DEPODART · WHITEPAPER · 2026

AI in mineral exploration

A framework for honest, auditable prospectivity mapping

Author DepoDart Geoscience (names to be confirmed before publication)

Date 2026

Version 0.1 (draft, internal review) depodart.com

Abstract

Mineral exploration is, in the end, an exercise in deciding where to drill. A drill program commits a large fraction of an annual budget to a small number of holes, on the basis of evidence that is always incomplete and frequently contradictory. The first wave of machine learning tools marketed to mining promised to resolve this asymmetry, and largely failed to do so. The failures were not failures of statistics. They were failures of transparency, calibration, validation, and respect for the geological context the tools were meant to support.

This paper sets out a framework for prospectivity mapping that takes those failures seriously. It describes a methodology built on positive-unlabelled learning, gradient-boosted ensembles, spatial cross-validation, Platt-scaled probabilities, and bootstrap uncertainty estimation. It argues that an uncalibrated probability map is worse than no map at all, that uncertainty must be a first-class output rather than a footnote, and that any honest system must disclose where it cannot help. The paper closes with a description of how DepoDart applies this framework inside a hands-on engagement that typically runs two to four weeks, the artifacts every client walks away with, and an honest account of where qualified geological judgment must override the model.

The intended audience is exploration managers, chief geoscientists, technical evaluators at mining majors and juniors, and academic geoscientists who follow the industry. The paper is technical where it needs to be, and plain where it can be.

Table of contents

1. The targeting problem

Exploration runs on incomplete information. That sentence is the starting point for every other observation in this paper.

A drill program commits a large budget to a small number of holes. The cost of targeting the wrong ground is measured in seasons, not days. A junior with a single project area may spend the better part of a year, and most of its treasury, on a campaign whose entire epistemic foundation is a handful of anomalies on a handful of maps. A major with a portfolio across several districts faces the same problem multiplied by jurisdiction, commodity, and partner. The decision is binary at the level of the drill bit, and it is made under uncertainty that the geophysics, the geochemistry, and the structural interpretation never fully resolve.

The information asymmetry is not a shortage of data. It is the opposite. Decades of geophysics, geochemistry, and drill logs already sit in client archives, in formats and coordinate systems that rarely line up. Public geological surveys publish at a scale and quality that would have been unimaginable a generation ago. The shortage is in the human bandwidth required to bring those layers into coherent relationship at the scale of a district.

A single anomaly is easy to read. A trained geoscientist can look at a magnetic survey and identify a feature of interest in minutes. The relationship between a dozen data layers, across a whole district, across coincident and conflicting signals, is not something the same geoscientist can hold in working memory. The combinatorial space is too large. The history of mineral exploration is, in part, a history of accepting that the integration step is necessarily partial.

The financial weight of this partiality is heavy. Industry estimates of the cost per discovery have risen consistently for two decades, even as the volume of exploration data per project has grown by orders of magnitude. The metals behind smart vehicles, robotics, and energy storage must be found faster than the world has ever found them. The current rate of discovery is not keeping up with the projected demand for copper, lithium, nickel, and the rare earth elements. The supply problem is not a problem of geology in the deep sense. It is a problem of where to look.

The promise of machine learning, in this context, is straightforward. A model can hold more data layers in working relationship than a person can. It can evaluate millions of pixel-level relationships simultaneously. It can score every square kilometre in a district against a learned signature of mineralisation, and rank the result. Whether that promise can be made good is the subject of the rest of this paper.

2. Why first-generation AI failed in mining

The first generation of machine learning tools marketed to the mining sector arrived with the same vocabulary as every other vertical receiving AI attention in the late 2010s. Deep learning, neural networks, end-to-end models, hands-off automation. A non-trivial number of exploration teams ran pilots with those tools. The results were mixed in some places and disappointing in most. The failures clustered around four issues.

The opacity problem.

Most first-generation tools were black boxes. A geoscientist would receive a heat map, a ranked list of targets, and an invitation to drill the top of the ranking. When asked why a particular pixel scored high, the tool could not say. The model had no facility for surfacing which inputs drove the prediction. The geoscientist was being asked to defend a recommendation to a board, to a partner, or to a regulator, on the strength of a vendor's assurance that the model was good. This is not a posture any working exploration manager can adopt. Opacity is one of the central problems mining has with AI, and it is the part of the problem that the design of any serious successor tool must deliberately set out to solve.

