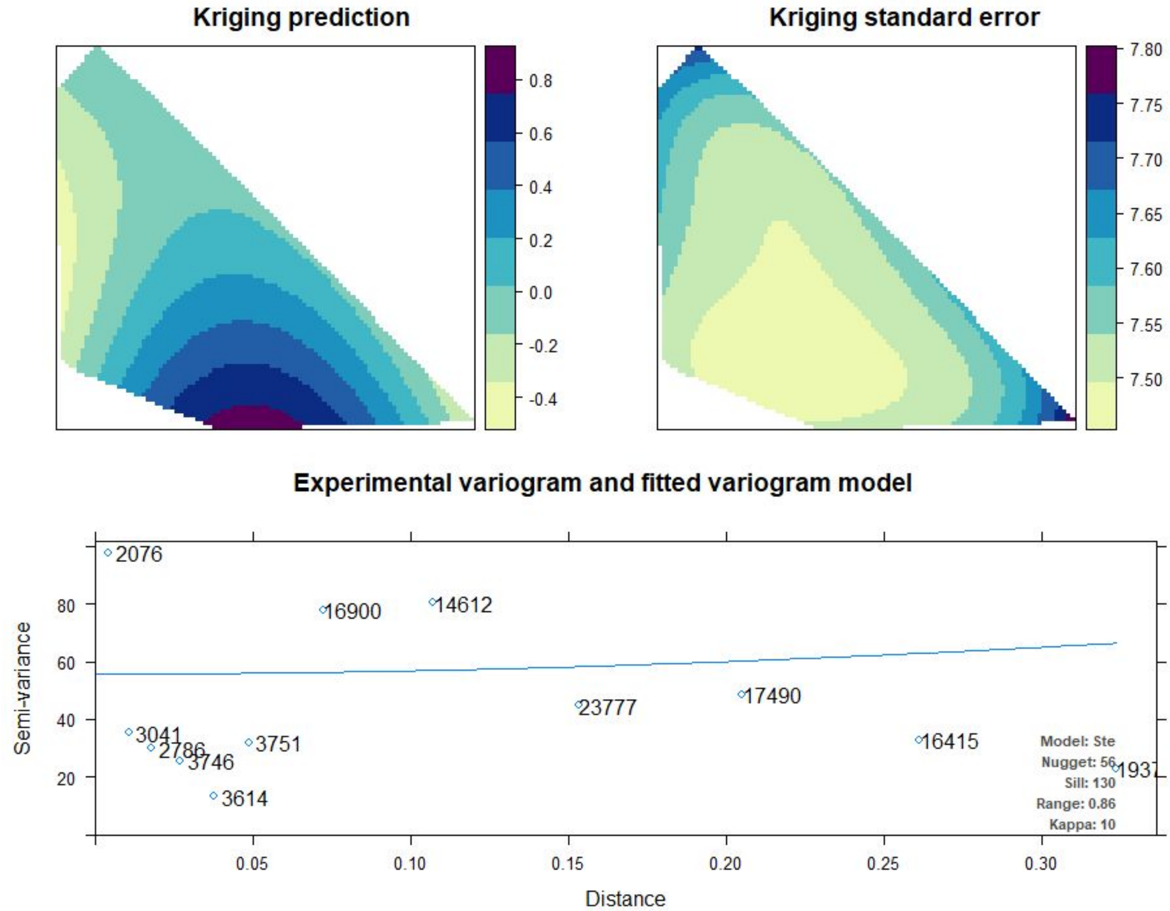

Figure 3: Ordinary Kriging with gold concentration on drill site

Due to lack of outcrops of crystalline basement rocks in the Southern Mountwood domain, any geological interpretation has to be based largely on potential field data and drilling information. The concentration of drilling data points appears centralized in the southwest region (Coverage Area : 135.07415771484 -29.998949050206, 135.07415771484 -28.999732931519,136.07940673828,-28.999732931519,136.07940673828 -29.998949050206). From our observations, the coverage area is relatively small and the landscape appears to be fairly homogeneous visually. This combination of using both machine learning and geological methods to help validate the prediction allow us more certainty with the area to prospect

Model Selection Process

Before choosing a combination of LightGBM and Kriging, other models were considered. For example, a 4-layer neural network was originally created and tested using Keras. The neural network predictions were not as accurate as the LightGBM model, and the model itself had no interpretability. The neural network also did not specify which features were important, or why the network was predicting the concentrations it was. Furthermore, the neural network was slow to train even after incorporating a GPU. Hence, we selected an LGB regression model over a neural network. Knowledge of the features that are important to the LGBM allows anyone else using our model to accurately conduct further EDA at future prospective sites.

