Machine learning as a tool for geologists

Antoine Caté[1,2], Lorenzo Perozzi[1,2], Erwan Gloaguen[1], and Martin Blouin[1,2]

Abstract

Machine learning is becoming an appealing tool in various fields of earth sciences, especially in resources estimation. Six machine learning algorithms have been used to predict the presence of gold mineralization in drill core from geophysical logs acquired at the Lalor deposit, Manitoba, Canada. Results show that the integration of a set of rock physical properties — measured at closely spaced intervals along the drill core — with ensemble machine learning algorithms allows the detection of gold-bearing intervals with an adequate rate of success. Since the resulting prediction is continuous along the drill core, the use of this type of tool in the future will help geologists in selecting sound intervals for assay sampling and in modeling more continuous ore bodies during the entire life of a mine.

Introduction

Since most outcropping deposits already have been discovered and mature mining camps have started to dry out, the discovery of new mineral deposits in the last 15 years has become increasingly expensive and risky (Schodde, 2011). New discoveries tend to be deeper and are located in more complex geologic settings. Hence, new geophysical, geochemical, and geologic data-collection tools are developed to compensate for the increasing difficulty of deposit discovery. In the next few years, logging tools with downhole sensors adapted to the mining industry (e.g., DET-CRC program in Australia) will permit the introduction of rock physical properties as standard data available during drilling campaigns. The historically knowledge-driven exploration industry will then be shifting toward a more data-driven approach. Indeed, new mining tools will generate gigantic amounts of data acquired at an almost continuous rate. This data carries a strong potential for helping geologists and mining engineers by providing tools to better log the drill core, predict the lithofacies from geophysical logs, predict mineralization, and optimize drilling and exploitation. However, existing data-management and interpretation tools cannot cope with the quantity and variety of data collected. New integration methods are needed to optimize the outcome of this expensive data and allow their effective use in the exploration process. Recently developed data mining and machine learning techniques allow one to identify patterns in large multivariate data sets and to make predictions based on them. These methods have great potential for data integration and can help in decision making for deposit modeling (e.g., Hill et al., 2014). However, little research has been focused on applying machine learning analysis for optimal and real-time mine management.

In this paper, we describe a workflow using rock physical properties and machine learning to predict the presence of gold in the drill core, which would help geologists optimize sampling for assaying. The objective is to evaluate the performance of machine learning algorithms to predict the presence of invisible gold in the drill core, using rock physical properties.

Geologic context and data

Area of interest. The Zn-Cu-Au Lalor deposit is a volcanogenic massive sulphide deposit exploited by HudBay Minerals (Hudbay) and located in central-north Manitoba, Canada, in the Snow Lake mining camp. The deposit is composed of at least 12 stacked ore lenses divided into base metal (Zn-rich), gold, and copper-gold ore lenses (Caté et al., 2015). Base metal and copper-gold lenses are composed mainly of massive to semimassive sulphides and are easily distinguishable in drill core. Gold lenses are composed of disseminated sulphides, which can be difficult to distinguish from the hydrothermally altered sulphide-bearing wall rocks. The nugget effect and high variance in gold mineralization make the identification of mineralized bodies and mapping of their continuity in space challenging. As is the case with many other gold deposits, gold-bearing mineral assemblages at Lalor can be difficult to discriminate in drill core, which can introduce errors in the process of selecting core intervals for assaying metal content. These errors can lead to an underestimation in the volume of ore zones and lead to the overlooking of economic zones during exploitation of the deposit.

Data. Combined drill-hole rock physical properties and metal assay data are available in a total of 14 drill holes intersecting lenses of the deposit. A typical data set along a section of a drill hole is presented in Figure 1.

Assay data was collected by Hudbay and analyzed for metals by Hudbay and ACME Laboratories. Assays were collected only in 0.2 to 1 m long intervals considered as potentially metal-bearing during the core-logging process. Eight elements were analyzed (Ag, As, Au, Cu, Fe, Ni, Pb, and Zn), and a significant part of the measurements is below the detection limit. Only gold values are used in this study. Analysis and QA/QC methods are provided in Carter et al. (2012).

