BEARD

Borehole Enhanced and Automated Realtime description

& Other AI Applications developed recently by Geolearn

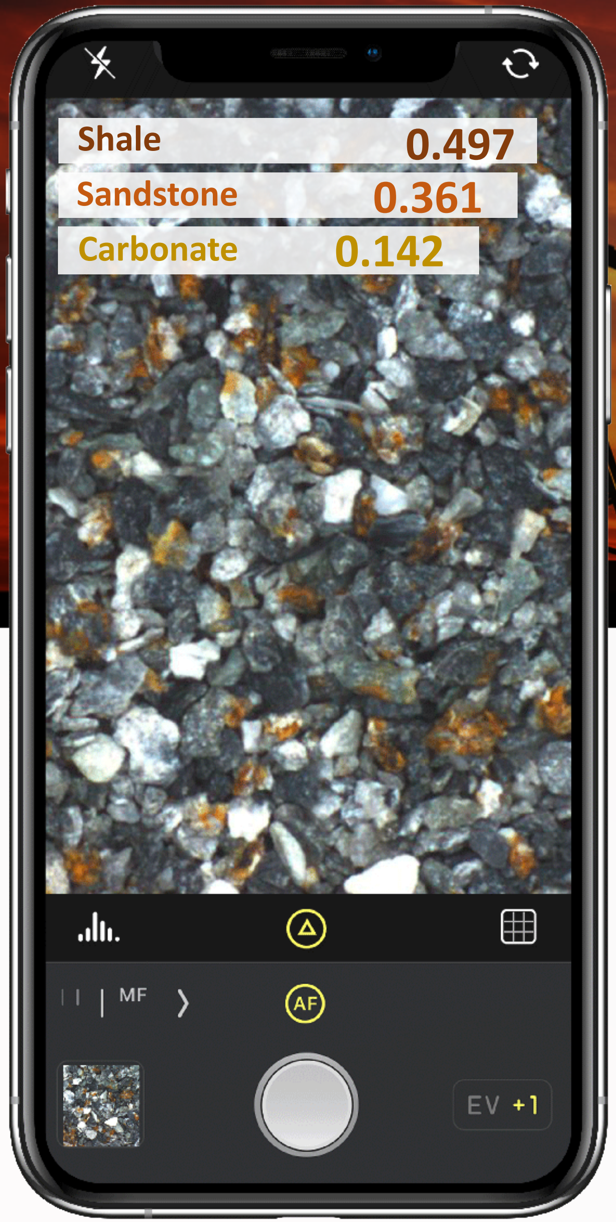
Jean-Sebastien Marcil, Product Development Manager
jsmarcil@geolearn.ai