The calibration problem.

A probability that does not behave like a probability is worse than no probability at all. Many early systems output scores in a 0 to 1 range that were called probabilities but were not. A score of 0.80 did not mean that the modelled location would host a deposit roughly 80% of the time across many such scored locations. It meant that the score was higher than 0.79 and lower than 0.81 in the model's internal ranking. The distinction matters. Drill program decisions, partner negotiations, and capital allocation are all sensitive to whether the number on the map is a probability or a relative ranking. When the two are confused, expectations are set badly, and pilots fail in ways that are blamed on the tool when they should be blamed on the lack of calibration.

The template problem.

A model trained on a global compendium of porphyry copper deposits and then applied to a Canadian Shield gold setting is being asked to do something it cannot do well. Deposit-type templates are a useful pedagogical device. They are a poor substrate for a probabilistic predictor at district scale. The geological signatures of mineralisation vary across terranes, across metamorphic histories, across structural settings. A generic template flattens that variation into a single learned pattern, and the regions that do not match the template are systematically underscored. A model that is the same model for every client is, by construction, the wrong model for most of them.

The provenance problem.

When a recommendation cannot be traced back to the input layers that drove it, the recommendation is not auditable. It cannot be defended in front of a technical committee. It cannot be reproduced six months later when the data changes. It cannot be falsified, which means it cannot be improved. The discipline of mineral exploration is, in the end, a documentary practice. Every interpretation lives or dies on the audit trail that supports it. A system that does not preserve provenance is incompatible with that practice, regardless of how good its statistical performance might be in the abstract.

These four failures were not failures of machine learning as a category. They were failures of how machine learning was packaged for a regulated, evidence-driven industry that does not tolerate hand-waving. Any framework that intends to succeed in mineral exploration has to address all four directly. The next section sets out how.

3. A framework for honest prospectivity mapping

This section is the technical core of the paper. It describes the methodology DepoDart applies to every pilot, with a focus on the design choices that distinguish it from the first-generation tools described above. The structure follows the practical sequence of a model build: problem framing, training data, feature engineering, model architecture, uncertainty quantification, and validation.

3.1 Problem framing

Mineral prospectivity mapping is a spatial classification problem with severe class imbalance. Known deposits in a district are rare, often one to three, sometimes none at all. Non-mineralised terrain is the overwhelming majority of the land surface. The standard machine learning treatment of class imbalance, including techniques like oversampling the minority class or undersampling the majority, assumes that the labels in both classes are reliable. In mineral exploration, that assumption does not hold.

The known deposits are reliable positive labels. The non-deposit locations are not reliable negative labels. A pixel that is unlabelled in the training data may host an undiscovered deposit. Treating every unlabelled pixel as a confirmed negative biases the model toward an artificially crisp boundary between mineralised and non-mineralised ground that does not exist in the geology. This is the structural problem at the root of much of the first generation's poor performance.

3.2 PU learning versus naive binary classification

DepoDart uses positive-unlabelled (PU) learning, an approach formalised in the machine learning literature by Elkan and Noto (2008) and extended in the years since. PU learning explicitly models the asymmetry between known positives and unlabelled locations. The unlabelled set is treated as a mixture of latent positives and true negatives, and the loss function is constructed to recover the underlying positive class distribution rather than to penalise predictions in unlabelled regions that may, in fact, be positive.

Naive binary classification, by contrast, treats every unlabelled pixel as a hard negative. Applied to mineral prospectivity, this approach produces models that are over-confident in declaring large areas non-prospective. The result is a map that looks decisive but encodes the wrong inductive bias. Pixels that should be flagged as prospective are systematically suppressed because the model has been trained to believe that any pixel without a known deposit nearby is a negative example. The cost of that error in an exploration context is the cost of not drilling ground that should have been drilled, which is functionally invisible and therefore particularly insidious.

PU learning is not a panacea. It introduces its own assumptions, including the assumption that the labelled positives are drawn from the same distribution as the latent positives in the unlabelled set. In mineral exploration, that assumption is approximately true at district scale and within a single deposit class, and approximately false across very different deposit types or geological terranes. The methodology is honest about this limitation and accounts for it in the way training data is partitioned and validated.