Physical rock properties were logged by DGI Geoscience for Hudbay. A total of 15 rock properties were collected at a 10 to 20 cm spacing (Figure 1). A significant part of the measurements for each rock property was not taken into account for various reasons and indicated as "not a value" in the data set. As an example, approximately 93% of the conductivity measurements were set at 0.5, a point at which no strong variation of the conductivity was encountered.

Data preprocessing

Joining assay and physical properties data sets. Gold assay measurements have been composited into 1 m long intervals using weighted averages in order to have a homogeneous data set and to avoid biased statistics during the modeling. Because physical properties have been logged with a 10 to 20 cm interval, between five and 10 measurements were taken within each composited gold assay interval. Parts of the physical properties were log-transformed to obtain unskewed distribution. All 0.5 values for conductivity were replaced by log(1/log(Resistivity_8inch)) as the

[1]INRS, Centre Eau Terre Environnement.
[2]geoLEARN (geolearn.ca).

http://dx.doi.org/10.1190/tle36030215.1.

two properties have a Pearson's r correlation coefficient of 0.71. This replacement enables retaining measured values of conductivity instead of completely dropping the property.

For each interval, we computed derived features (variables) from summary statistics (minimum and maximum value, mean, median, standard deviation, and variance) of each physical property to enhance the high-resolution information contained in the geophysical logs.

The resulting data set contains several "not a value" (empty cells). A common strategy consists of removing the corresponding features (columns) or samples (rows) from the data set entirely (Raschka, 2015). However, by applying this correction, part of the information that could be valuable for the final prediction may be lost.

A solution, when using ensemble machine learning methods, is to replace the missing values with an out-of-range value, usually -99999, that tell the algorithms to ignore the missing values. This strategy has been used for all statistically derived features.

Cut-off value for gold. The objective of this study is to evaluate the presence of gold in the rock (discriminate between low- and high-gold content in the rock) and not to precisely evaluate gold grade, which is done precisely by assaying. A cut-off gold grade is chosen to differentiate between the background gold values within the deposit and high gold values related to gold lodes. The 1 g/t cut-off value has been retained arbitrarily, as it is higher than the background values, lower than economical values, and still corresponds to approximately 10% of the composited assay values.

Classification strategy

Classification algorithms. Machine learning is an application of statistical learning that identifies patterns in data and then makes predictions from those patterns. Among the three types of machine learning methods (supervised, unsupervised, and reinforcement learning), supervised learning is the best suited for this work as its main goal is to learn a model from labeled training data that allows us to make a prediction (Raschka, 2015). Here, the term "supervised" refers to a set of samples where both the desired output signals (label) and the predictive variables (logs and derived statistics) are already known. In this case, the label is a binary classification of samples having a gold value higher (positive class) or lower (negative class) than 1 g/t. A total of five machine learning algorithms are tested here: (1) the k-nearest neighbors (k-NN) method, which uses labeled neighboring points in the Euclidian space formed by the input features to predict classes; (2) the naïve Bayesian method, which uses Bayes theorem to evaluate the probability of an event (class) to occur given the value of the input data; (3) support vector machine, which is a discriminative classifier formally defined by a separating hyperplane; (4) classification trees, which are decision trees built by using thresholds on input features at each split; and (5) ensemble algorithms. The latter is particularly well suited for this kind of problem as they combine the predictions of several base estimators built with a given learning algorithm in order to improve robustness over a single estimator. These algorithms are particularly resistant to noisy data and to outliers (Breiman, 2001). Here, the random forest and the gradient tree boosting algorithms are tested. Both algorithms use decisions trees as base estimators.

All algorithms can be tuned with a series of various algorithm-specific parameters that significantly contribute to the robustness, variance, and bias of the classification. The choice of the best parameters is done through the training/validating process. Here, the algorithms and the tuning tools proposed by ScikitLearn (Pedregosa et al., 2011) with Python have been used for the implementation.

Choosing the training and testing data sets. The data set was split into a training set composed of the data from 11 drill holes and a testing set composed of the data from the remaining three drill holes (Figure 2). Both the training and testing sets were chosen so as to be representative of the geology of the deposit and of the different types of gold mineralization.