July 23nd, 2019 – Denver, Colorado

www.geolearn.ai

Borehole Enhanced and Automated Realtime Description

BEARD



database traininig

Machine learning algorithms are trained on a database of core libraries photos

deployment on the field

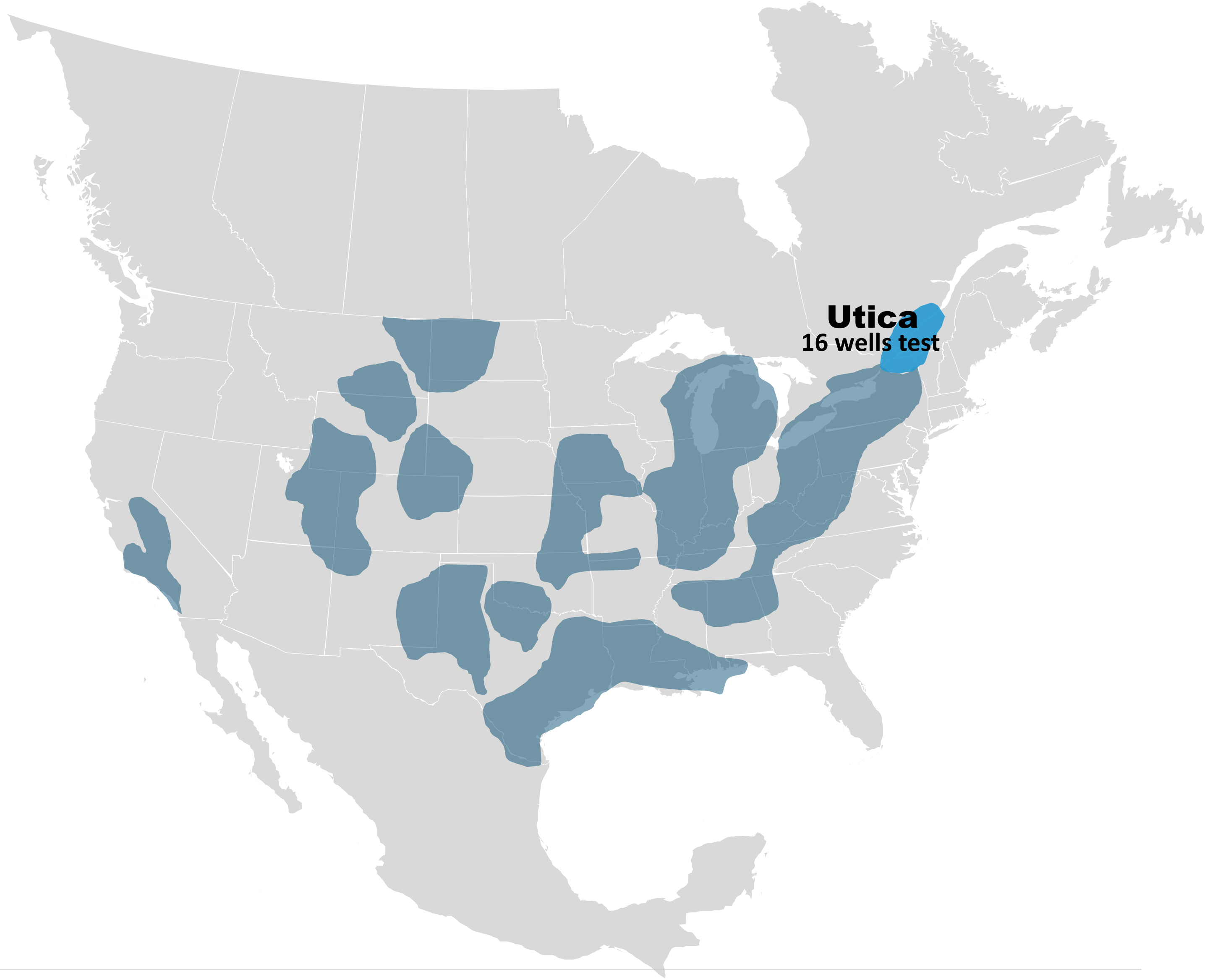
BEARD automatically describe rock chips from photos

Why we are here

BEARD actually under development/deployment

Short term potential for BEARD

Long term potential for BEARD



AAPG UPITCH INVENT AND INVEST

Meet the team

Martin Blouin, PhD

Cofounder and CEO

As a CEO, researcher and scientist, Martin leads the majority of the projects related to artificial intelligence.
He has a 10-year experience in geophysics and data integration applied to geosciences.

Lorenzo Perozzi, PhD

Cofounder and VP

As a researcher and entrepreneur committed to ethics in the implementation of AI, Lorenzo was recruited to participate in the co-construction of the Montreal Declaration for a Responsible Development of Artificial Intelligence . He ensures that this aspect is being reflected in all activities of Geolearn.

Jérome Simon

Technical Director

Jérome ensures the integrity of databases and the reliability of the results. Propelled by his experience with digital methods in science, he develops interpretation techniques combining computer vision and artificial intelligence

JS Marcil, ING., MSc

Product Development Manager

With 20 years of E&P experience, JS is in charge of product development at Geolearn. Based on its field background, JS directs product development activities to Geolearn's various customers, from design through to production

Our vision

Geodata, **enhanced!**

We research, design & build AI technologies to help our clients solve real-world problems using geodata and reducing uncertainties.

