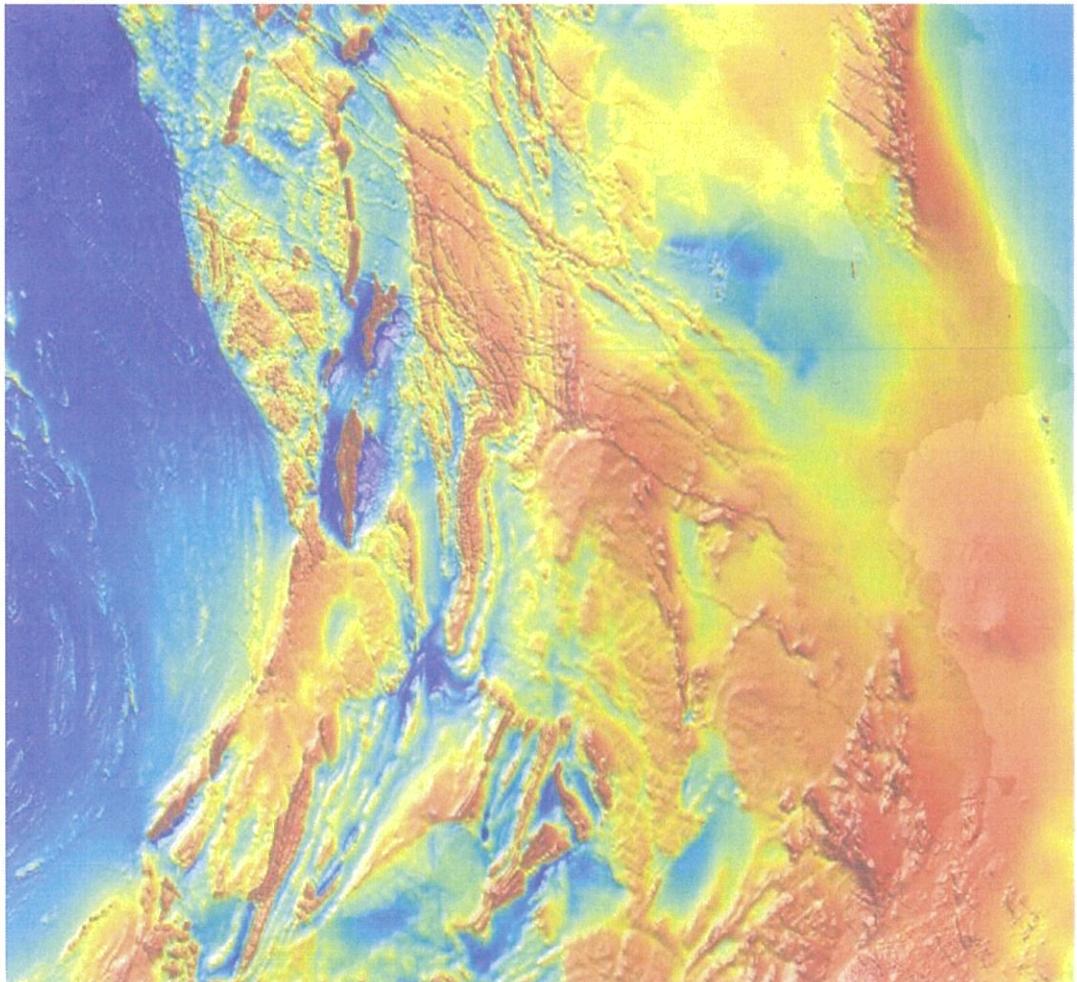


Title ↴

Data Driven Mineral Exploration



Evaluation of data mining and transformation methods predicting mineralisation using only gravity and magnetic maps in the Gawler Craton

} subtle
optimal

{

} team
members
names

Contents

1	Introduction	1
2	Data Summary	2
2.1	Target variable	2
2.2	Predictor variables	2
3	Literature Review	3
3.1	Focus of review	3
3.1.1	Weight of Evidence and Logistic Regression	5
3.1.2	Random Forest	5
3.1.3	Naïve Bayes	6
3.1.4	Support Vector Machine	6
3.1.5	Neural Networks	7
4	Methods and Transformations	9
4.1	Other methods trialled but results not reported	9
4.2	Dimensionality reduction	9
4.2.1	Factor analysis	10
4.2.2	First Principal Component of Clustered Variables	10
4.3	Binning and normalisation	11
4.3.1	Weight of Evidence (WoE) Binning	11
4.3.2	Min-Max by record layer	12
4.3.3	Z-score normalisation	12
4.4	Dealing with class imbalance	12
4.4.1	Oversampling techniques	12
4.4.2	Profit matrix as performance measure	13
4.5	Cross validation	13
4.6	Hyperparameters adopted for methods	14
4.6.1	Random Forest	14
4.6.2	Support Vector Machine	14
4.6.3	Convolutional Neural Network	15
4.7	Decision threshold	15
4.8	Ensemble model	16
5	Insights obtained on variable importance	16
6	Summary of Results	18
6.1	Relationship between class imbalance and profit	19
7	Conclusion	22
7.1	Recommendations and areas for future research	23
7.1.1	Circumstances where data drive approach should be useful	23
7.1.2	Areas of future research	24
	References	25

Tables

Table 1 - Dimensionality reduction techniques applied	9
Table 2 - Variable binning and/or normalisation.....	9
Table 3 - Factor loadings using the Minchi extraction method with Varimax rotation.....	10
Table 4 - Approximate binning for Cluster 1 of training set for Fold 1	11
Table 5 - Final binning for cluster 1 of training set 1	12
Table 6 - Profit matrix adopted as a performance measure.....	13
Table 7 - Summary of cross validation folds	14
Table 8 - Traditional SVM hyperparameter grid search range	14
Table 9 - One-Class SVM hyperparameter grid search range	15
Table 10 - Adjustments to Decision Thresholds.....	16

Figures

Figure 1 - Study area – Eastern Gawler Craton.....	1
Figure 2 - OCSVM for Outlier Detection (Source: Amer et. al.).....	7
Figure 3 - Grouping of variables into 10 clusters.....	11
Figure 4 – Relative importance of each variable type using Naive Bayes model	17
Figure 5 - Relative importance of each cell and layer in predicting mineralisation (logistic regression)..	17
Figure 6 - Distribution of total profit for various methods using 10-fold cross validation	18
Figure 7 - Distribution of profit rate for various methods using 10-fold cross validation	19
Figure 8 – Profit of ensemble model and “drill all” strategy by class imbalance	20
Figure 9 - Region of model's improvement over 'drill none' or 'drill all' strategies	20
Figure 10 - Profit for selected models vs. "drill all" strategy by class imbalance	21
Figure 11 - Profit rate of best model and "drill all" strategy by class imbalance.....	21
Figure 12 - Profit rate of selected models by class imbalance	22

1 Introduction

Mineral exploration is an expensive and risky endeavour offering infrequent large payoffs when mineralisation is discovered in sufficiently high concentration. Traditionally, the location of prospective drillholes is determined by experts through inference of subsurface composition. This inference is based on theories of geological and geochemical processes, developed after observations of surface materials, core drilling samples and maps of deviations in gravity and magnetic fields.

In the Eastern Gawler Craton, this observational data is scarce due to both overburden obscuring the basement geology west of the Flinders Ranges and relatively sparse core drilling samples over a vast area. Gravity and magnetic surveys are, however, available at high resolution throughout the study area.



Figure 1 - Study area – Eastern Gawler Craton

An alternative to predicting the location based on a geological model is to treat the mineral prospecting problem as a straightforward classification task with inputs and outputs. Given the available information, how can we best predict whether mineralisation will be present for an exploration drillhole at any given point on a map?

This report evaluates the comparative effectiveness of various classification and anomaly detection techniques for this task. In our case, we have limited our inputs to gravity and magnetic maps due to their completeness and availability.

Using a profit matrix to evaluate models, where the payoff from a successful drillhole was considered equivalent to the cost of drilling 50 holes, a wide range of predictive models were found to outperform naïve, default strategies of drilling either no drillholes or all drillholes in the study area. One caveat is that the relative performance of all models is dependent on the proportion of drillholes with mineralisation in the total set of drillholes being classified.

2 Data Summary

The data used in our analysis comprise mostly publicly available information sourced from the South Australian Resources Information Gateway (SARIG) website (Department of State Development 2017). This includes the location of known exploration drillholes and gravity and magnetic data relating to the study area. This data has been supplemented with the locations of known mineral deposits within the study area supplied directly by the Department of State Development.