The training set is used to tune/optimize algorithm parameters and evaluate the prediction success of the algorithms. The data set is highly unbalanced (negative class with Au < 1 g/t is much greater than the positive class with Au > 1 g/t), and to cope for it, the

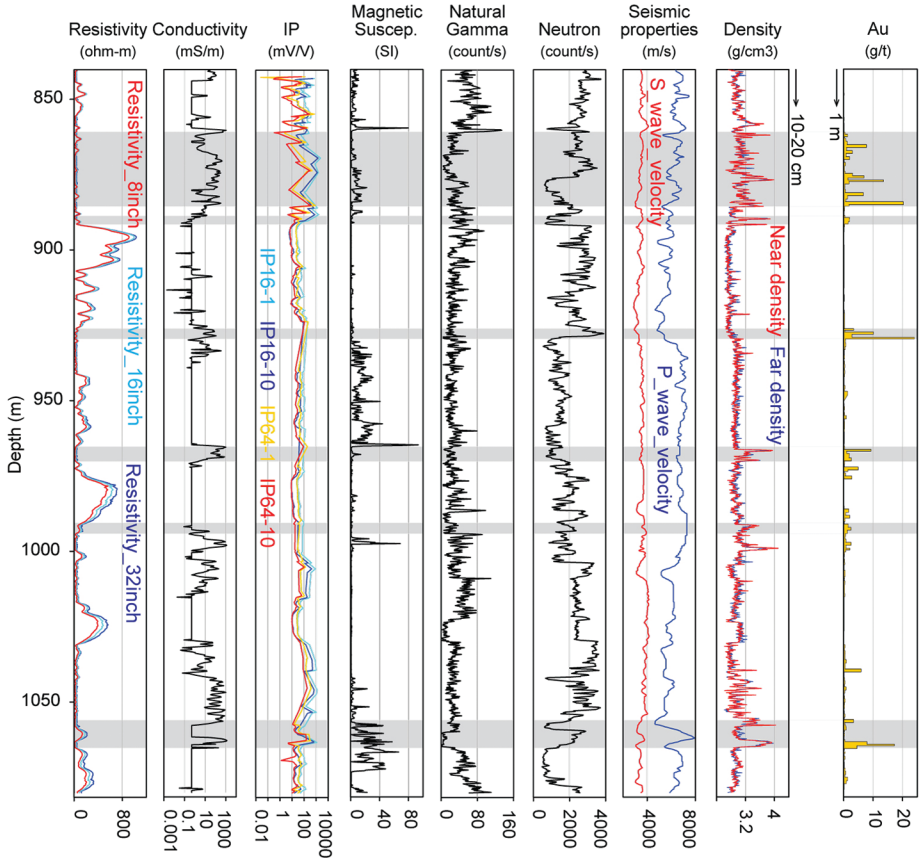


Figure 1. Typical Lalor data set. Physical properties logs are sampled at 10–20 cm, and assay composites are 1 m long. Gray background indicates the ore lenses defined by Hudbay.

negative class has been randomly undersampled to match the positive class (Chawla, 2005). The resulting training data set is composed of 382 entries equally divided into positive and negative classes.

The testing set is used as a blind data set to evaluate the final prediction success of the model.

Training and validation. The training step is used to choose the best-suited algorithm and combination of parameters for the classification problem. The goal is to optimize classification-scoring metrics (e.g., accuracy, precision, recall, and *f*1) by selecting a model with both low variance and bias. *K*-fold cross validation is a useful technique to obtain reliable estimates of the model generalization error (i.e., how well the model performs on unseen data). This technique consists of randomly splitting the data set into *k* folds, where *k*-1 folds are used to build the model and one fold is used for validation (estimating the prediction success). The *k* folds are randomly built from the training set, and the training/validating process is repeated for each fold (Figure 2). Scoring metrics are averaged for each validation fold and used to assess the predictive model. The objective of the exercise is to predict the presence of gold mineralization greater than 1g/t (positive class) in the drill core while minimizing false positives. The scoring metric selected for model selection and tuning is the *f*1 score of the positive class. The *f*1 score is a measure of the accuracy combining the precision and the recall. It gives a general idea of the success rate of the prediction (Raschka, 2015). Each algorithm was tested with a range of parameter combinations by cross validation. This step is critical as it can significantly improve the success rate of the final prediction.