UPITCH INVENT AND INVEST

Our mission

Delivering more than products and services building blocks for a growing community

At Geolearn, we believe that Artificial Intelligence has the power to disrupt the geoscience industry in a positive fashion, if done the right way. Our team is dedicated at making sure your data go the extra mile with those technologies and at getting you involved in the process.

AI in geosciences

AI and in particular machine and deep learning have been applied with success for different geosciences tasks such as:

- geomechanical properties characterization (Keynejad et al.(2017));
- automatic fault interpretation (Araya-Polo et al.(2017), Bugge et al.(2018), Guitton et al.(2017));
- machine learning to predict core physical properties (Caté et al.(2017)) and;
- geophysical data inversion (Araya-Polo et al (2018), Smith et al.(2010))

UPITCH
INVENT AND INVEST

Machine learning as a tool for geologists

The Leading Edge · April 2017

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 have already been discovered and mature mining camps have started to dry out, the discovery of new mineral deposits has become increasingly expensive and risky in the last 15 years (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 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 amount 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. Similarly to 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. A total of eight elements were analyzed (Ag, As, Au, Cu, Fe, Ni, Pb, and Zn), and a significant part of the measurements are 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.

http://dx.doi.org/10.1190/tle36030064.1.

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 comprised 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 sample (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 a lot 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, which 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 six 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 algorithms parameters and evaluates 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

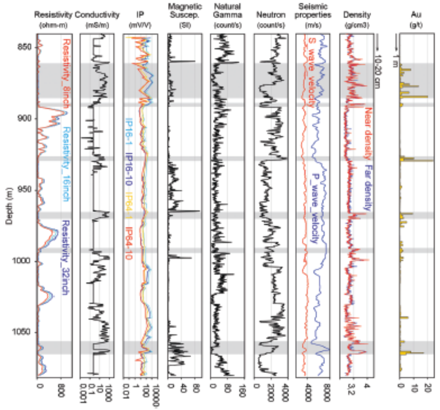


Figure 1. Typical Lalor data set. Physical properties logs are sampled at 10–20 cm, and assay composites are 1 m long. Gray background indicates the ore lenses defined by Hudbay.

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. The prediction results obtained on one of the test drill holes are presented in Figure 5. Intervals located above an ~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 be 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

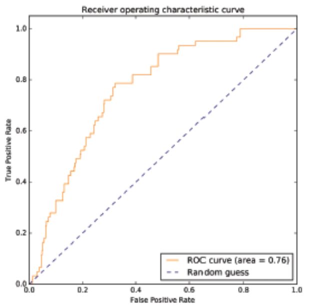


Figure 4. Receiver operating characteristic (ROC) curve of the prediction of the presence of gold in test drill holes.

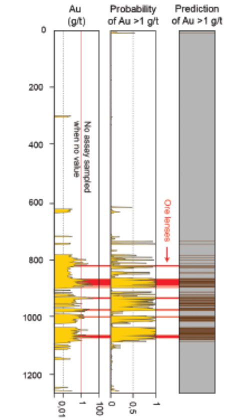
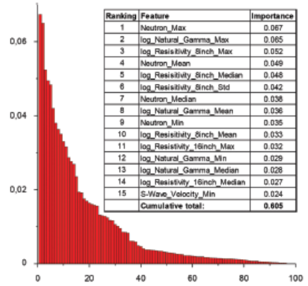


Figure 5. Results of the prediction on a test drill hole. The values of gold concentration on assayed intervals (left), the predicted probability of Au>1 g/t (middle), and the result of the classification (right) are presented. The location of ore lenses, defined by Hudbay during the exploration stage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



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.