Based on advice from the Department, from the variety of possible geophysical surveys, we have limited our predictive variables to the following layers:

- Gravity (WPA-SGRV-EGP) – hereafter “grav”
- Gravity 1VD (WPA-SGRV-EGP) – hereafter “grav1VD”
- Total Magnetic Intensity – TMI RTP (WPA-SGRV-EGP) – hereafter “mag”

The dataset represents 3,401 drillholes where the mineralisation status is known, due to drilling extending below the basement level and where gravity and magnetism readings are available. There are no known missing values in the dataset and no duplicates were detected.

2.1 Target variable

The target variable is binary indicating the presence (1) or absence (0) of mineralisation at each location. The dataset is highly imbalanced. Less than 2% of all known drillholes in the study area in mineralisation in commercial quantities. Compounding the class imbalance problem is that for practical purpose, we are far more interested in the accurate detection of the presence of mineralisation, rather than its absence. Our methods to mitigate this problem are outlined in Section 4.4 below.

An interesting question is how this ratio of successful/unsuccessful drillholes would generalise to all possible drillhole locations throughout the study area. Obviously, the location of each drillhole was not chosen at random but instead chosen by an expert to maximise the chance of finding minerals. Assuming their efforts were successful, as is likely, then our sample is biased. If our sample instead reflected locations drilled at random then a lower success rate would be expected. This bias in our sample may have implications for the usefulness of our model and this is discussed further in Section 6.1 **Error! Reference source not found.** below.

2.2 Predictor variables

The predictor variables each measure deviation in gravity or magnetism at a point directly above the drillhole or at surrounding locations. The `grav1VD` variable is derived from the rate of change in the `grav` variable. We also understand that the `mag` variable is adjusted to account for the direction of the magnetic field and height above ground that the reading was taken. Both `grav1vd` and `mag` had been calculated in the dataset supplied to us and did not form part of our analysis.

The resolution of the gravity and magnetic layers is typically approximately 35 by 35 metres.

All readings are from a continuous scale with both positive or negative values possible. As would be expected for spatial data, the gravity and magnetic values of neighbouring points are highly correlated. The effect is that there is vastly more variation in gravity between points corresponding to different drillholes than between points surrounding drillholes.

The standard set of predictor variables available to each model comprised 25 of the `grav`, `grav1VD` and `mag` variables; hence 75 in total for each drillhole. Each subset of 25 represents measurements from a 5 by 5 grid of points on and surrounding the location of drillhole for which the class is predicted.

The convolutional neural network model takes inputs from a larger 32 by 32 grid of points, hence 3,072 input variables in total per drillhole.

The support vector machine model takes as input only the value of the point directly above the drillhole.

3 Literature Review

We reviewed literature specifically on the application of machine learning on mineral exploration.

We have reviewed the literature to determine:

- The extent to which data mining has already been used to improve the accuracy of mineral prospecting;
- Those techniques which are considered most effective at predicting the location of mineral deposits using gravity and magnetism maps as inputs; and
- Data transformations which improve models' predictive accuracy.

3.1 Focus of review

Porwal and Carranza (2015) make a distinction between data-driven and knowledge driven mineral prospecting models. Knowledge driven models have parameters determined using expert knowledge. Data driven models find spatial statistical relationships between known deposits and predictor features. The literature to date suggests that data driven models can outperform knowledge driven models in situations where multiple predictive variables are available (Harris et al. 2015; Porwal & Carranza 2015; Rodriguez-Galiano et al. 2015). Our research is concerned with data-driven models.

Porwal and Carranza (2015) further classify data-driven models into four separate approaches:

- Probabilistic,
- Regression based,
- Artificial Intelligence (AI) based, and

Referencing in text³
e.g. Lastname (YYYY)

Lastname 1 and
Lastname 2 (YYYY,
et al.)
year

- Dempster-Shafer belief theory based.

Our project is limited to probabilistic, regression based and artificial intelligence (AI) based methods. The successful drill holes have been classified into ordinal categories describing their scale and ordinal regression was considered as a method to predict these categories. Ultimately, due to the small number of successful training examples overall (67), there are also very few examples of each ordinal category. For this reason, we have excluded regression based methods and focussed on a binary classification of successful or unsuccessful. Of the AI methods, Porwal and Carranza (2015) distinguish fuzzy set based expert systems from machine learning systems including neural networks, decision trees, support vector machines and random forests. Our research on AI methods excludes fuzzy set based expert systems as we lack expert geological knowledge.

Harvey and Fotopoulos (2016) discuss the use of machine learning techniques to infer geological conditions although this is beyond the scope of our research.

Granek (2017) summarises the problem characteristics of applying machine learning methods to mineral prospectivity mapping as:

- **Limited training data** - the area which has been explored is very small relative to the area unexplored. The limited sample size makes variance within the positive training data more important. The ability of different techniques to deal with limited training data is studied by Carranza and Laborte (2015) and Dutta et al. (2010) although the latter was to estimate the scale of a known ore body;
- **Imbalanced training set** - The number of drill holes successful in detecting mineral deposits at a commercial scale is generally quite small relative to the number of unsuccessful drill holes. This can lead to lazy models which target the majority class and fail to generalise;
- **Uncertain training labels** - The method of classification of drill holes as successful or unsuccessful is often arbitrary. A classification of "No Mineralisation" does not necessarily mean that none exists (perhaps the drill was not deep enough). Our method attempted to mitigate this by considering only drill holes which intersected the basement, although it is possible that mineralisation could occur well below basement level;
- **Uncertain training data** - There is more uncertainty associated with the input data than other fields, for example those associated with detection limits and processing procedures for magnetic field measurement; and
- **Data variety** - Input data can include disparate data types including sensors, spatial data, geological interpretation of age, rock type, primary minerals and fault structure.

We have chosen a variety of methods described in the literature based on their ability to deal with these challenges.

3.1.1 Weight of Evidence and Logistic Regression

Logistic regression is a widely used learning algorithm effective in both binary or categorical classification problems (Allison 2012; Cox & Snell 1989). A logistic regression model outputs a likelihood occurrence output ranging from 0 to 1. When applied to our problem, the model provides an estimate of the probability of a drillhole returning mineralisation. The technique works by transforming the predictor variable and modelling a linear relationship between target and predictor variables, using the log odds function below:

$$\text{Log odds} = \log\left(\frac{P(\text{drillhole is successful})}{P(\text{drillhole is unsuccessful})}\right)$$

Subsequently, a regression model is built to generate estimated coefficients of each predictor. The logistic regression learning process is strongly related to maximising likelihood while observing the relationship between predictor and target variables.

(Schaeben & Semmler 2016) in their research article “The quest for conditional independence in prospectivity modelling: weights-of-evidence, boost weights-of-evidence, and logistic regression” confirmed that Weight of Evidence (WoE) is a well-established method used by geologists, specifically in modelling potential maps of mineral prospects. However, the general weight of evidence calculation assumes conditional independence between the predictors. In other words, when the conditional independence assumption is not satisfied, the model tends to output poor results. This tendency is mitigated when using logistic regression, which is well-known for its robustness to correlated features.

In practice, WoE is a binning transformation method applied to predictor variables. By transforming continuous predictors into WoE values, they are standardised to obtain a weight scale before feeding into logistic regression model. Observations represented by each WoE value bin are assumed to have similar relationship to the target variable.

The formula for WoE involves calculation of the conditional probability, as follows:

$$WoE = \ln\left(\frac{\% \text{ of non-events}}{\% \text{ of events}}\right)$$

WoE explicitly transforms variables to a logistic scale. In some cases, where distribution of bad and distribution of goods are of the same number, the WoE value will be zero. In this case, the respective category bin it represents can be interpreted as undefined or insignificant. Thus, it can be merged with the other bin or discarded.

3.1.2 Random Forest

Random Forest™ is a trademark of Leo Breiman and Adele Cutler (2001). The Random Forest classifier is a type of ensemble modelling, which combines the decisions of different predictive models to improve prediction accuracy. Random Forest builds many single Decision Trees, combines each tree’s decision and output the mode of the predicted class for each observation.

This method is not frequently mentioned in the literature as a method for mineral prospectivity mapping although Carranza and Laborte (2015) note that the method is well suited due to its ability to elegantly handle missing values and can be adapted for spatial data by using distances to geological features as inputs.

3.1.3 Naïve Bayes

This classifier is based on the Bayes Theorem with a strong (naïve) assumption of independence. The Bayes Theorem using multiple predictor variables is expressed as follows:

$$P(H|E_1, E_2, \dots, E_n) = \frac{P(E_1, E_2, \dots, E_n|H)P(H)}{P(E_1, E_2, \dots, E_n)}$$

In the above equation H represents the predictor variable and E represents evidence. With the independence assumptions, the theorem can be written as:

$$P(H|E_1, E_2, \dots, E_n) = \frac{P(E_1|H)P(E_2|H) \dots P(E_n|H)P(H)}{P(E_1, E_2, \dots, E_n)}$$

A new tuple is classified as a specific class if it maximises the posterior probability $P(H|E_1, E_2, \dots, E_n)$.