Results of the cross-validation process for each algorithm using the best sets of parameters are presented in Figure 3. In general, ensemble algorithms score significantly better than the other algorithms. The gradient boosting algorithm yields a lower variance in the scores and is thus chosen for the final prediction.

A total of four parameters of the gradient boosting algorithm from ScikitLearn (Pedregosa et al., 2011) were optimized for *f*1 with cross validation (n_estimators, learning_rate, max_depth, min_samples_leaf). The best set of parameters obtained after tuning are: n_estimators=200, learning_rate=0.1, max_depth=10, min_samples_leaf =10. The validation scores for the positive class on the training set obtained with a 10-fold cross validation are presented in Table 1. All scoring metrics are above 0.7, but their standard deviation is relatively high (>0.1), which is probably related to the small training set size.

Classification of the test data

Scoring metrics. After selecting the best algorithm and optimal set of parameters, the predictive model was built using the entire training set (Figure 2). The predictive model was run on the test data set to estimate the generalization error. Table 2 shows the classification report resuming the scores for each class. Scores for the positive class are similar to that obtained for validation, which suggests that the generalization error is not important. The moderately high recall for the positive class and high precision for the negative class indicate that the number of true positives misclassified as negative is low. The low precision for the positive class indicates a high rate of false positives. However, the estimated precision of the geologist for the positive class (number of assays > 1 g/t divided by total number of assays) is of 0.14, which is half of that of the model. This difference can be

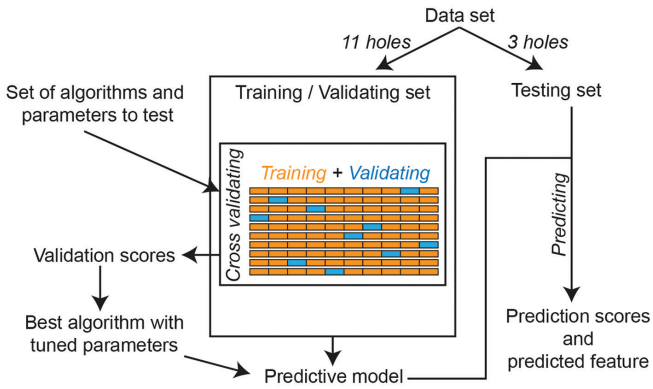


Figure 2. Schematic diagram showing the process of machine learning model building.

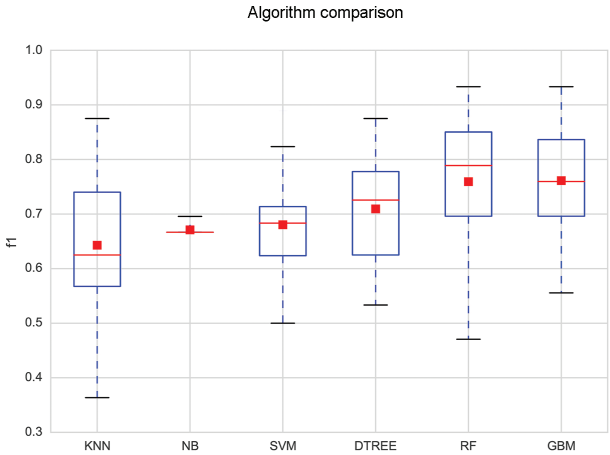


Figure 3. Boxplots of the *f*1 score obtained from cross validation performed over the best set of parameters for each algorithm. The red dots represent the mean, the red lines the median; samples comprised within the first quartile (Q1) to third quartile (Q3) range are included in the blue boxes; crosses represent outliers (further than 1.5*[Q3–Q1] from the box). Extreme values that are not outliers are included within the whiskers.

Table 1. Classification metrics obtained by cross validation over the training set using the tuned gradient tree boosting algorithm.

	Precision	Recall	*f*1-score
Mean	0.75	0.75	0.74
Std	0.11	0.15	0.11

Table 2. Classification report resuming scores obtained with the prediction over the testing set using the tuned gradient tree boosting algorithm.

	Precision	Recall	*f*1-score	Support
Negative (Au < 1 g/t)	0.94	0.68	0.79	372
Positive (Au > 1 g/t)	0.27	0.72	0.39	61
Avg/total	0.84	0.69	0.73	433

explained by geologists oversampling for assaying so as to minimize the risk to overlook mineralization.