Feature Importance

Our team decided to focus on a purely data driven approach to discover feature importance. Because we are using a LightGBM model to make our predictions, we can visualize and quantify feature importance in the final model. The benefit of a data driven approach is that we can methodically determine the most important features, alongside traditional geological knowledge. This strategy will give us an objective view into how the model operates, as well as what features to measure when predicting mineral concentrations at new locations.
When predicting gold concentration, our final set of feature importances are as follows:

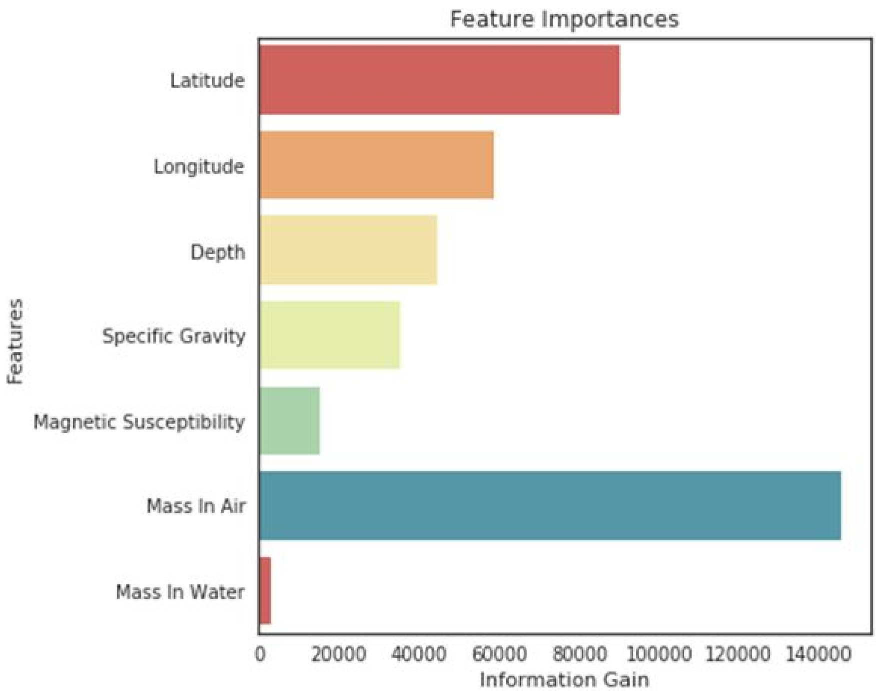

Figure 4.1: Feature Importance of Gold Concentration

Latitude, longitude, and depth are important predictors for gold concentration. Mineral deposits are usually found in “groups”, meaning spatial correlation is expected. However, by combining latitude, longitude, and depth with other predictors such as specific gravity, magnetic susceptibility, and mass in air, our model is able to generate its final prediction.
The copper concentration model:

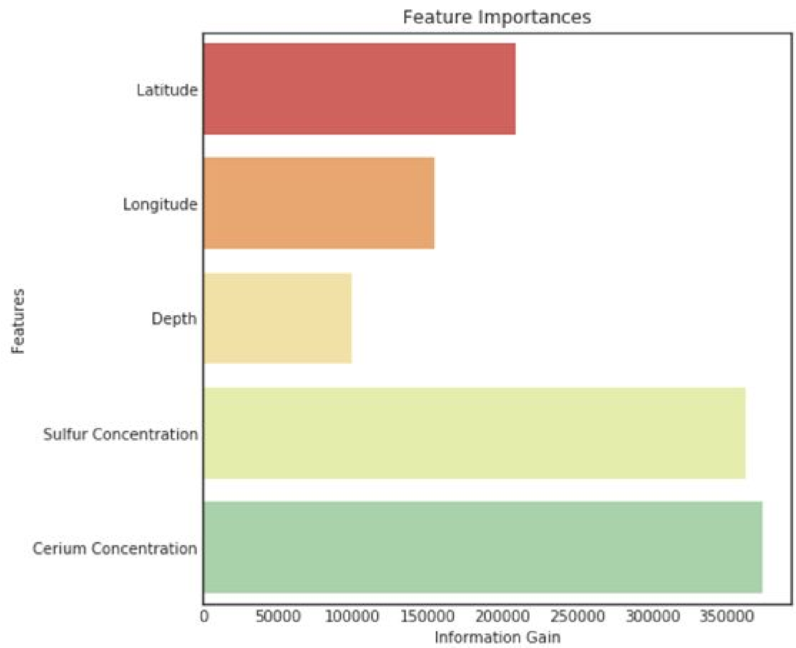

Figure 4.2: Feature Importance for copper concentration

Similarly, latitude, longitude, and depth show up as important predictors for copper concentration values. However, in comparison to the gold predictive model, sulfur and cerium concentrations are shown to be strong predictors for copper concentration. Interestingly, we found that the strongest predictors of gold are not the strongest predictors of copper.

Of course, we cannot ignore traditional geological teachings. By layering features we believe are important mineral predictors on top of a map with our model's predictions, we can validate our chosen features and resulting predictions.

Results

After filtering down our results to the top locations with high mineral concentrations, the following maps for gold and copper were created. The top 200 locations with the highest concentrations were taken and layered on top of a satellite map to visualize the locations with highest potential. The top 10 sites were with the highest potential are highlighted:

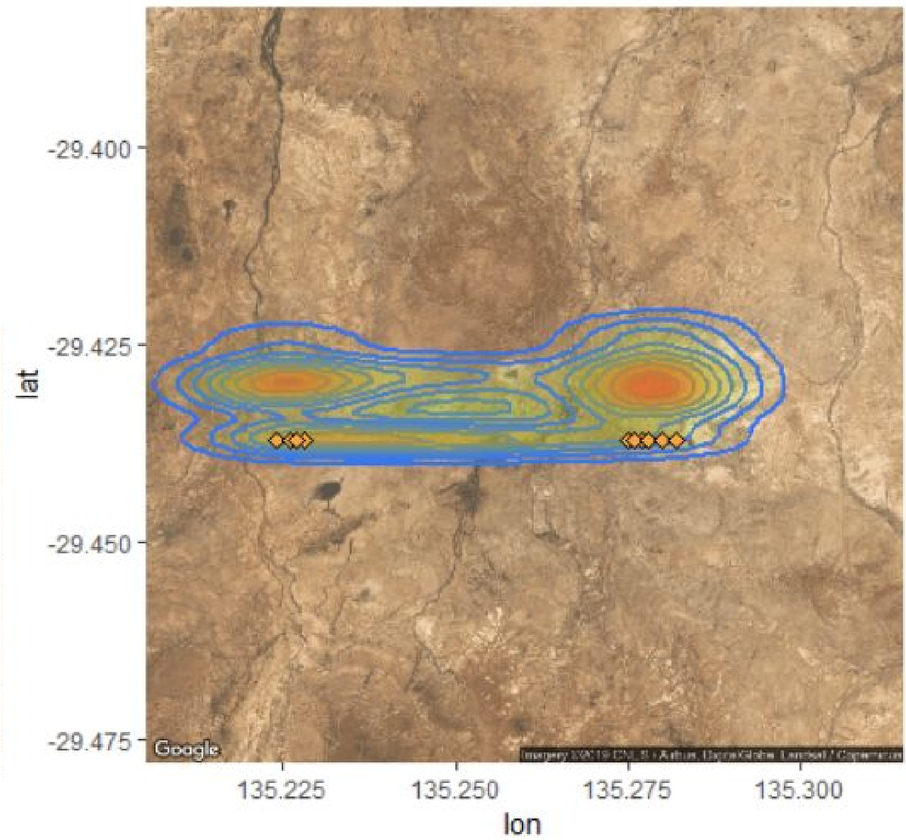

Figure 5.1: 2D Prospectivity Map of Top 10 Copper Sites

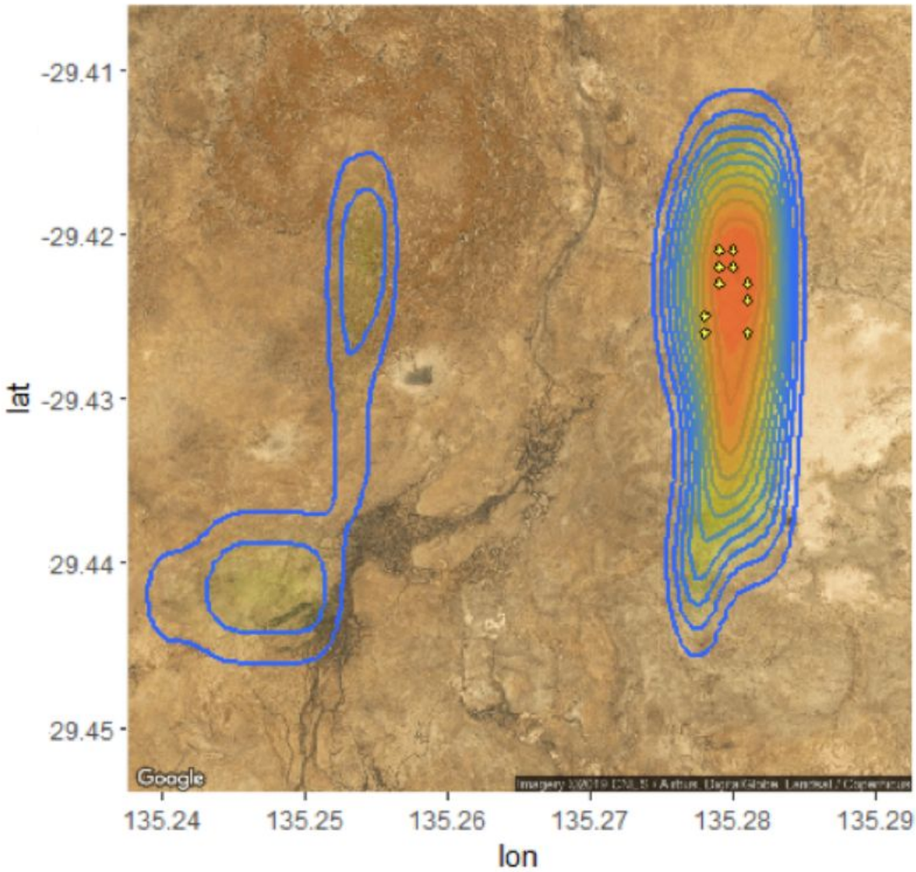

Figure 5.2: 2D Prospective Map of Top 10 Gold Sites

Looking at the plots above, the predicted sites with highest potential for copper and gold are in close proximity to each other. The sites highlighted in the above maps have the highest predicted mineral concentration after depth has been aggregated.

Model Validation using Kriging

Copper and gold domains were generally interpolated using either Ordinary Kriging (OK) or Inverse Distance squared (ID2) methods.[6] We used the spatial prediction along with the Gaussian process regression to help validate the result. Figures below are generated using hydrochemical data from Hydrogeochemistry of South Australia: Data Release[7].Total dissolved solids (TDS) is a measure of the combined total of organic and inorganic substances contained in a liquid.[8] Minerals in the surface water could be an indicator for the high concentration minerals underground.