There are some benefits of using the Naïve Bayes models. They are robust to noise as the probabilities are calculated from all the data. Further, the models work well if there are irrelevant attributes as they do not impact the posterior probability calculation. However, as the model assumes independence between the predictor variables, it does not perform well if the attributes are correlated.

3.1.4 Support Vector Machine

Traditional SVM

Support Vector Machines (SVM) are discriminative classifiers that transform data and based on these transformations, obtain boundaries or hyperplanes to separate data into groups with respect to the classes of the target variable. When new data is introduced into the models, these hyperplanes are used to classify the new observations accordingly.

Zuo and Carranza (2011) investigate the use of Support Vector Machine models to predict locations of mineral deposits and achieved very good results but in an area with good geological and geochemical input data. The geochemical data was found to be more predictive than geological data. However, in our study we focus primarily on gravitational and magnetism data.

An SVM, and other algorithms can incorporate spatial data but in a more rudimentary way, creating a matrix of neighbouring locations of size $n \times md$ where n is the number of sample locations, m is the number of neighbouring points and d is the number of data channels. The downside is that the size of the data grows rapidly.

One-Class SVM

One-Class SVM is an anomaly detection technique, also useful in classification problems with extreme class imbalance. Schölkopf et al. (2000) in their research proposes One-class Support Vector Machines (OCSVM) for anomaly detection. Their research overcomes the limitation of traditional SVM, which has difficulty in dealing with outliers and highly imbalanced classes. OCSVM is a semi-supervised technique as model training is performed on the normal class only and hence its name. OCSVM learns a decision function and similar to traditional SVM, applies a kernel function to separate the “class” and “not the class”. The main difference between the One-class SVM and traditional SVM is that the origin is set as the reference point and outliers are data points closer to the origin in a higher dimensional feature space. Figure 2 - OCSVM for Outlier Detection (Source: Amer et. al.) illustrates how OCSVM detect outliers through “rotation” of data into higher dimensional space (Amer, Goldstein & Abdennadher 2013).

To the best of our knowledge, there are no literature on the applications of OCSVM in the mineral exploration domain. However successful OCSVM applications include,

- The winning solution for Task 2 of KDD Cup 2002 Yeast Gene Regulation Prediction (Kowalczyk & Raskutti 2002) in which the researchers detail OCSVM outperforms traditional classifiers where the classes are heavily unbalanced.
- One-Class SVMs for Document Classification (Manevitz & Yousef 2001) in which the researchers found that OCSVM method is superior than other classifiers when trying to identify positive examples in the absence of negative examples using the standard Reuters data set.
- Support Vector Novelty Detection Applied to Jet Engine Vibration Spectra (Hayton et al. 2001) in which the researchers successfully used OCSVM to detect abnormal jet engine vibrations using normal jet engine vibrations in the absence of rare abnormal jet engine vibrations.

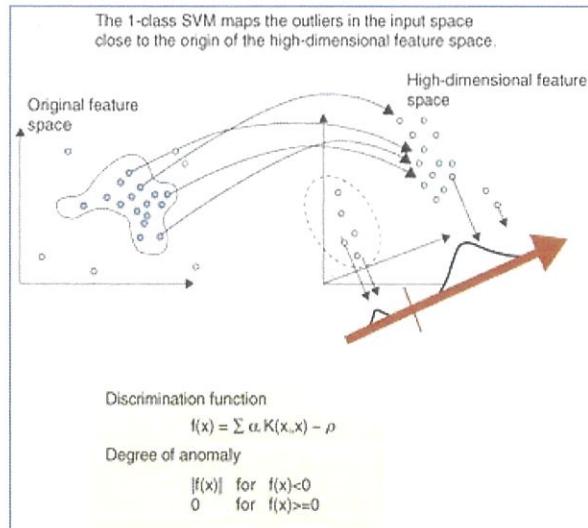


Figure 2 - OCSVM for Outlier Detection (Source: Amer et. al.)

3.1.5 Neural Networks

Neural Network classifiers try to emulate the human brain in solving machine learning problems. In its simplest form, the predictors are input nodes connecting to hidden neurons within the hidden layer(s), which process a set of Boolean variables, and output the results to

the set of output nodes (target classes). The hidden neurons are “fired by” the strength of the signals carried by connections between the input nodes and the hidden neurons.

Variants of neural networks have been shown in various papers to have good success in interpreting spatial patterns in gravity and magnetic maps (Albora et al. 2001; Granek 2017; Rodriguez-Galiano et al. 2015; Vallance 2016). Convolutional neural networks have recently emerged as the most effective technique for image classification problems which, like our problem, use spatial patterns in matrices to make predictions.

Granek’s (2017) demonstration of the use of a convolutional neural network for mineral prospectivity mapping is highly relevant to our problem where the usefulness of geochemical data is limited. By using only 3 layers (gravity, magnetic and fault lines) the model achieved over 80% success rate on the validation data. A support vector machine was also used but this had access to 91 layers of data.

Albora et al. (2001) similarly used a cellular neural network to identify iron deposits only using Bouger gravity anomaly maps. The results better distinguished known iron ore deposits than traditional gravity derivative maps.

4 Methods and Transformations

The predictive models applied in this paper, and the respective dimensionality reduction techniques and variable transformations applied are summarised in Table 1 and Table 2 below:

Classification technique	First Principal Component of Clustered Variables	Factor Analysis
Logistic regression 1	✓	
Logistic regression 2		✓
Random forest		✓
Naïve Bayes		✓
Support vector machine		
Convolutional neural network		

Table 1 - Dimensionality reduction techniques applied

Classification technique	WoE binning	Min-max by record/layer	Z-score
Logistic regression 1	✓		
Logistic regression 2			
Random forest			
Naïve Bayes			
Support vector machine			✓
Convolutional neural network		✓	

Table 2 - Variable binning and/or normalisation

Each transformation technique is described in further detail below.

4.1 Other methods trialled but results not reported

Other methods trialled but not reported in detail due to unpromising results include:

- Neural network using 5x5 grid implemented with the H2O.ai package

4.2 Dimensionality reduction

The default set of predictor variables comprise three types of geophysical readings (gravity, gravity1VD and magnetism) from a 5 x 5 grid of locations above and surrounding each drillhole. The total number of raw predictor variables was therefore 75, comprising:

$$\# \text{ of Predictor variables} = 3 \text{ layers} \times 25 \text{ readings} = 75$$

Many classification techniques, including some adopted in this report, do not work well with high dimensional data. To address this problem, one of two dimensionality reduction techniques have been adopted for all models excluding the support vector machine and the convolutional neural network models:

1. Factor analysis; or
2. Clustering with principal component analysis to determine the cluster prototype

The support vector machine model was trained with only one variable from each layer. The convolutional neural network is tolerant of high dimensionality and takes inputs from a larger 32 by 32 grid of points, hence 3,072 input variables in total.

4.2.1 Factor analysis

Factor analysis is often used to find latent variables which are difficult to measure. Factor analysis gathers the common features into the descriptive groups. Another benefit of the technique is dimensionality reduction and for this reason we have used it with some of our classification techniques.

The implementation we have adopted uses the Min Chi Factoring technique with a Varimax rotation (Chi et al. 2011). The Min Chi algorithm minimizes the sample size weighted chi-square when treating pairwise correlations with a different number of subjects per pair. The outcomes of factor analysis in the first iteration are illustrated in Table 1. In fact, the results of all ten iterations are just similar. There are three common factors that explain all 75 variables.

Factor Loadings	Factor		
	1	2	3
Raw Gravity 1		0.73	
Raw Gravity 2		0.74	
⋮		⋮	
Raw Gravity 25		0.73	
1VD Gravity 1			0.66
1VD Gravity 2			0.66
⋮			⋮
1VD Gravity 25			0.66
Magnetism 1	0.71		
Magnetism 2	0.70		
⋮	⋮		
Magnetism 25	0.70		

Table 3 - Factor loadings using the Minchi extraction method with Varimax rotation

All variables belonging to the same type of data are loaded to the same factor. Therefore, the names of three factors can be given as the Magnetism, Raw Gravity and 1VD Gravity, respectively.

4.2.2 First Principal Component of Clustered Variables

This method makes use of the ‘ClustOfVar’ package in R (Chavent et al. 2012) to reduce data dimensionality for modelling while minimising loss of information from all variables in the data. The main objective is to cluster groups of variables, each with strong relationship to each other, hence likely to represent similar information (Chavent et al. 2012). The method groups similar variables using the k-means algorithm. Principal Components Analysis (PCA) is undertaken within each cluster and first principal component is used as a prototype to represent all members of the cluster. Hence the procedure reduces the 75 input variables to 10.