The receiver operating characteristic (ROC) curve gives insight on the precision/recall tradeoff for the positive class by giving the true positive rate (i.e., recall) at different false positive rate thresholds (Figure 4). Increasing the recall above 0.8 would be at the expense of dramatically increasing false positive rates, which would lead to a precision score similar to that of a geologist (as estimated above).

Prediction along drill holes. Prediction results obtained on one of the test drill holes are presented in Figure 5. Intervals located above ~750 m depth generally have a low probability of bearing gold, while intervals below are predicted as more likely, in general, to be gold bearing, which corresponds to the observed assay results. All zones with high gold values according to assays have been classified as gold bearing by the prediction model in these drill holes. The probability of an interval being gold bearing is distributed as high-value intervals centered on the actual high-grade gold zones and with a more smoothed distribution than measured gold grades (presented on a log scale in Figure 5). However, a few intervals detected by the model as potentially gold bearing have not been assayed or have been assayed and include only gold values below 1 g/t.

Feature importance. The gradient tree boosting algorithm allows the evaluation of the importance of the features used for the classification. The result is expressed as the individual contribution of each feature for building the predictive model. The feature-importance histogram is presented in Figure 6. An inflexion in the feature-importance curve is visible at the 15th feature, and the combined 15 first features account for 60% of the classification power.

As expected, the most informative features are derived from neutron activation, natural gamma, and resistivity logs. Both neutron activation and natural gamma give insights on the elemental composition of the host rocks. Their classification powers are probably derived from the variations in rock composition, in part

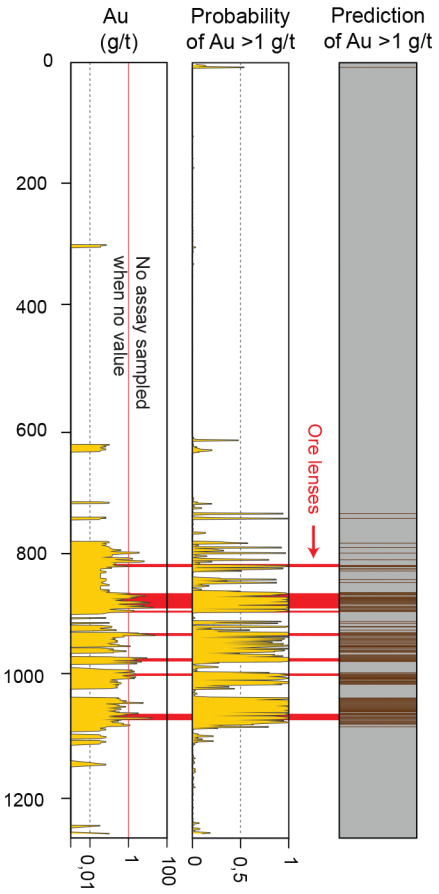


Figure 5. Results of the prediction on a test drill hole. The values of gold concentration on assayed intervals (left), the predicted probability of Au>1 g/t (middle), and the result of the classification (right) are presented. The location of ore lenses, defined by Hudbay during the exploration stage (not up to date data, refer to Hudbay's website for information on reserves and resources), is given for reference.