3.3 Training data

Training data for each pilot is composed of two layers. The first is client project data: drill logs, geophysical surveys, geochemical assays, structural interpretations, and any prior geological work the client is willing to share. The second is public geological survey data, drawn where available from the Geological Survey of Canada, the United States Geological Survey, Geoscience Australia, and the relevant state and provincial surveys. Known mineral occurrences from client drill logs or from public databases such as MINFILE in British Columbia, the Mineral Resources Data System in the United States, and the Mineral Industry Resource Information System equivalents in other jurisdictions serve as positive labels.

The decision to combine client data with public data is deliberate. Public data alone tends to produce a model that performs well on public training labels but generalises poorly to the client's specific geological setting. Client data alone tends to be too sparse to train a robust model, particularly for under-explored project areas. The combination produces a model that benefits from the breadth of regional data and the depth of local data simultaneously.

3.4 Feature engineering

Raw rasters are not fed directly into the model. The geological information content of a magnetic survey, for example, is not in the total magnetic intensity at each pixel but in the derived quantities that geologists already use to interpret the survey: reduced-to-pole anomalies, vertical and tilt derivatives, analytic signal amplitudes. Each of those derivatives encodes a different aspect of the underlying magnetic source distribution, and each contributes independently to the model's ability to learn a mineralisation signature.

The same logic applies across data modalities. Geochemical inputs are transformed into element ratios, pathfinder anomalies, and multivariate anomaly indices that capture the relationships between elements rather than their absolute concentrations. Structural inputs are transformed into proximity-to-fault rasters, lineament density grids, and measures of structural complexity. Remote sensing inputs are transformed into spectral indices appropriate to the commodity in question.

The feature engineering layer is the place where geological knowledge is most directly encoded into the modelling pipeline. It is the difference between a model that learns from data and a model that learns from data the way a geologist would.

3.5 Model architecture

DepoDart uses an ensemble of gradient-boosted decision trees as the prediction layer. The choice is deliberate, and the reasoning is worth setting out.

Gradient boosting performs well on tabular geoscientific data without the very large training sets that deep learning typically requires. Mineral prospectivity datasets are tabular at the level of the engineered features and modest in size by deep learning standards. A district may have a few million pixels of input data, which is small relative to the data volumes that justify a deep neural network and that suit a tree ensemble well.

Gradient boosting provides native feature importance rankings that geoscientists can interpret. When a pixel is scored highly, the model can be queried for the inputs that contributed most to that score, both in aggregate across the district and locally at the pixel. This is the property that addresses the opacity problem identified in section 2. The ranking is not a post hoc explanation grafted onto an opaque model. It is a structural property of the model itself.

Gradient boosting is robust to the mixed data types that characterise geoscientific datasets. A single pipeline can ingest continuous geophysical fields, ordinal lithological codes, and binary structural flags without the elaborate normalisation that other architectures require. This robustness is a practical advantage that compounds across the data fusion stage of the pipeline.

The predictions from the ensemble are calibrated with Platt scaling, a logistic regression fit to the model's raw scores against the validation labels. The output of the calibration step is a probability in the formal sense: a pixel scored at 0.80 should host a deposit approximately 80% of the time across many such pixels. The calibration is the part of the methodology that addresses the calibration problem identified in section 2. Without it, the map is a ranking. With it, the map is a probabilistic prediction that supports defensible decision-making.

3.6 Spatial cross-validation versus random holdout

Validation is the part of the methodology where most prospectivity work goes wrong. The standard machine learning practice of random holdout, in which a fraction of the training data is held out at random and used to evaluate model performance, is inappropriate for spatial data. The reason is straightforward: pixels that are geographically close are statistically dependent. A random holdout splits the data without respecting that dependence, and the result is a validation set that contains pixels next to pixels in the training set. The model is, in effect, being tested on data it has already seen.

The metric this produces is optimistic, sometimes wildly so. A model that appears to achieve excellent accuracy under random holdout may collapse to chance performance when applied to a new district. The literature on this problem is mature; Roberts et al. (2017) is a useful and widely cited reference. The corrective is spatial cross-validation, in which the holdout set is constructed as geographic blocks rather than random pixels. The model is trained on a subset of geographic blocks and tested on the remaining blocks, so that the validation set is genuinely independent of the training set in the spatial sense.