The clusters found in the training data are shown in Figure 3 below. Each cluster consists of same type of data but from cells in different positions relative to the drillhole.

Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Variable	Correl	Variable	Correl	Variable	Correl	Variable	Correl	Variable	Correl
mag_3	1.00	mag_17	1.00	mag_19	1.00	grav_3	1.00	grav_9	1.00
mag_8	1.00	mag_16	1.00	mag_20	1.00	grav_2	1.00	grav_14	1.00
mag_2	1.00	mag_18	1.00	mag_25	1.00	grav_4	1.00	grav_8	1.00
mag_9	1.00	mag_12	1.00	mag_14	1.00	grav_1	1.00	grav_13	1.00
mag_4	1.00	mag_22	1.00	mag_24	1.00	grav_5	1.00	grav_15	1.00
mag_7	1.00	mag_23	1.00	mag_15	1.00			grav_10	1.00
mag_10	1.00	mag_11	1.00					grav_7	1.00
mag_1	1.00	mag_21	1.00						
mag_6	0.99	mag_13	1.00						
mag_5	0.99								
Cluster 6		Cluster 7		Cluster 8		Cluster 9		Cluster 10	
Variable	Correl	Variable	Correl	Variable	Correl	Variable	Correl	Variable	Correl
grav_11	1.00	grav_24	1.00	grav1VD_6	0.99	grav1VD_9	0.99	grav1VD_23	0.98
grav_16	1.00	grav_19	1.00	grav1VD_7	0.99	grav1VD_4	0.98	grav1VD_18	0.97
grav_17	1.00	grav_23	1.00	grav1VD_1	0.97	grav1VD_10	0.97	grav1VD_22	0.97
grav_12	1.00	grav_18	1.00	grav1VD_12	0.97	grav1VD_8	0.97	grav1VD_17	0.97
grav_6	1.00	grav_25	1.00	grav1VD_2	0.96	grav1VD_14	0.97	grav1VD_19	0.96
grav_21	1.00	grav_20	1.00	grav1VD_11	0.96	grav1VD_5	0.96	grav1VD_24	0.96
		grav_22	1.00			grav1VD_3	0.96	grav1VD_21	0.93
						grav1VD_13	0.95	grav1VD_25	0.92
						grav1VD_15	0.95	grav1VD_20	0.92
								grav1VD_16	0.91

Figure 3 - Grouping of variables into 10 clusters

4.3 Binning and normalisation

4.3.1 Weight of Evidence (WoE) Binning

All attributes were discretised to increase speed of the iterative model building. The first stage of the WoE binning process is to provisionally assign each of the 10 variables (following dimensionality reduction) to a bin. Then the WoE value is calculated for each bin, as shown in Table 4 below.

Segment number	from	to	woe	iv
1	-11.16352	-0.893963	0.8484778	6.795771
2	-3.893963	-0.680898	-0.1898487	0.359501
3	-2.680898	-1.825958	-0.6159052	3.680191
4	-1.825958	-1.071555	-0.657697	4.171929
5	-1.071555	-0.3488391	-0.5794043	3.268243
6	-0.3488391	0.4908	-0.2711794	0.7312525
7	0.4908	1.503269	-0.2409943	0.577284
8	1.503269	2.946137	0.2603569	0.6741918
9	2.946137	5.75437	0.6711392	4.342533
10	5.75437	28.85832	0.8301809	6.51038

Table 4 - Approximate binning for Cluster 1 of training set for Fold 1

And then, the segmentations with similar WoE values were merged since they have similar prediction power. By repeating this process, the 10 clustering variables in the 10 training sets

were each separately binned. Since the binning process maximises the weight of evidence based on the training data, binning categories vary each time a new model is trained.

Segment number	from	to	woe	iv
1	-11.16352	-.893963	0.8484778	6.795771
2	-.893963	1.503269	-.4219014	10.52424
3	1.503269	28.85832	0.5793651	9.792432

Table 5 - Final binning for cluster 1 of training set 1

Finally, we used WoE value to replace the segmentations as transformed attributes. In the end, this makes the predictors representation shift from the original variables value to a WoE value.

4.3.2 Min-Max by record layer

This transformation is done specifically for the convolutional neural network model. As noted in Section 2.2, the input data has high levels of spatial autocorrelation. To counter this, and focus the analysis on local patterns, the min-max standardisation was performed within each grid layer for each record. For example, the maximum of all values of `grav` for points surrounding a drillhole is scaled to 1 and the minimum scaled to 0. All other `grav` input variables are scaled to be within this range. The effect is to standardise and maximise the variance between variables in each layer, enhancing the convolutional neural network's ability to find spatial patterns in each grid layer.

4.3.3 Z-score normalisation

This transformation is undertaken only on the support vector machine model. The mean and standard deviation of each variable is calculated, then within each variable the mean is subtracted and the remainder is divided by the standard deviation. The normalised variables each have a mean of zero and a standard deviation of 1.

4.4 Dealing with class imbalance

4.4.1 Oversampling techniques

Random oversampling examples (ROSE)

The Logistic Regression 1 model uses a hybrid resampling technique in R, namely Random Over-Sampling Example (ROSE), which was implemented on the training data. ROSE is known as a package in R which generates artificial balanced samples using smoothed bootstrap approach (Lunardon, Torelli & Menardi 2014).

Synthetic minority oversampling technique (SMOTE)

Like ROSE, this approach attempts to create a more balanced training set by synthetically creating an enlarged set of the minority class. The difference is that while ROSE creates synthetic values in the space around the minority class, SMOTE instead creates values on a line between the minority class and its nearest neighbour (Chawla et al. 2002). It has been adopted by all other methods except the Logistic Regression using clustered variables.

4.4.2 Profit matrix as performance measure

Another method of overcoming the tendency of classification models trained on imbalanced classes to over-predict the majority class is to use a profit matrix giving a higher reward to correct classification of the minority class. In our case, this method has the added benefit of better reflecting the true economics of mining exploration. The profit matrix adopted is summarised in Table 6 below:

Profit Matrix		Predict	
		Successful	Unsuccessful
Actual	Successful	(TP) 50	(FN) 0
	Unsuccessful	(FP) -1	(TN) 0

Table 6 - Profit matrix adopted as a performance measure

If the model predicts that a drillhole will be unsuccessful, then the simulated outcome is that drilling does not occur, hence both the cost and return is zero, representing the two outcomes on the right of the profit matrix.

If the model predicts that a drillhole will be successful, then the payoff of a successful prediction is 50 times greater than the cost of an unsuccessful prediction. Our rudimentary enquiries indicate that this ratio is roughly accurate, although the payoff and cost are of course highly variable.

The adopted ratio also results in a result very close to break-even if all potential drillholes in our dataset are drilled, given there are 67 successful drillholes out of 3,401;

$$\text{Revenue if all holes drilled} = 67 \times 50 = 3,350$$

$$\text{Cost if all holes drilled} = 3,401 \times 1 = 3,401$$

$$\text{Profit if all holes drilled} = 3,350 - 3,401 = -51$$

Our methods will each attempt to maximise profit through a combination of high recall of drillholes containing mineralisation and high precision avoiding the cost of falsely predicting mineralisation where it does not exist.

A related performance measure is the profit rate. This enhances the realism of the profit measure by dividing the total number of holes drilled (predictions made). This additional calculation accounts for the capital required to be put at risk by a miner to achieve their simulated profit. If all holes in the dataset are drilled, using our profit matrix, the profit would be close to zero (-0.15%). Our models will attempt to maximise the profit rate by more selectively choosing the most likely locations to drill, thereby minimising cost.

4.5 Cross validation

Although the dataset is not small, comprising 3,401 records, within these there are very few examples of the successful drillholes, only 67. To better demonstrate each methods ability to generalise on new data, each model is trained using 10-fold cross validation.

Stratified sampling was deliberately not adopted so the class imbalance varies in each fold. This provides useful information on the relative performance of each method in differing conditions of class imbalance. The size and class imbalance of each of the 10 folds is summarised in Table 7 below.

	Number of drillholes	#/% of successful drillholes
Fold 1	352	7 / 1.99%
Fold 2	319	5 / 1.57%
Fold 3	322	6 / 1.86%
Fold 4	337	9 / 2.67%
Fold 5	331	8 / 2.42%
Fold 6	353	2 / 0.57%
Fold 7	339	5 / 1.47%
Fold 8	350	12 / 3.43%
Fold 9	373	4 / 1.07%
Fold 10	325	9 / 2.77%
Total	3,401	67 / 1.97%

Table 7 - Summary of cross validation folds

A desirable feature of a predictive model would be an ability to adjust its decision threshold in sparse regions although this would require prior knowledge of the extent of mineralisation, so is not possible.