[6] "Prominent Hill Mineral Resource and Ore Reserve ... - OZ Minerals." 30 Jun. 2018, https://www.ozminerals.com/uploads/media/181112_Prominent_Hill_Mineral_Resource_and_Ore_Reserve_Statement_as_at_30_June_2018.pdf. Accessed 31 May. 2019.
[7] "CSIRO Data Access Portal - Home Page." https://data.csiro.au/collections/. Accessed 31 May. 2019.
[8] "What is the acceptable Total Dissolved Solids ... - Berkey Water Filter." https://theberkey.com/blogs/water-filter/what-is-the-acceptable-total-dissolved-solids-tds-level-in-drinking-water. Accessed 31 May. 2019.

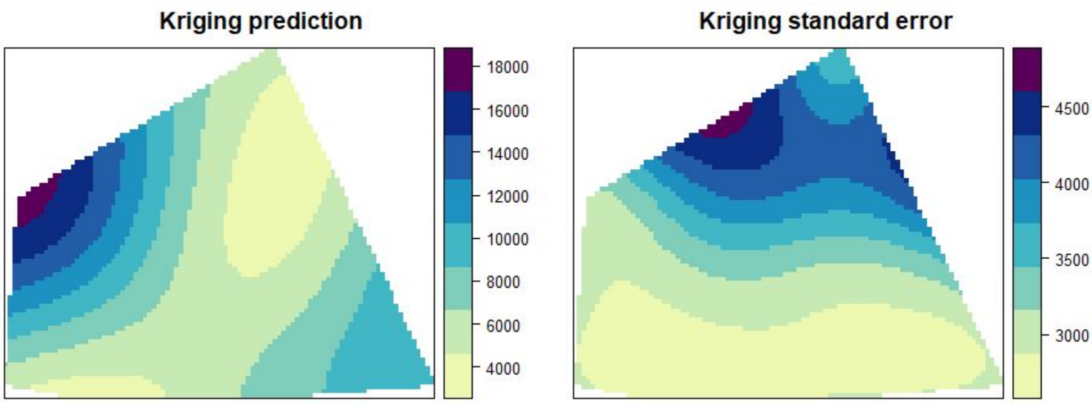

Figure 6: Model Validation using Total Dissolved Solids (TDS) Concentration

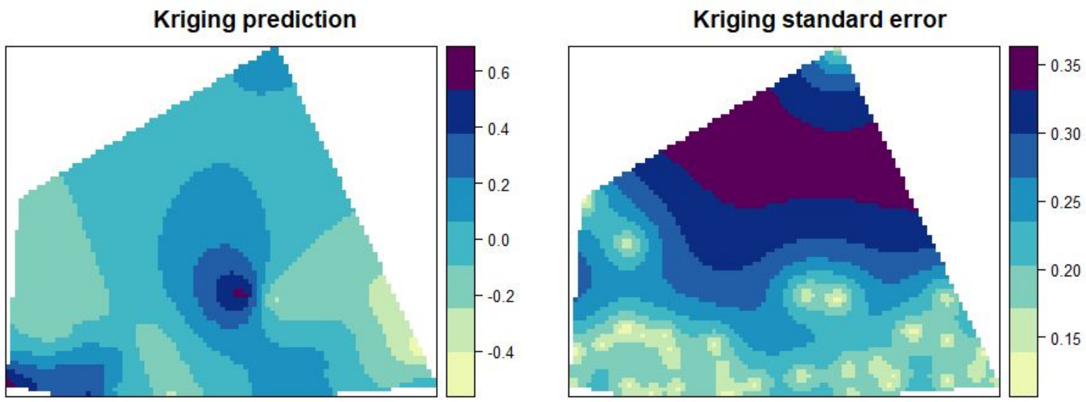

Figure 7: Model Validation using Magnesium(Mg), Sodium(Na) Surface Water

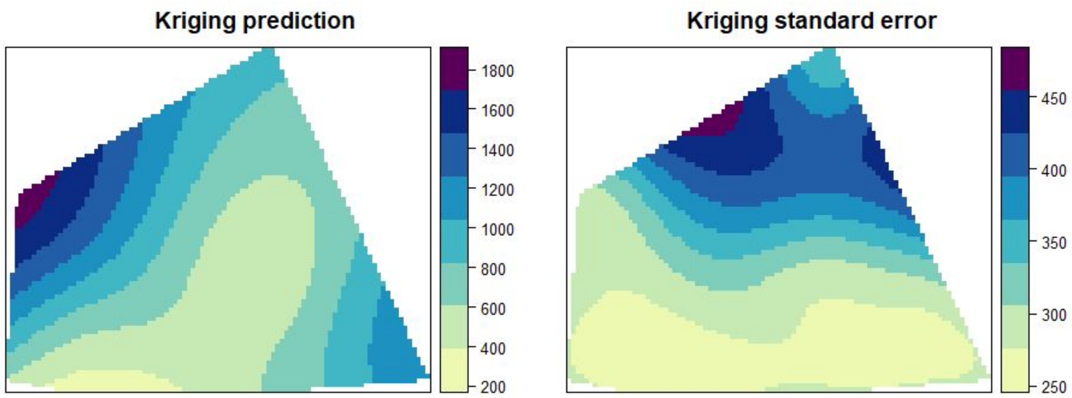

Figure 8: Model Validation using SOI4CID

Top Prediction Results

Based on the feature importance, the most important predictors for the Cu and Au concentration are latitude, longitude, depth-from, and depth-to. We have predicted the top 10 sites for each metal.

	Latitude	Longitude	Depth	Predicted Gold Concentration
0	-29.722	135.574	0	7.967636
1	-29.722	135.575	0	7.967636
2	-29.722	135.574	100	7.850216
3	-29.722	135.575	100	7.850216
4	-29.721	135.575	400	5.661776
5	-29.721	135.574	400	5.661776
6	-29.720	135.574	400	5.179574
7	-29.720	135.575	400	5.179574
8	-29.719	135.575	400	4.994691
9	-29.719	135.574	400	4.994691

Table 1.1: Top 10 Gold Predictions by Latitude, Longitude, and Depth

	Latitude	Longitude	Depth	Predicted Copper Concentration (ppm)
0	-29.723	135.563	400	10.493921
1	-29.722	135.563	400	10.493921
2	-29.724	135.563	400	10.026416
3	-29.725	135.563	400	10.026416
4	-29.776	135.563	400	8.845558
5	-29.720	135.563	400	8.794481
6	-29.721	135.563	400	8.794481
7	-29.723	135.563	300	8.769163
8	-29.723	135.563	200	8.673746
9	-29.725	135.563	300	8.378422

Table 1.2: Top 10 Copper Predictions by Latitude, Longitude, and Depth

Cost Justification