GEOLEARN

Example of success: Predikor

A change of pace in core logging

Predikor

Results

Photo

Model: Lithologie principale

Table: Predikor-litho

From	To	Code 2		Prediction
45.27	45.66	I2J	Predikor-litho2	I2J - 75%, TLB - 25%
45.66	45.85	TLB	Predikor-litho2	TLB - 73%, I2J - 27%
45.85	47.19	TLB	Predikor-litho2	TLB - 80%, I2J - 20%
47.19	47.38	TLB	Predikor-litho2	TLB - 75%, I2J - 25%
47.38	48.13	TLB	Predikor-litho2	TLB - 85%, I2J - 15%
48.13	48.29	TLB	Predikor-litho2	TLB - 91%, I2J - 9%
48.29	49.88	TLB	Predikor-litho2	TLB - 100%, I2J - 0%
49.88	50.42	TLB	Predikor-litho2	TLB - 100%, I2J - 0%
50.42	57.26	TLB	Predikor-litho2	TLB - 100%, I2J - 0%
57.26	57.45	TLB	Predikor-litho2	TLB - 85%, I2J - 15%
57.45	58.40	TLB	Predikor-litho2	TLB - 96%, I2J - 4%
58.40	58.59	TLB	Predikor-litho2	TLB - 90%, I2J - 10%
58.59	59.53	TLB	Predikor-litho2	TLB - 100%, I2J - 0%

Depth referenced core photos

Predictions

GEOLEARN

About visual descriptions of core

- Subjective and qualitative to semi-quantitative at most.
- Descriptions are not reproducible.
- Quality assessment and quality control is hard to implement.
- Interpretation is frozen in time.

Automated drill core descriptions opportunities

✓ Provide geologists with a highly valuable tool for logging :

- Speed up the process;
- Standardize descriptions;
- Increase description quality.