4.6 Hyperparameters adopted for methods

No hyperparameters are required to be set for logistic regression, or naïve Bayes. The hyperparameters for the random forest, support vector machine and convolutional neural network models are described below.

4.6.1 Random Forest

The random forest model was limited to 200 trees.

4.6.2 Support Vector Machine

Traditional SVM

The hyperparameters for the support vector machine were optimised using a grid search over the following range:

Parameters	Values
SVM type	C-svc
Kernel type	Linear, Sigmoid, Radial Basis Function
Gamma	0 to 1,000 in 10 steps
C	0 to 100 in 10 steps

Table 8 - Traditional SVM hyperparameter grid search range

The grid search involved testing 363 possible combinations of hyperparameter in each fold. Consequently, the optimal hyperparameters varied in each fold. The radial basis function

kernel was consistently selected. Gamma was usually 1,000 but sometimes 900 and infrequently 800. The cost function varied between 40 and 100.

One Class SVM

As an alternative to the traditional SVM, a one-class SVM model was trialled. Similar to the traditional SVM, a grid of possible hyperparameters was searched to optimise profit and applied to the test set. The range searched is as follows:

Parameters	Values
SVM type	One-class
Kernel type	Sigmoid
Gamma	0 to 2 in 20 steps
C	1×10^{-9} to 1 in 20 steps

Table 9 - One-Class SVM hyperparameter grid search range

The grid search involved testing 441 possible combinations of hyperparameter in each fold. The optimal value for gamma was zero in all but one fold where it was 0.1. The value of C varied from 1×10^{-9} to 0.16.

4.6.3 Convolutional Neural Network

The ‘tensorflow’ and ‘keras’ packages implemented in R were used to develop the model. The model architecture comprises in order:

1. 2 x 2-dimensional convolutional layers, each with 32 feature maps with a size of 3 by 3 and a rectifier activation function;
2. 1 x maximum pooling layer to reduce overfitting with size 2 by 2;
3. 2 x 2-dimensional convolutional layers, each with 32 feature maps with a size of 3 by 3 and a rectifier activation function,
4. 1 x additional maximum pooling layer with size 2 by 2;
5. 1 x dense fully connected layer with 512 nodes and rectifier activation function, and
6. Final binary (2 unit) output layer with a softmax nonlinear activation function.

The model uses a stochastic gradient descent optimiser with a learning rate of 2.5×10^{-4} , momentum of 5×10^{-6} and decay of 1×10^{-6} . The loss function adopted is binary cross-entropy. Image data augmentation, including rotation of layers was not used to maintain consistent polarity for the magnetic layer.

Each model was trained in batches of 32 records at a time over 300 epochs.

The model’s architecture is chosen due to its proven success on classification of small colour images with 32 by 32 pixel resolution.

4.7 Decision threshold

The decision thresholds adopted for each model’s probabilities to determine class membership are summarised as follows:

	Decision threshold (profit)	Decision threshold (profit rate)
Logistic Regression 1	32%	66.27%
Logistic Regression 2	60%	93%
Naïve Bayes	51%	98%
Random Forest	34%	78%
Support Vector Machine	50%	50%
Convolutional Neural Network	50%	50%

Table 10 - Adjustments to Decision Thresholds

4.8 Ensemble model

For the ensemble model, a simple majority voting scheme was adopted. If more than three models indicated mineralisation then the ensemble predicted the same, otherwise the ensemble predicted no mineralisation.

The models in the ensemble optimised for highest **profit** were:

- Weight of Evidence Logistic Regression with decision threshold of 32%;
- Weight of Evidence Logistic Regression with decision threshold of 66.27%;
- Naïve Bayes using factor analysis for dimensionality reduction with all data with decision threshold of 51%;
- Naïve Bayes as above but using only magnetic and raw gravity data with decision threshold of 51%;
- Logistic Regression with same data as Naïve Bayes and decision threshold of 60%; and
- Random Forest with the same data and decision threshold of 34%.

The models in the ensemble optimised for highest **profit rate** were:

- Weight of Evidence Logistic Regression with decision threshold of 66.27%;
- Naïve Bayes using factor analysis for dimensionality reduction with all data with decision threshold of 98%;
- Naïve Bayes as above but using only magnetic and raw gravity data with decision threshold of 51%; and
- Logistic Regression with same data as Naïve Bayes and decision threshold of 60%.

5 Insights obtained on variable importance

Only two models provide insight into the relationship between the target variable and the predictor variables; logistic regression and naïve Bayes. Both were fairly consistent in identifying the `mag` and `gravIVD` layers being more predictive of the presence of mineralisation than the `grav` layer although to varying degrees.

Figure 4 shows the variable importance of the Naïve Bayes model for the training data predicting the first fold (so only approximately 90% of the data). The analysis indicates that both magnetism and the first vertical derivative gravity are important predictors and raw gravity is much less significant. Analysis from other folds indicated very similar results.

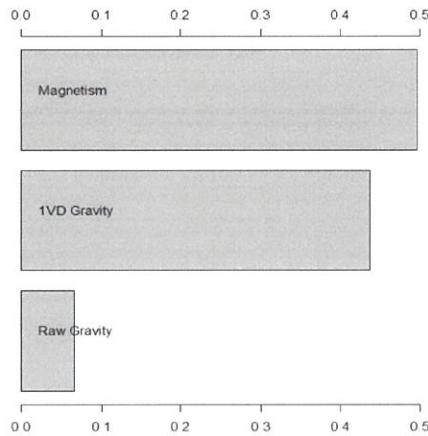


Figure 4 – Relative importance of each variable type using Naive Bayes model

The clustering approach used by logistic regression identified groups of contiguous cells on and around the drillhole location which were correlated. The variable importance of each of these clusters suggest interesting spatial patterns in Figure 5 below.

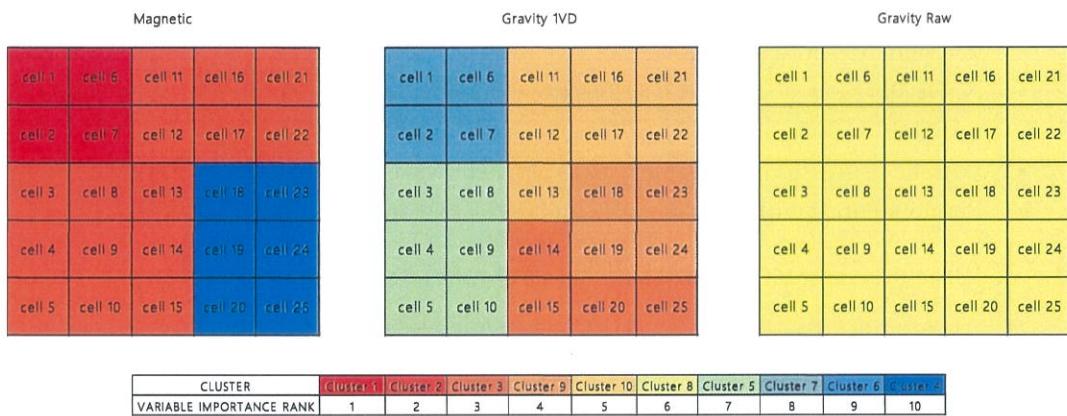


Figure 5 - Relative importance of each cell and layer in predicting mineralisation (logistic regression)

The magnetism readings for cells situated to the north-west of a proposed drillhole are considered the most predictive of the presence of mineralisation. Interestingly, the magnetic readings for cells to the south east are the least predictive, also relative to both gravity layer readings. We lack sufficient geological expertise to provide a plausible theory for this effect although it may be an artefact of processing procedures for magnetic field measurement as identified by Granek (2017). The remainder of the cells in the magnetic grid, excluding the south-east corner are also highly predictive of mineralisation although not to the same extent as the north-west.

The spatial pattern is inverted for the gravity 1VD data with cells to the south-east being more highly predictive than cells to the north-west. The raw gravity readings are all moderately predictive of mineralisation.

6 Summary of Results

As discussed in Section 4.4.2 above, the use of a profit matrix as a performance measure assists to overcome two major problems; firstly, the higher class imbalance between successful and unsuccessful drillholes, and secondly, the differing weight miners attach to accurately predicting the presence of mineralisation rather than its absence.

Using profit as a measure, we attribute:

- 50 units to correct classifications of drillholes with mineralisation, and;
- -1 unit to drillholes incorrectly classified as having mineralisation present when none exists

Each models' results, presented as box plots of the 10 folds in cross validation are summarised in Figure 6 below. The "Drill All" boxplot is a baseline measurement indicating the profit using the naïve strategy of drilling all holes in the study area.