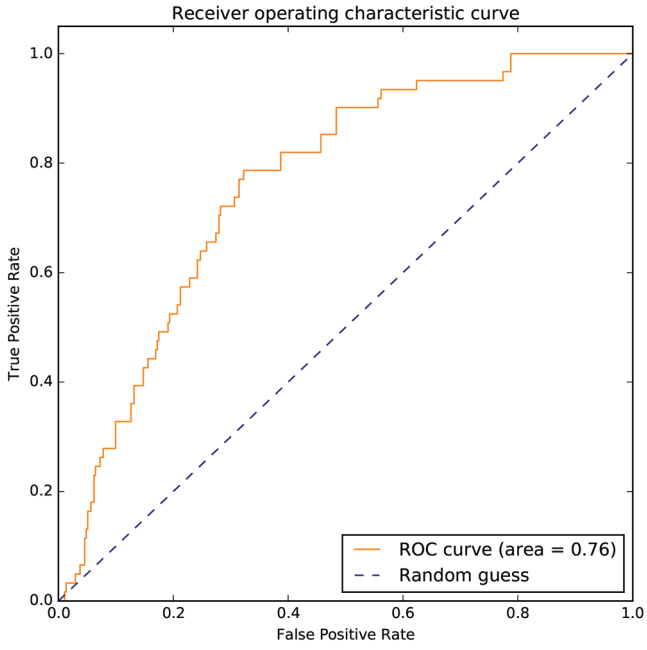
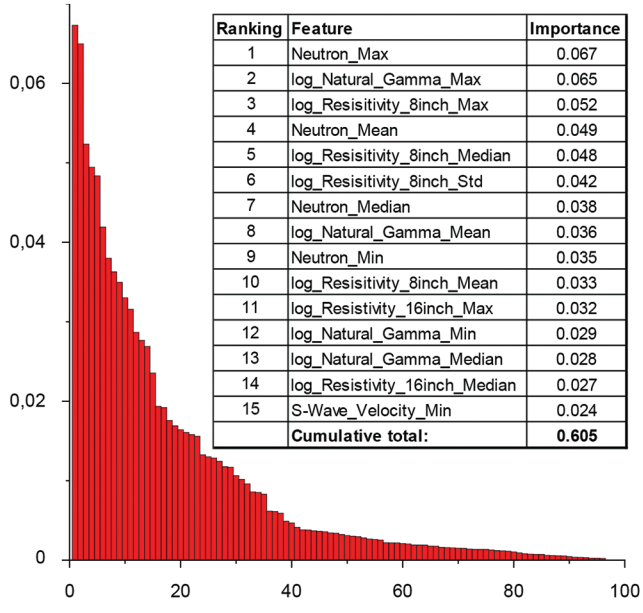


Figure 4. Receiver operating characteristic (ROC) curve of the prediction of the presence of gold in test drill holes.



Ranking	Feature	Importance
1	Neutron_Max	0.067
2	log_Natural_Gamma_Max	0.065
3	log_Resisitivity_8inch_Max	0.052
4	Neutron_Mean	0.049
5	log_Resisitivity_8inch_Median	0.048
6	log_Resisitivity_8inch_Std	0.042
7	Neutron_Median	0.038
8	log_Natural_Gamma_Mean	0.036
9	Neutron_Min	0.035
10	log_Resisitivity_8inch_Mean	0.033
11	log_Resistivity_16inch_Max	0.032
12	log_Natural_Gamma_Min	0.029
13	log_Natural_Gamma_Median	0.028
14	log_Resistivity_16inch_Median	0.027
15	S-Wave_Velocity_Min	0.024
	Cumulative total:	**0.605**

Figure 6. Relative importance of input features (descending order) for classification with the gradient tree boosting algorithm. The 15 most important features are presented in the table.

due to the presence of alteration associated with gold mineralization (Caté et al., 2015). Resistivity is a proxy for the presence of conductive minerals such as sulfides (e.g., pyrrhotite and chalcopyrite), which are associated with the presence of gold.

Most statistical calculations are represented among the best features, but the standard deviation is present only once, which indicates that this statistical information has less classification power. The three features contributing the most to the prediction are the maximum values for neutron activation, natural gamma, and resistivity (8 inches). This indicates that maximum values have a high classification power.

Discussion

Ensemble algorithms, and the gradient tree boosting algorithm in particular, yield better results than the other algorithms. This is in agreement with the general performance of such algorithms. The difference in performance between the random forest and the gradient tree boosting algorithms is minimal (Figure 3) and might evolve as the training set size changes. This illustrates that the choice of algorithm is not absolute.

Gold distribution is not homogeneous in drill core, and it is subject to a high local variability (nugget effect), which makes ore-bodies modeling difficult. The more evenly distributed results of the prediction can be used as a more spatially continuous proxy to improve 3D modeling as suggested by Hill et al. (2014).

Features derived from neutron activation and natural gamma measurements appeared in the main contributors to the classification. These measurements are very approximate proxies for rock composition. Quantitative to semiquantitative tools measuring element concentrations in rock (e.g., portable XRF) have the potential to measure with a greater quality the variations in major elements forming the rock. Because the presence of gold in rocks is usually associated with specific rock formations (e.g., banded iron formation or intrusive rocks), alteration, and the presence of veins, information on rock composition is critical to the prediction of gold mineralization.