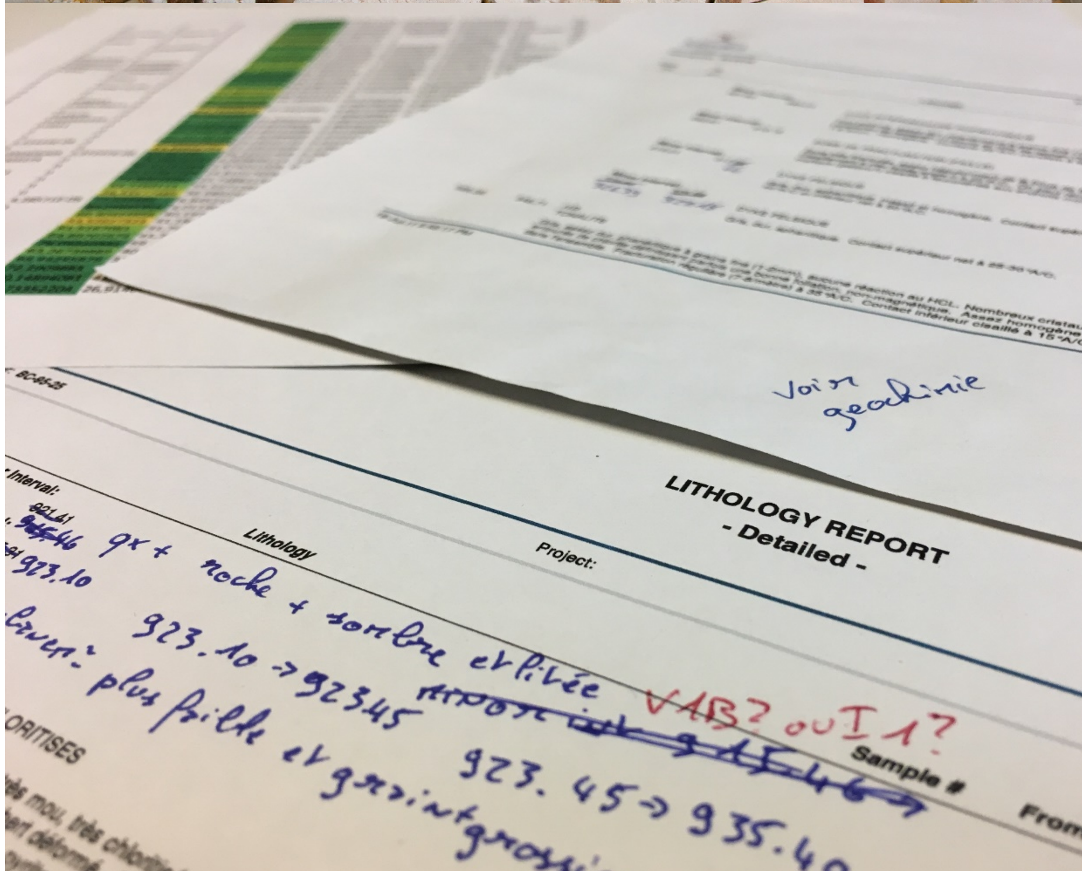
Need for data

Acquiring new high value data:

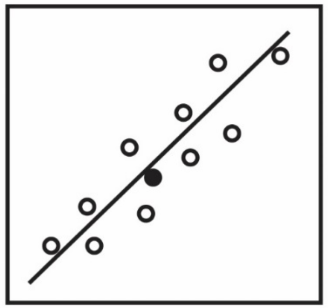
- Ex: Hyperspectral images;
- Ex: Geophysical logs.

Use available data:

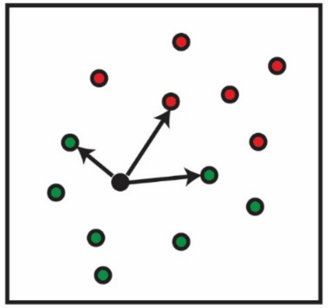
- Core boxes photos;
- Compiled geological descriptions.