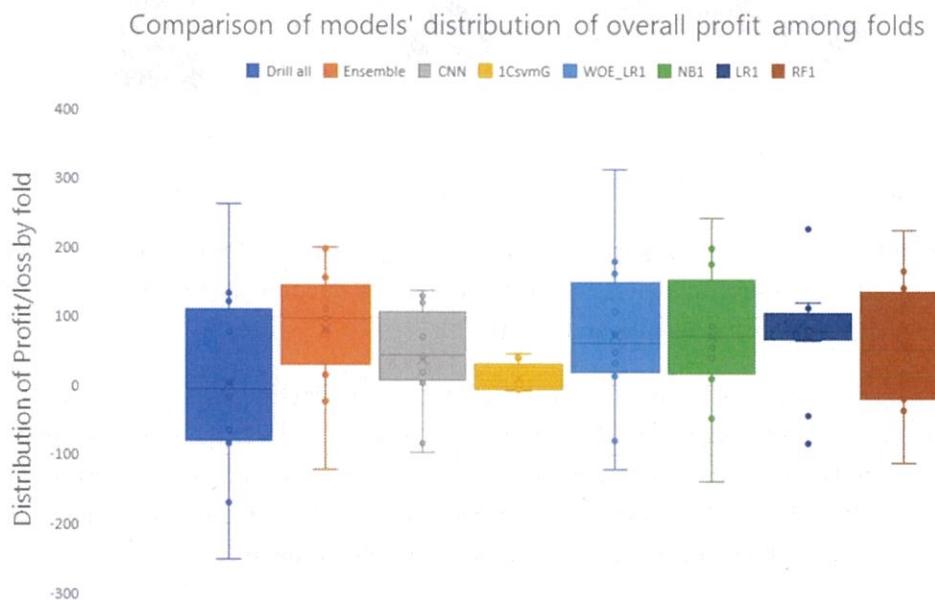


Figure 6 - Distribution of total profit for various methods using 10-fold cross validation

Figure 9 shows that compared to the naïve 'drill all' strategy, represented by the leftmost boxplot, with a mean and median profit of approximately zero, a variety of other models performed better. The logistic regression model using variable clustering for dimensionality reduction and weight of evidence binning (WOE_LR1) achieves the highest profit in one single fold although the ensemble method has the highest mean and median profit. The one-class SVM appears to offer a conservative low-risk/low-return strategy in contrast to the other models.

Another measure is the profit rate which also considers the amount of capital required to be put at risk during exploration. The results using this measure are summarised in Figure 7 below.

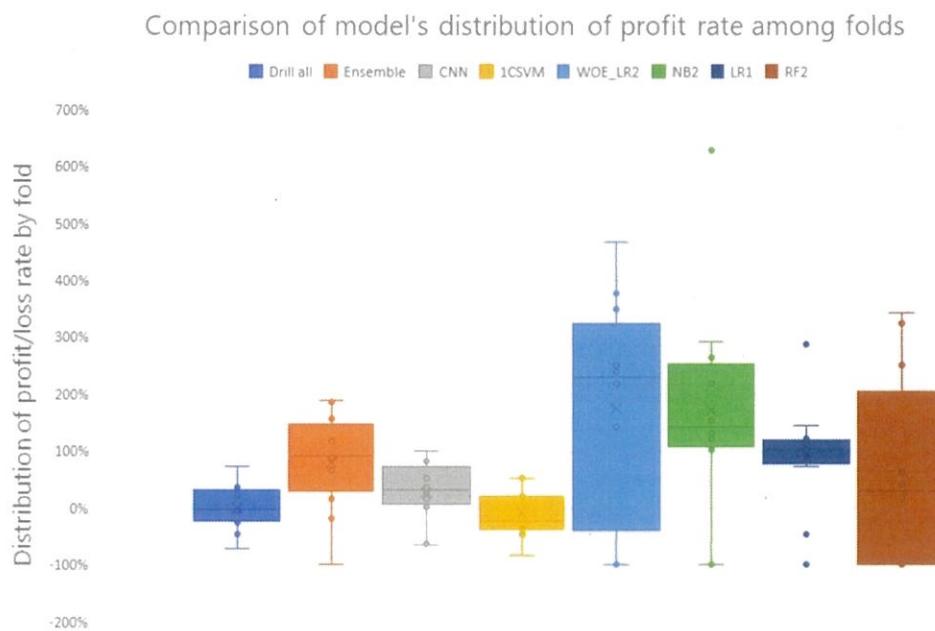


Figure 7 - Distribution of profit rate for various methods using 10-fold cross validation

The use of a profit rate instead of total profit has decreased the variance for the baseline “drill all” strategy as all profits are divided by the largest possible denominator relative to other strategies (the total number of drill holes in the set).

Again, there is a wide variety of models which all seem to outperform the base strategy in profit rate most of the time. The highest median profit is clearly WOE_LR2, the same model as WOE_LR1 described above, but with the decision threshold tuned to achieve the highest profit rate. The downside of this model is its risk, with the lower band of the interquartile range less than zero indicating that the model makes a loss more than 25% of the time.

The Naïve Bayes model (NB2), also with a decision threshold tuned for this purpose, also has a strong median profit but with less risk. The LR1 model, representing logistic regression on variables using factor analysis, has a remarkably low variance with more than half of models achieving close to 100% profit and one outlier achieving more than 600% profit, although on 2 out of 10 folds the model results in a loss.

6.1 Relationship between class imbalance and profit

As expected, there is an almost perfectly linear relationship between the performance of the profit/loss using the “Drill All” strategy and the proportion of successful drillholes in the dataset. This is a function of the profit matrix where the payoff from successfully finding mineralisation offsets the cost of drilling if more than 1 in 50 holes (2%) of drillholes

investigated contain mineralisation. One of the best performing model is simple majority voting ensemble of the seven highest-performing individual models. The improvement in expected profit using this model over a “Drill All” strategy is shown in Figure 8 below.

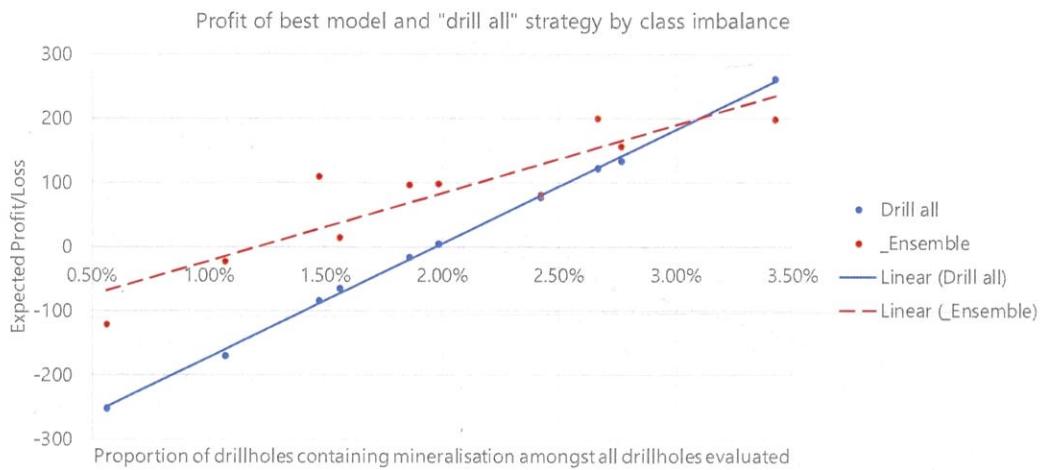


Figure 8 – Profit of ensemble model and “drill all” strategy by class imbalance

This chart suggests that using the ensemble model will, on average, reduce a miner’s loss and increase their profit in regions where the expected success rate is lower than 3% of all holes drilled. In the dataset supplied, the average success rate is close to 2%. These percentages are of course completely dependent on the ratio adopted in the profit matrix.

One way of interpreting this plot is to consider that the triangular shaded region bounded by the zero line, representing a “drill none strategy”, the linear fit to the “drill all strategy” and the linear fit to the model provides an estimate of the model’s estimated improvement over either strategy.