According to InfoMine, Gold Price is 1,302.35 USD/ozt (1,167.71 EUR/ozt) as of 31 May 2019[9]. Copper Price is 2.66 USD/lb (5,865.40 USD/t | 5,256.20 EUR/t) as of 30 May 2019[10].

[9] "Gold Prices and Gold Price Charts - InvestmentMine - InfoMine." http://www.infomine.com/investment/metal-prices/gold/. Accessed 31 May. 2019.
[10] "Copper Mining Stocks, Companies, Prices and News - InfoMine." http://www.infomine.com/investment/copper/. Accessed 31 May. 2019.

When considering the top 10 sites, factors like waste management, production per day, energy cost and supplies and materials and labour should all be taken into consideration in addition to the commodity grade.

We suggest conducting Preliminary Economic Assessment and In-depth feasibility study to determine whether the mineral resource can be mined economically.

Risk Analysis

The following table listed the common risks, the level of severity, and frequency in mineral detections.

Risk	Severity	Frequency	Mitigation	Justification
Uncertainty of the the deposit size, grade etc.	Medium	Frequent	Drill more holes. Step out drilling: These holes are drilled to see how far our predicted deposits goes along a trend. Infill drilling: Drilling holes in between will boost our confidence.	Boost confidence
Minerals cannot be extracted from the ore economically	Medium	Low	Metallurgical test can be conducted	Test results contain information of the structure and alloy examples.
Accidents & Injuries			Use AI for hazards and failure predictions	

Conclusion

In our approach, we combined the geological interpretations with machine learning to generate predictions on the spatial distributions of ore concentrations. With our understanding of the drill hole database, external hydrochemical data, soil type data, and other internal data sources gathered from SARIG, we used boosting algorithms to predict the top 10 gold and copper deposits in the Mount Woods Area. We also validated our predictions on the top 10 sites using Gaussian Process Regression.