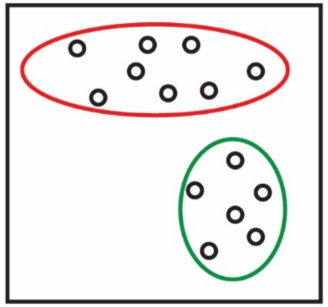
Data-driven approaches



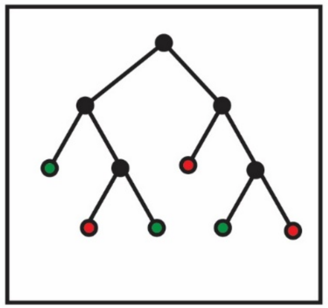
Regression



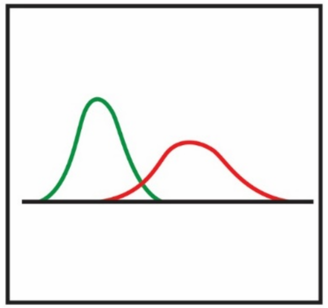
Euclidian Distances



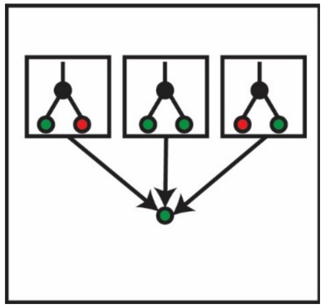
Clustering



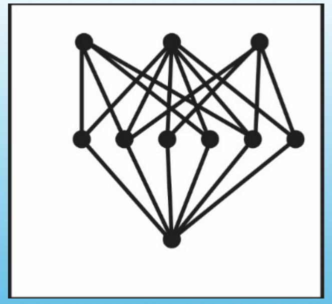
Decision Trees



Bayesian Probability

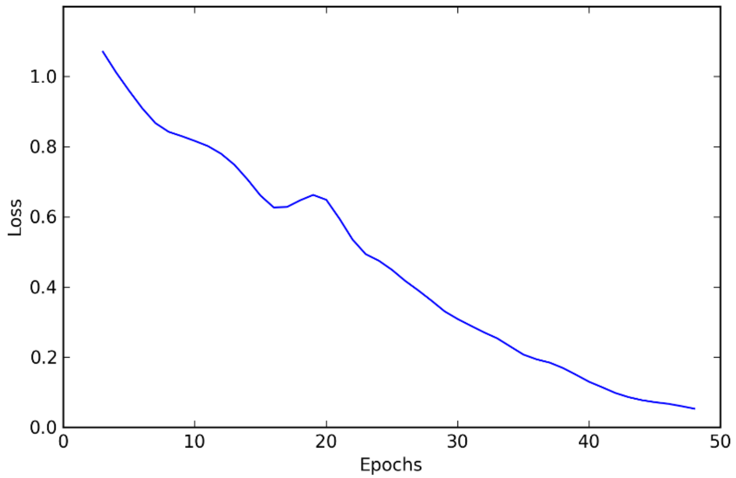


Ensemble Methods

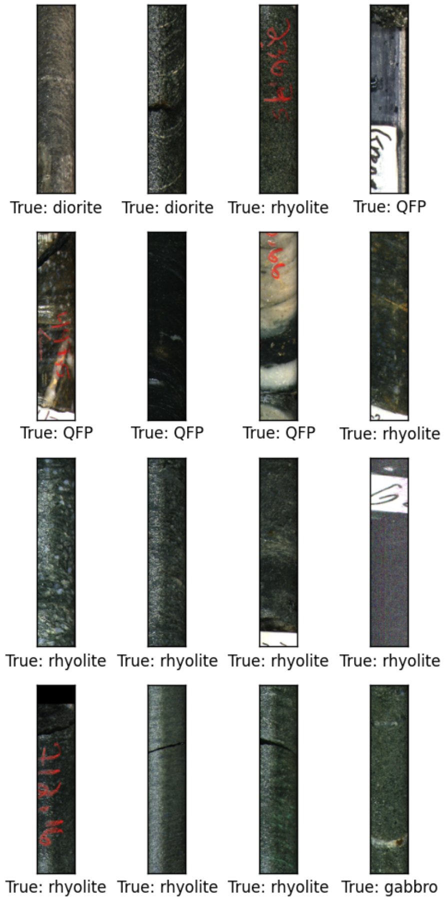


Neural networks (NN) and DNN

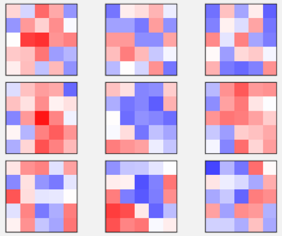
Training the network



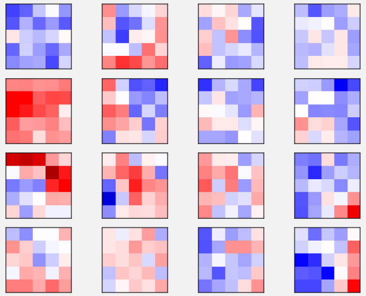
Training images



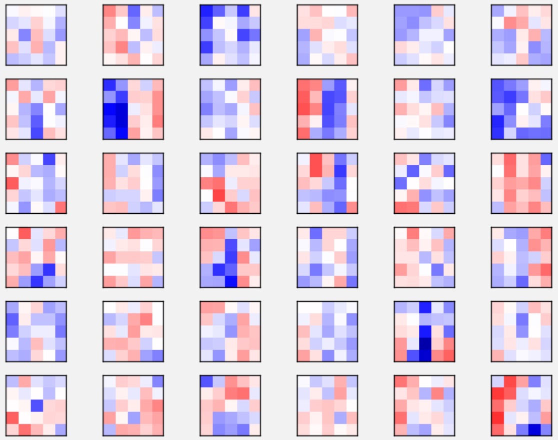
Filter-Weights Layer 1
(5 x 5 pixels)