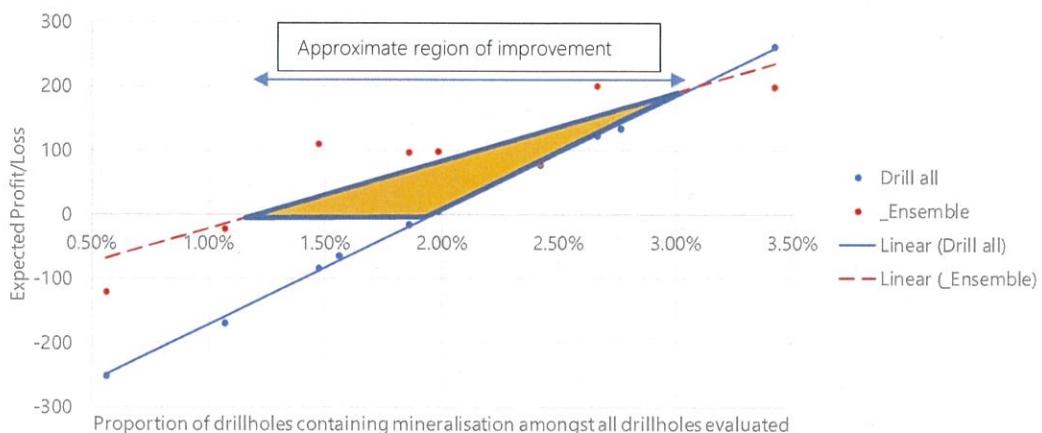


Figure 9 - Region of model’s improvement over ‘drill none’ or ‘drill all’ strategies

This plot indicates that the model is most likely to improve performance when the proportion of drillholes evaluated is between approximately 1% and 3%.

Under 1%, a better strategy would be to not drill at all while over 3% the best strategy would be to drill all. The highest expected improvement is towards the middle of this range.

A similar plot with a larger selection of models is shown in Figure 10 below.

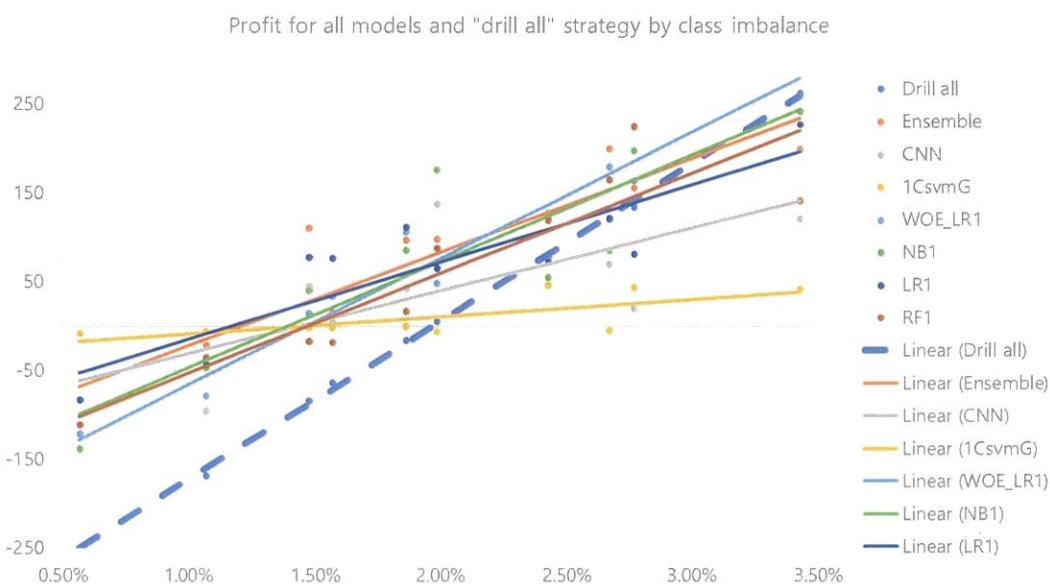


Figure 10 - Profit for selected models vs. "drill all" strategy by class imbalance

When considering the profit rate, rather than overall profit, the best model (using the highest median profit rate in all 10 cross validation folds as a measure) is the logistic regression model using weight of evidence binning and variable clustering.

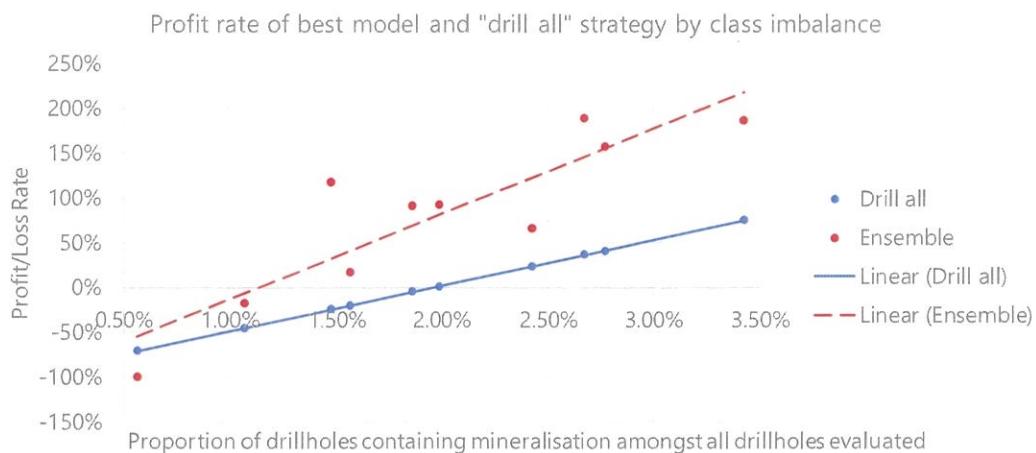


Figure 11 - Profit rate of best model and "drill all" strategy by class imbalance

This evaluation measure encourages models to be more cautious in predicting the presence of mineralisation.

Although the ensemble model failed to identify either of the two drillholes with mineralisation in the most imbalanced fold (Fold 6), the linear trend suggest that this is likely due to chance.

Overall, the model is expected to enhance a miner's profit rate in all scenarios, albeit not the overall profit. The linear trend in expected profit rate of all models is considered in Figure 12 below.

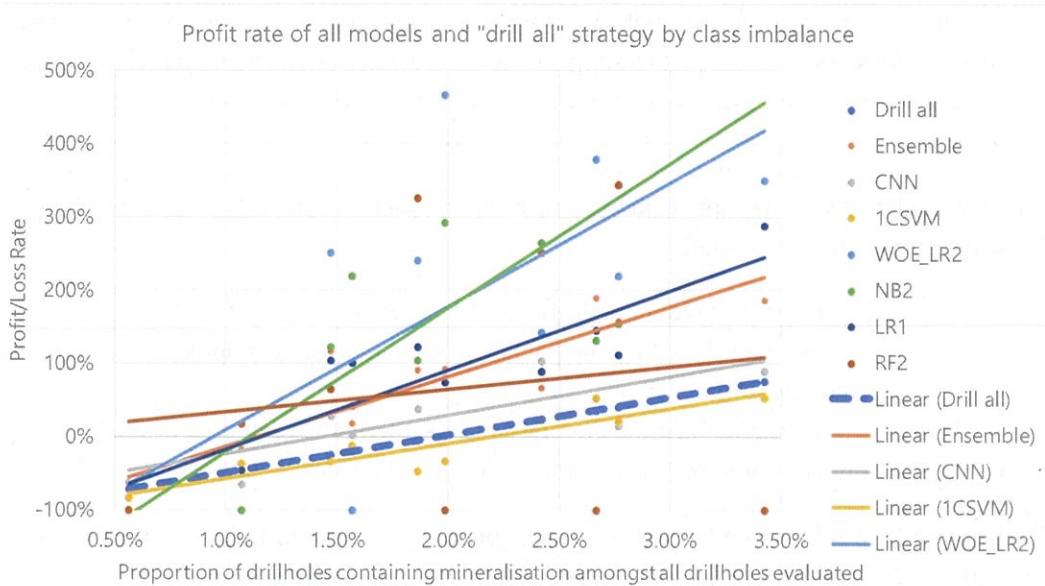


Figure 12 - Profit rate of selected models by class imbalance

The linear trend for the profit rates of the WOE_LR2 (logistic regression with clustered variables) and NB2 (naïve Bayes with factor analysis on variables) models show a distinct trend of higher profitability than other models when the proportion of successful drillholes exceeds approximately 1.25% of the set evaluated.

Variance from the linear trend is undesirably high for most models. In this regard, LR1 (logistic regression with variables consolidated using factor analysis) is the lowest risk option giving good results.

7 Conclusion

This primary research question for this report considers the potential for data driven methods to detect the presence of mineralisation in the Eastern Gawler Craton region using only gravity and magnetic maps.

Our dataset included all drillholes in the study area known to intersect the basement. Of these, approximately 2% were classified as 'successful', containing sufficient mineralisation to be commercial viable mines. This class imbalance poses a challenge to most classification algorithms and encourages them to overpredict the majority class. For our business case, it is more valuable to correctly classify the minority class, drillholes with mineralisation. This challenge was addressed by the use of oversampling techniques, particularly SMOTE (Chawla et al. 2002) and ROSE (Lunardon, Torelli & Menardi 2014). Also, the use of a profit matrix

assisted in addressing the class imbalance by rewarding correct classification of the minority with a higher payoff.