DepoDart applies spatial cross-validation to every model, with block sizes calibrated to the spatial autocorrelation length of the input data. Where the number of known occurrences is sufficient, leave-one-out validation is also performed at the level of individual deposits: the model is trained with one known deposit held out, and the held-out deposit's location is checked against the resulting prospectivity surface. The metric this produces is conservative, and it is the metric that should be reported to clients and to readers of this paper.

3.7 Uncertainty quantification

Every prospectivity surface ships with a companion uncertainty map. The uncertainty estimate is constructed through bootstrap resampling, following the resampling tradition formalised by Efron (1979). The training data is resampled with replacement, a model is fit to each bootstrap sample, and the variance of the predictions across the bootstrap ensemble is recorded at each pixel.

The variance is interpreted as an estimate of epistemic uncertainty: the uncertainty arising from the model's limited knowledge of the underlying process, which would in principle decrease with additional training data. High variance at a pixel means the model's prediction at that pixel is unstable across bootstrap samples, which means the prediction is sensitive to which subset of the training data is used. Low variance means the prediction is stable, which means the data is sufficient to support the prediction with confidence.

The role of the uncertainty map in drill planning is the subject of section 4.

3.8 Validation and disclosure

The methodology is honest about its limits. In districts with fewer than three known occurrences, validation is inherently limited. The leave-one-out check is not statistically meaningful with one or two deposits, and the spatial cross-validation may be dominated by the holdout block that happens to contain the deposit. In these settings, the model can still be informative, but the prospectivity surface should be read as exploratory rather than confirmatory, and the uncertainty map should be weighted accordingly.

Every pilot report includes a disclosure section that names the validation regime applied, the number of positive labels used, and the residual uncertainty in the validation estimate itself. This is not a defensive footnote. It is part of the deliverable.

4. The role of uncertainty in drill planning

> *A probability without an uncertainty estimate is worse than no probability at all.*

The reason is that a probability without an uncertainty estimate invites the reader to act as though the probability is reliable, while concealing the information needed to assess whether it is. A pixel scored at 0.85 is operationally identical to a pixel scored at 0.65 if the uncertainty around the 0.85 prediction is wide enough that the true probability could plausibly be 0.5. The map shows the higher number; the underlying evidence does not support acting on the difference.

The two-map workflow is the practical response to this problem. Every DepoDart pilot delivers a prospectivity surface and a companion uncertainty surface as separate, registered GeoTIFFs. The intended reading is joint: every decision derived from the prospectivity map should be informed by the corresponding region of the uncertainty map. The two surfaces support different categories of decision.

High probability, low uncertainty.

These are the pixels where the model is confident and the evidence is sufficient to support the confidence. They are the most defensible drill targets in the conventional sense. A drill program designed against these pixels is being placed where the model has the strongest claim to be right, and the bootstrap variance is low enough that the claim is not an artefact of any particular subset of the training data. Most pilot reports rank these locations as Tier 1 targets.

High probability, high uncertainty.

These are the pixels that deserve more thought. The model is predicting that mineralisation is likely, but the prediction is unstable across bootstrap samples, which means that the prediction is sensitive to the available data. The honest reading is that the pixel is interesting and the data is insufficient to fully support the interpretation. The appropriate response is not to drill, or at least not to drill first. The appropriate response is to collect additional data: an infill geochemical survey, a new geophysical line, a structural reinterpretation in the surrounding area. The uncertainty map, in this regime, is a planning tool for data acquisition rather than for drilling.

Low probability, low uncertainty.

These are the pixels the model is confident in deprioritising. They are the parts of the district that can be set aside with reasonable assurance. In a large project area, the practical value of this category is significant. It allows exploration budgets to be concentrated on the parts of the district where the evidence is most promising, without spending field season time on areas that are unlikely to repay the effort.

Low probability, high uncertainty.

These are the pixels the model is uncertain about in a way that suggests neither drilling nor deprioritisation is justified. They are most often pixels in geological settings that are poorly represented in the training data. The honest response is to flag them and to consult the geological team about whether the setting is one in which the model can be expected to perform.

The joint reading of the two maps is the operational core of the methodology. It is the part that distinguishes a defensible drill plan from a heat map ranking, and it is the part that most directly addresses the failure modes of first-generation tools. A drill program planned against the prospectivity surface alone is a drill program planned without the information needed to assess its own risk. A drill program planned against the joint surface is a drill program that can be defended in front of a technical committee, a partner, or a regulator.