Filter-Weights Layer 2
(5 x 5 pixels)

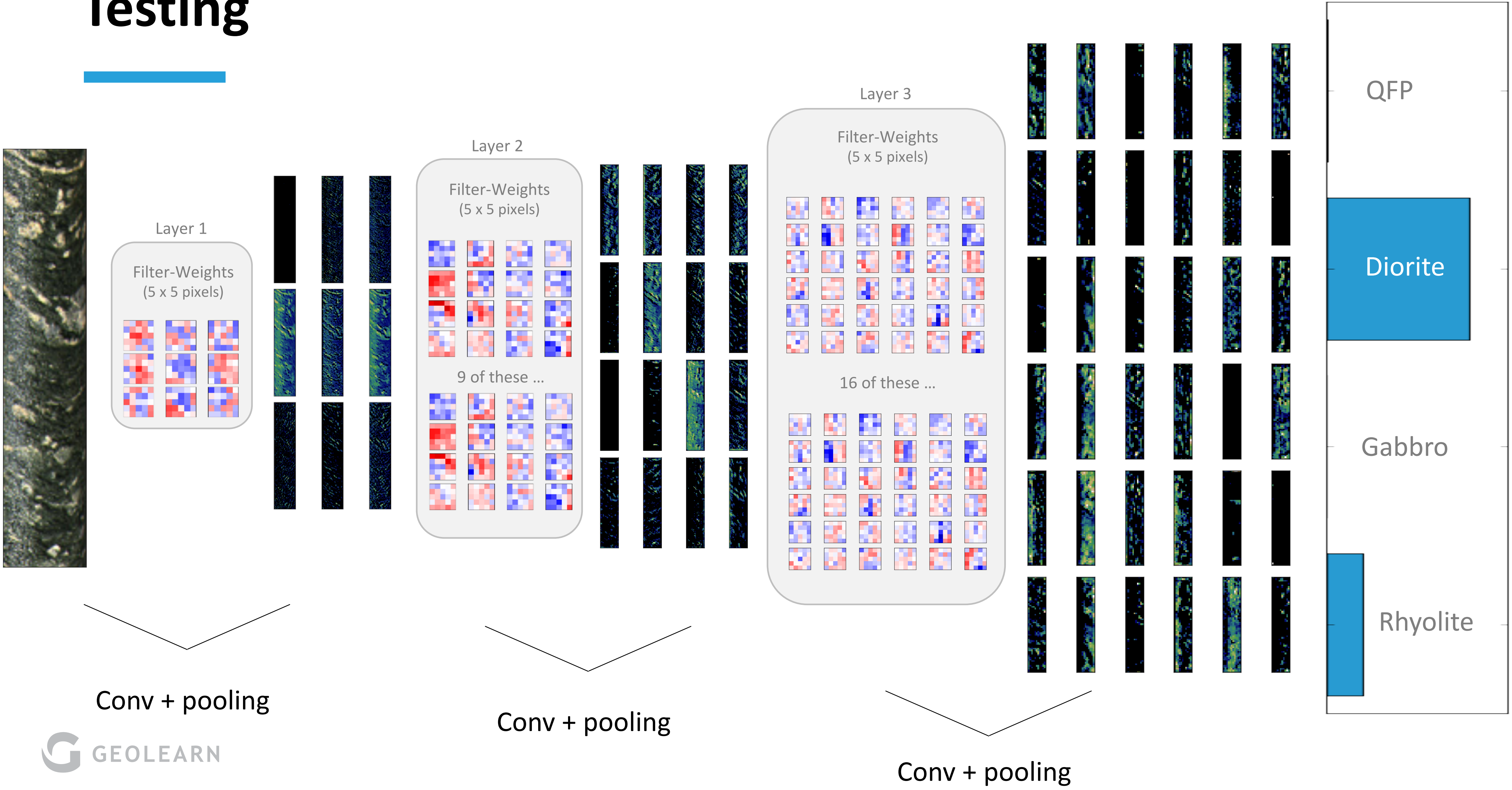


Filter-Weights Layer 3
(5 x 5 pixels)



GEOLEARN

Testing



GEOLEARN

Automatic descriptions from photographs

Successes:

- ✓ Standardization.
- ✓ Increased description pace.