The use of statistics gives information on the texture of the data, but further feature engineering might improve the classification power of the various physical rock properties.

Conclusion

Contrary to base metals (e.g., Zn and Cu), the presence of gold in a rock is difficult to assess, even for the trained eye of an experienced geologist. The combination of a set of rock physical properties measured at closely spaced intervals along the drill core with machine learning allows the detection of gold-bearing intervals with an adequate success rate close to that of a geologist. The predicted recall, precision, and accuracy show room for further improvement, which could be obtained by the collection of more training data or different data types and/or improved feature engineering.

Because the resulting prediction is continuous along the drill core, this kind of tool may help the geologist when making decisions on which intervals to select for assay sampling. Missed gold-bearing intervals could be reduced significantly, which in turn could potentially increase the reserve.

Along with predicting the presence of metals in rocks, physical properties combined with machine learning have the potential to classify lithologies, characterize hydrothermal alteration, and estimate exploration vectors and geotechnical information in the drill core (e.g., Ross et al., 2013). The combination of these predictions will significantly improve the work of logging geologists, the quality of geologic interpretations, and the decisions made during a drilling campaign and during the exploitation of a mine.

This method has been developed on a restricted set of data (14 drill holes). The success rate of predictions will increase as increasing amounts of data are collected. This method should be applied from the very beginning of the exploration stage (i.e., starting from the discovery hole) so that the initial model can be trained and continuously updated with new drill holes. TLE

Acknowledgments

This study has been completed with funds from the MITACS program, Mira Geoscience Ltd., and the Canada Chair on stochastic geologic modeling using geophysical and geologic data assimilation. Data was provided by Hudbay and Mira Geoscience Ltd.

Corresponding author: lorenzo.perozzi@gmail.com

References

Breiman, L., 2001, Random forests: Machine Learning, **45**, no. 1, 5–32, http://dx.doi.org/10.1023/A:1010933404324.

Carter, R. P., T. Schwartz, S. West, and K. Hoover, 2012, Pre-feasibility study technical report, on the Lalor deposit, Snow Lake, Manitoba, Canada: HudBay Minerals, http://s1.q4cdn.com/305438552/files/doc_downloads/our_business/Lalor%20PreFeasandTechnicalReport_Mar30_v001_o9by3l.pdf, accessed 30 January 2017.

Caté, A., P. Mercier-Langevin, P. S. Ross, S. Duff, M. Hannington, B. Dubé, and S. Gagné, 2015, Geology and Au enrichment processes at the Paleoproterozoic Lalor auriferous volcanogenic massive sulphide deposit, Snow Lake, Manitoba, *in* J. Peter and P. Mercier-Langevin, eds., Targeted Geoscience Initiative 4: Contributions to the understanding of volcanogenic massive sulphide deposit genesis and exploration methods development: Geological Survey of Canada Open File 7853, 131–146.

Chawla, N. V., 2005, Data mining for imbalanced data sets: An overview, *in* O. Maimon, and L. Rokach, eds., Data mining and knowledge discovery handbook: Springer, 853–867, https://dx.doi.org/10.1007/0-387-25465-X_40.

Hill, E. J., N. H. Oliver, L. Fisher, J. S. Cleverley, and M. J. Nugus, 2014, Using geochemical proxies to model nuggety gold deposits: An example from Sunrise Dam, Western Australia: Journal of Geochemical Exploration, **145**, 12–24, http://dx.doi.org/10.1016/j.gexplo.2014.05.008.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, 2011, Scikit-learn: Machine learning in python: JMLR, **12**, 2825–2830, http://www.jmlr.org/papers/v12/pedregosa11a.html.

Raschka, S., 2015, Python machine learning: Packt Publishing Ltd.

Ross, P.-S., A. Bourke, and B. Fresia, 2013, A multi-sensor logger for rock cores: Methodology and preliminary results from the Matagami mining camp, Canada: Ore Geology Reviews, **53**, 93–111, http://dx.doi.org/10.1016/j.oregeorev.2013.01.002.

Schodde, R., 2011, Recent trends in gold discovery: Presented at NewGenGold Conference.