The two-map workflow also changes what a successful pilot looks like. The conventional measure is the number of high-probability targets identified. The more honest measure is the number of high-probability, low-uncertainty targets identified, plus the number of high-probability, high-uncertainty regions for which a clear data acquisition plan can be designed. The second number is at least as valuable as the first, because it converts an underexplored area into a structured programme rather than a guess.

5. What it looks like in practice

The methodology described in sections 3 and 4 is delivered to clients inside a focused, hands-on engagement. This section describes how a pilot is structured in practice, what inputs are accepted, and what the client receives at the end.

At this stage, DepoDart is deliberately small, and every engagement is shaped around what the client's team actually needs. Our geoscientists and data engineers work alongside the client throughout. The platform is the tool we use to do the work; it is not, today, a self-serve product the client logs into. Over time more of this becomes self-serve, but the work that earns trust now is collaborative by design.

5.1 Inputs

The platform handles any format, any vintage, and any coordinate system. The full inventory accepted in a standard pilot covers four broad categories.

Geophysical inputs include airborne magnetic surveys in XYZ or GDB format, gravity surveys, ground and airborne electromagnetic data, induced polarisation and resistivity, and radiometric data. Geochemical inputs include rock samples, soil and sediment samples, assay tables of any schema, lithogeochemical data, and stream sediment surveys, in CSV, XLS, or proprietary formats. Geological inputs include drill logs in LAS, CSV, or DLOG format, geological maps as shapefiles or geodatabases, lithology logs, structural interpretations, and core imagery. Remote sensing inputs include multispectral and hyperspectral GeoTIFFs, SRTM and DEM elevation, Landsat and Sentinel scenes, and SAR derivatives.

The intent of the broad input acceptance is to remove the data preparation burden from the client. A pilot does not require the client to convert formats, reproject coordinate systems, or harmonise schemas. Our team handles the conversion work, alongside the client's geologists where helpful, as part of the engagement.

5.2 The three stages of the work

Every engagement moves through three stages.

Data fusion. Our team ingests and normalises multi-source geoscientific data regardless of format, coordinate system, or age. Drill logs, geophysics, geochemistry, satellite imagery, and structural maps are resolved into a unified spatial model. The client is not required to preprocess anything before sending it across.

AI prospectivity mapping. We use the platform to learn the geological signature of known mineral deposits within the dataset, then apply that learned pattern across the full area of interest. Every point on the map receives a ranked mineralisation probability score, producing a continuous prospectivity surface at district scale and below.

Ranked drill targets. Every output includes a prioritised list of drill targets with confidence tiers, estimated depths, mineral type classifications, and uncertainty bounds. Full data provenance is included so geologists can audit and defend every recommendation.

5.3 Typical engagement timeline

A pilot typically takes two to four weeks, depending on dataset volume, complexity, and how iterative the review with the client's team needs to be. The stages, in order, are data sharing (any format, any vintage, any coordinate system; NDA available before transfer), ingestion and normalisation (with a data receipt summarising what was received and how it was resolved), the AI model run, QC and geoscientist review (often in dialogue with the client's geologists), and delivery.

5.4 What the client walks away with

Every engagement returns four artifacts.

1. Prospectivity surface as a GeoTIFF. A colour-coded raster giving ore-forming probability per pixel, registered to the client's coordinate system.
2. Ranked drill targets as a PDF and CSV. A prioritised list with probability scores, mineral type classification, estimated depth, and confidence tier.
3. Uncertainty map as a GeoTIFF. A companion raster quantifying model uncertainty per pixel, registered to the same grid as the prospectivity surface.
4. Data provenance report. A full audit trail showing which input layers drove each recommendation, suitable for technical committee review.

5.5 Case studies and comparison to baseline

This is the part of the paper that, in a mature version of the document, would carry the most weight with technical readers. It is also the part that cannot be written in the absence of pilot results that DepoDart is authorised to share publicly. The placeholders below mark exactly what is missing.

> [NEEDS REAL PILOT DATA — placeholder removed before publication]
>
> One anonymised case study describing a completed pilot. Required content: jurisdiction at the level of country or province; commodity; dataset volume and modalities supplied; number of known occurrences used as training labels; resolution of the resulting prospectivity surface; number of ranked drill targets returned; brief description of the geological setting; client-permitted summary of what the pilot identified and how it informed subsequent exploration. No client names, no project names, no deposit names unless the client has explicitly authorised disclosure.