A wide variety of classification models were attempted including, logistic regression, naïve Bayes, support vector machines (both traditional and one-class), random forest and neural networks. All methods except the support vector machine could detect mineralisation to achieve higher profit and profit rate than the naïve strategies of drilling all or no drillholes in the set evaluated.

Another challenge was the high dimensionality of the dataset. To address this problem, the following approaches were trialled:

- Use of dimensionality reduction techniques including:
 - cluster analysis with principal component analysis to summarise each cluster
 - factor analysis
- Use of a convolutional neural network which is tolerant to high dimensionality
- Use of only one value from each gravity or magnetic layer.

The best results were obtained using the dimensionality reduction techniques which made the data suitable for analysis with a wide range of classification algorithms.

Each data driven techniques ability to outperform the ‘drill all’ or ‘drill nothing’ strategies was dependent on the proportion of drillholes containing mineralisation in the dataset being evaluated. Within the study area, the proportion of drillholes containing mineralisation was approximately 2%. The linear trends from each classification models’ results suggest that improvement over a “drill all” or “drill nothing” strategy is possible when the estimated proportion of drillholes being evaluated is between approximately 1.5% and 3.0%. These proportions will vary if the profit matrix is adjusted.

Evaluating models using the profit rate as a measure of success changes this conclusion. Most classification model improves the profit rate if the estimated proportion of drillholes in the study area was higher than approximately 1% with no upper bound on the proportion where improvement is obtained.

7.1 Recommendations and areas for future research

7.1.1 Circumstances where data driven approach should be useful

One major advantage of data driven methods over manual expert analysis in mineral prospecting is its ability to scale up across a wide geographical area. With sufficient computing resources, the models developed could be easily applied to all points in the study area to create a prospectivity map. Uncertainty exists however whether this map would be useful. Assuming our adopted ratios in the profit matrix are correct then our data driven models outperform a ‘drill all’ strategy (also equivalent to a random drilling strategy) if the prior probability of success is between 1% and 3%. The success rate of drillholes currently in the study area is approximately 2%. It is unclear how much of this prior success is due to expert

knowledge. Would drilling randomly throughout the study area result in a similar proportion of successful drillholes? If expert knowledge has, for the sake of argument, doubled the proportion of successful drillholes found, then drilling randomly would result in approximately 1% of drillholes finding mineralisation. In this case, and assuming our profit matrix is reasonable, then using our data driven approach is unlikely to result in a profitable exploration endeavour, although it should mitigate loss.

Nevertheless, if human experts create a set of candidate locations for drilling and within this set approximately 2% contain drillholes with mineralisation, then our classification techniques should be useful in selecting a subset likely to yield higher profit.

Also, creating a prospectivity map with using our classification algorithms may help focus the efforts of human experts and make their search more efficient and productive.

7.1.2 Areas of future research

The most useful methods for adapting traditional classification techniques to this problem involved addressing the class imbalance and dimensionality reduction.

Considering the class imbalance issue, using profit and profit rate as a measure assisted in comparing the relative performance of classification models. One way the performance of classifiers may be improved is by incorporating this profit/cost measure into classification algorithms training.

Regarding the dimensionality reduction problem, our principal component of each cluster approach arbitrarily adopted 10 clusters. It would be useful to adopt a more systematic approach in future research to consider the effect or more or fewer clusters on classification accuracy.

Our approach was limited to using gravity and magnetic maps as inputs for classification models. Although much of the study area is sparsely explored, in regions where geochemical data exists from previous drillhole exploration, this further information would likely hugely improve model accuracy in more densely explored areas.

References

→ Referencing — use one of the standard methods of referencing

Albora, AM, Ucan, ON, Ozmen, A & Ozkan, T 2001, 'Separation of Bouguer anomaly map using cellular neural network', *Journal of Applied Geophysics*, vol. 46, no. 2, 2001/02/01, pp. 129-142.

Allison, PD 2012, *Logistic Regression Using SAS: Theory and Application*, SAS Institute Inc., North Carolina.

Amer, M, Goldstein, Markus & Abdennadher, Slim 2013, 'Enhancing One-class Support Vector Machines for Unsupervised Anomaly Detection', *KDD 2013*. viewed 16 Oct 2017, <http://www.outlier-analytics.org/odd13kdd/papers/slides_amer.goldstein.abdennadher.pdf>.

Breiman, L & Cutler, A 2001, *Random Forest*, University of California, Berkeley, viewed April 2017, <https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm>.

Carranza, EJM & Laborte, AG 2015, 'Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines)', *Computers & Geosciences*, vol. 74, 2015/01/01/, pp. 60-70.

Chavent, M, Liquet, B, Kuenst-Simonet, V & Saracco, J 2012, 'ClustOfVar: An R Package for Clustering of Variables', *Journal of Statistical Software*, vol. 50, no. 13, pp. 1-16.

Chawla, NV, Bowyer, KW, Hall, LO & Kegelmeyer, WP 2002, 'SMOTE: synthetic minority over-sampling technique', *Journal of artificial intelligence research*, vol. 16, pp. 321-357.

Chi, M, Koedinger, KR, Gordon, GJ, Jordon, P & VanLahn, K 2011, 'Instructional Factors Analysis: A Cognitive Model For Multiple Instructional Interventions', in M Pechenizkiy et al (eds), *Proceedings of the 4th International Conference on Educational Data Mining*, pp. 61-70.

Cox, DR & Snell, EJ 1989, *Analysis of binary data*, CRC Press,

Department of State Development, tGoSA 2017, SARIG, *South Australian Resources Information Gateway*, viewed 10 September 2017, <<https://map.sarig.sa.gov.au/>>.

Dutta, S, Bandopadhyay, S, Ganguli, R & Misra, D 2010, 'Machine Learning Algorithms and Their Application to Ore Reserve Estimation of Sparse and Imprecise Data', *Journal of Intelligent Learning Systems & Applications*.

Granek, J 2017, *Application of machine learning algorithms to mineral prospectivity mapping*,

Harris, JR, Grunsky, E, Behnia, P & Corrigan, D 2015, 'Data- and knowledge-driven mineral prospectivity maps for Canada's North', *Ore Geology Reviews*, vol. 71, 2015/12/01/, pp. 788-803.

Harvey, AS & Fotopoulos, G 2016, 'Geological mapping using machine learning algorithms', *The International Archives of the Photogrammetry*.

Hayton, PM, Schölkopf, B, Tarassenko, L & Anuzis, P 2001, 'Support vector novelty detection applied to jet engine vibration spectra', *Advances in neural information processing systems*, pp. 946-952. viewed 16 Oct 2017, <<http://papers.nips.cc/paper/1887-support-vector-novelty-detection-applied-to-jet-engine-vibration-spectra.pdf>>.

Kowalczyk, A & Raskutti, B 2002, 'One class SVM for yeast regulation prediction', *ACM Sigkdd Explorations Newsletter*, vol. 4, no. 2, pp. 99-100.

Lunardon, N, Torelli, N & Menardi, G 2014, 'ROSE: A package for binary imbalanced learning', *R Journal*, vol. 6, no. 1, pp. 79-89.

Manevitz, LM & Yousef, M 2001, 'One-class SVMs for document classification', *Journal of Machine Learning Research*, vol. 2, no. Dec, pp. 139-154.

Porwal, A & Carranza, EJM 2015, 'Introduction to the Special Issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration', *Ore Geology Reviews*, vol. 71, 2015/12/01, pp. 477-483.

Rodriguez-Galiano, V, Sanchez-Castillo, M, Chica-Olmo, M & Chica-Rivas, M 2015, 'Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines', *Ore Geology Reviews*, vol. 71, 2015/12/01/, pp. 804-818.

Schaeben, H & Semmler, G 2016, 'The quest for conditional independence in prospectivity modeling: weights-of-evidence, boost weights-of-evidence, and logistic regression', *Frontiers of Earth Science*, vol. 10, no. 3, 2016/09/01, pp. 389-408.

Schölkopf, B, Williamson, RC, Smola, AJ, Shawe-Taylor, J & Platt, JC 2000, 'Support vector method for novelty detection', *Advances in neural information processing systems*, pp. 582-588. viewed 16 Oct 2017, <<http://papers.nips.cc/paper/1723-support-vector-method-for-novelty-detection.pdf>>.

Vallance, S 2016, *Deep Gold: Using Convolution Networks to Find Minerals*, Hackernoon, viewed 7th Aug 2017, <<https://hackernoon.com/deep-gold-using-convolution-networks-to-find-minerals-aafdb37355df>>.

Zuo, R & Carranza, EJM 2011, 'Support vector machine: A tool for mapping mineral prospectivity', *Computers & Geosciences*, vol. 37, no. 12, 2011/12/01/, pp. 1967-1975.