From our models, the top predicted sites were found to be at a latitude and longitude of (-29.723, 135.563) for copper and (-29.722, 135.574) for gold.

Citation

1. The 4th Industrial Revolution: How Mining Companies Are Using Ai, Machine Learning And Robots Bernard Marr - https://www.forbes.com/sites/bernardmarr/2018/09/07/the-4th-industrial-revolution-how-mining-companies-are-using-ai-machine-learning-and-robots/#20f3304e497e, Date Accessed May 20, 2019
2. CSIRO Data Access Portal, https://data.csiro.au/collections/#collection/CIcsiro:17862v1, Date Accessed May 23, 2019
3. Risk Assessment and Management for Mines and Mining Plants, Ministry of Labour - https://www.labour.gov.on.ca/english/hs/pubs/gl_risk_assessment.php, Date Accessed May 23, 2019
4. Models and Exploration Methods for Major Gold Deposit Types Robert, F. ,Brommecker, R. Bourne, B. T. Dobak, P. J. McEwan, C. .J. , Rowe, R. R., Zhou, X. https://www.911metallurgist.com/blog/wp-content/uploads/2015/10/Models-and-Exploration-Methods-for-Major-Gold-Deposit-Types.pdf, Date Accessed May 13, 2019
5. 9 Reasons Mining Investors Are Looking At Yukon Companies, Jeff Desjardins - https://www.visualcapitalist.com/9-reasons-mining-investors-are-looking-at-yukon-companies/ Date Accessed May 13, 2019
6. How Ai and Big Data Will Unlock the Next Wave Of Mineral Discoveries, Nicholas LePan - https://www.visualcapitalist.com/ai-big-data-mineral-discoveries/ Date Accessed May 27, 2019
7. World's Largest Silver Mines: Suffer Falling Ore Grades & Rising Costs, Money metals - https://seekingalpha.com/article/4141630-worlds-largest-silver-mines-suffer-falling-ore-grades-and-rising-costs Date Accessed May 27, 2019
8. Evolution Of the Mount Woods Inlier, Northern Gawler Craton, Southern Australia: an Integrated Structural and Aeromagnetic Analysis, **Peter** Betts-Rick Valenta-Jim Finlay - Tectonophysics - 2003
9. Rock Mass Quality Q-System, Bhawani Singh, R.K. Goel, in Engineering Rock Mass Classification, 2011
10. World's Largest Silver Mines: Suffer Falling Ore Grades & Rising Costs, Money metals - https://seekingalpha.com/article/4141630-worlds-largest-silver-mines-suffer-falling-ore-grades-and-rising-costs
11. OZ minerals analyst visit presentation part 1 - exploration overview, marcel van eck hodgkison& jim, 2011
12. Gold and Silver Prospection Using Magnetic, Radiometryand Microgravity Methods in the Kıs¸ladag ˘ Province of Western Turkey https://onlinelibrary.wiley.com/doi/pdf/10.1111/rge.12024
13. Opengeohub Foundation, Envirometrix OpenGeoHub, Land GIS https://landgis.opengeohub.org/#/?base=Stamen%20(OpenStreetMap)&center=-30.7030,143.3260&zoom=6&opacity=80&layer=lcv_land.cover_esacci.lc.l4_c&time=2015

14. National Map https://nationalmap.gov.au/
15. Mining Companies Using Ai, Machine Learning And Robots, Alishba Imran-Alishba Imran - https://blog.prototypr.io/mining-companies-using-ai-machine-learning-and-robots-e6dcdebaccc3
16. Prominent Hill2018Mineral Resourceand Ore ReserveStatementand Explanatory Notes, OZ Minerals Limited https://www.ozminerals.com/uploads/media/181112_Prominent_Hill_Mineral_Resource_and_Ore_Reserve_Statement_as_at_30_June_2018.pdf