> [NEEDS REAL PILOT DATA — placeholder removed before publication]
>
> One comparison-to-baseline figure. Required content: a quantitative comparison between the DepoDart prospectivity surface and the prior exploration interpretation for the same area. Suggested form: a confusion-matrix-style summary showing how many of the known occurrences fell in the top decile, top quartile, and top half of the DepoDart prospectivity surface, alongside the same statistic for the prior interpretation. Alternatively, a pixel-wise comparison of the two surfaces with a calibration plot. The figure should be accompanied by an honest commentary on the limits of the comparison: small sample sizes for the known occurrences, the unknown frequency of false negatives, and the differences in the question each method was designed to answer.

A second case study, ideally from a different jurisdiction and a different commodity, would strengthen the section further. The same disclosure constraints apply.

6. Limitations and an honest agenda

AI prospectivity mapping does not replace geological judgment. The models identify statistically anomalous feature combinations. They do not understand geological processes. There is no causal model of mineralisation inside a gradient-boosted tree ensemble. There is a pattern recognised in the training data and projected across the area of interest. The recognition can be useful, sometimes very useful, but it is not the same thing as an explanation.

The implications of this distinction are practical. In geological settings with no analogues in the training data, model performance degrades and uncertainty is high. A model trained on the geology of one terrane and applied to a fundamentally different terrane will produce a prospectivity surface that may look plausible at a glance and is in fact unreliable. The uncertainty map is the first line of defence against this failure mode. A high-uncertainty surface in a new geological setting is the model's way of disclosing that it does not know what to do with the data it has been given. The second line of defence is the geoscientist. Every pilot output is reviewed by a qualified geologist before delivery, and the report names the geologist's review explicitly.

There are several categories of decision in which the model should not be the deciding voice. The first is any decision about whether a particular geological setting is permissive for the target commodity at all. That is a question of regional geology and of deposit-type theory, and it is upstream of anything the model can contribute. The second is any decision about the structural controls on a particular target. The model can flag a pixel as anomalous; the interpretation of why the pixel is anomalous, and whether the anomaly is consistent with a defensible structural framework, is a question for the geologist. The third is any decision about whether the project area is suitable for drilling on grounds that are not represented in the training data: land access, environmental constraints, social licence, partner agreements. The model has nothing to say about these and should not be asked.

DepoDart discloses these limits in every pilot. The disclosure is not a footnote. It is part of the report, in the same typographic register as the recommendations. The intention is to make it as easy as possible for a technical committee to read the report and to come to its own view about which recommendations to act on and which to set aside.

The agenda for the next two years follows from these limits. The team is working to extend the methodology to a wider range of deposit types, to improve the uncertainty quantification in low-data settings, to develop better disclosure formats for the audit trail, and to formalise the workflow by which qualified geoscientists review and adjust the model output. Progress in each of these areas will be described in subsequent versions of this document.

7. Conclusion

Mineral exploration is the work of deciding where to drill on the basis of incomplete information. The first generation of AI tools marketed to mining promised to resolve that incompleteness and largely failed, for reasons that were structural rather than statistical. A successor framework has to address those failures directly: opacity, calibration, generic templates, and provenance. It has to produce probabilities that behave like probabilities, uncertainty estimates that behave like uncertainty estimates, and an audit trail that a working geoscientist can defend in front of a technical committee.

The framework described in this paper is the result of taking those requirements seriously. It uses positive-unlabelled learning to model the asymmetry between known deposits and unlabelled ground, gradient-boosted ensembles to produce interpretable predictions from tabular geoscientific data, spatial cross-validation to produce honest performance estimates, Platt scaling to produce calibrated probabilities, and bootstrap resampling to produce uncertainty maps that support drill planning rather than decorating it. The whole is delivered inside a hands-on engagement that typically runs two to four weeks, with four concrete artifacts the client walks away with and a disclosure section that names what the model does not know.

Exploration teams interested in a pilot can begin the conversation at **depodart.com**.

Appendix A. Methodology references

The following references support the methodological claims made in the body of the paper. The list is selective rather than exhaustive, and is intended to allow technical readers to verify the foundations of the approach.

1. Elkan, C., and Noto, K. (2008). Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 213 to 220. ACM.
2. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), 1189 to 1232.
3. Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913 to 929.
4. Platt, J. C. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In *Advances in Large Margin Classifiers*, pages 61 to 74. MIT Press.
5. Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1), 1 to 26.
6. Bekker, J., and Davis, J. (2020). Learning from positive and unlabeled data: a survey. *Machine Learning*, 109(4), 719 to 760.
7. Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785 to 794. ACM.
8. Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing. In *Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 5372 to 5375. IEEE.
9. Lawley, C. J. M., McCafferty, A. E., Graham, G. E., Huston, D. L., Kelley, K. D., Czarnota, K., Paradis, S., Peter, J. M., Hayward, N., Barlow, M., Emsbo, P., Coyan, J., San Juan, C. A., and Gadd, M. G. (2022). Data driven prospectivity modelling of sediment hosted Zn-Pb mineral systems and their critical raw materials. *Ore Geology Reviews*, 141, 104635.
10. Carranza, E. J. M., and Laborte, A. G. (2015). Random forest predictive modelling of mineral prospectivity with small numbers of prospects and data-driven analysis of feature importance. *Computers and Geosciences*, 74, 60 to 70.

Appendix B. Glossary

Prospectivity mapping

The practice of producing a continuous spatial surface, typically a raster, in which each pixel is scored for the likelihood of hosting a mineral deposit. The surface is the principal decision support output for siting drill programmes at the district scale.

Positive-unlabelled (PU) learning

A machine learning approach designed for problems in which the training data contains confirmed positive examples and a much larger pool of unlabelled examples that may include further latent positives. The approach treats the unlabelled set as a mixture rather than as a set of negatives.

Ensemble model

A model that combines the predictions of many constituent models to produce a single output. In DepoDart's case, the ensemble is a set of gradient-boosted decision trees whose predictions are aggregated to produce the prospectivity score.

Gradient boosting

A machine learning method that builds a sequence of decision trees in which each new tree is fit to the residual errors of the trees already in the ensemble. The result is a strong predictor built from many weak learners. Formalised by Friedman (2001).

Platt scaling

A post hoc calibration method in which a logistic regression is fit to a model's raw scores against held-out validation labels, producing a calibrated probability output. Introduced by Platt (1999) for support vector machines and now widely applied to ensemble methods.

Epistemic uncertainty

Uncertainty arising from the model's limited knowledge of the underlying process. In principle, epistemic uncertainty decreases as more training data becomes available. Distinguished from aleatoric uncertainty, which arises from inherent randomness in the process and does not decrease with more data.

Aleatoric uncertainty

Uncertainty arising from the inherent variability of the process being modelled. Cannot be reduced by collecting more data of the same kind. In a mineral exploration context, this includes the unavoidable noise in geophysical and geochemical measurements.

Spatial cross-validation

A validation regime in which the holdout set is constructed as geographic blocks rather than as a random sample of pixels. The blocks are sized to exceed the spatial autocorrelation length of the data, so that the validation set is statistically independent of the training set in the spatial sense.

Leave-one-out validation

A validation regime in which the model is retrained with one positive example held out, the held-out example's location is checked against the resulting prediction surface, and the procedure is repeated for every positive example in turn. Particularly useful in low-positive settings characteristic of mineral exploration.

Multivariate anomaly index

A geochemical feature constructed from the joint distribution of multiple elements rather than from any single element. Captures relationships between elements that are diagnostic of particular mineralisation styles and that are not visible in univariate anomaly maps.

Tilt derivative

A magnetic field derivative defined as the arctangent of the ratio of the vertical derivative to the horizontal gradient. Used to enhance the visibility of magnetic source edges and to suppress amplitude variations between shallow and deep sources.

Reduced-to-pole anomaly

A transformation applied to total magnetic intensity data that mathematically relocates the magnetic source to a magnetic pole position, producing an anomaly map in which the peak is centred over the source rather than offset by the inducing field direction.

MINFILE

The mineral occurrence database maintained by the British Columbia Geological Survey. A standard source of positive labels for prospectivity modelling in British Columbia.

MRDS

The Mineral Resources Data System maintained by the United States Geological Survey. A standard source of mineral occurrence records for prospectivity modelling in the United States.

MIRIS

An umbrella term used in this paper for the Mineral Industry Resource Information System and equivalent national or state-level mineral occurrence registries outside North America. Conventions vary by jurisdiction.

Source: marketing/_source/brief.md · Version 0.1 · Draft, review before publication