Limits:

- ✓ Size matters.
- ✓ Garbage in, garbage out.
- ✓ Requires linearized photos.

Other applications developed

UnBOX

Automatic linearization of box core pictures

Automatically detection of cores in pictures

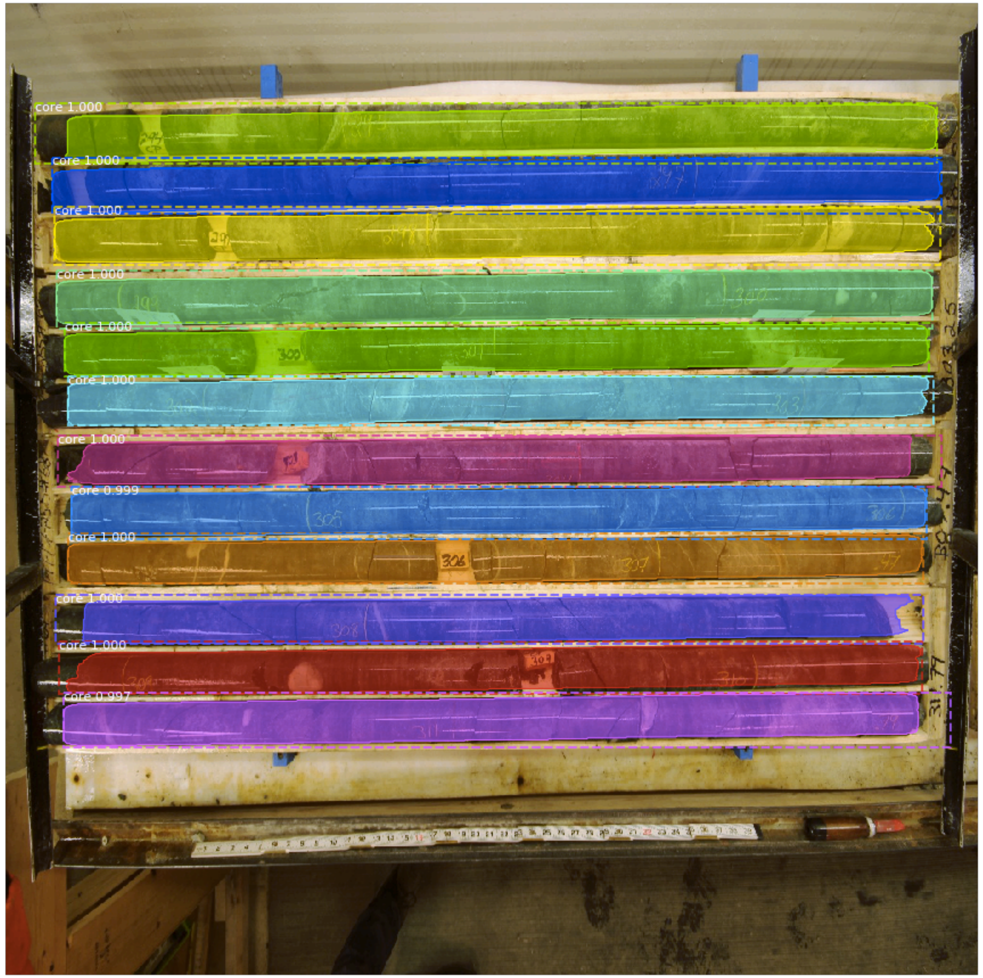
Core linearization

- Why is it necessary?
  - People expect linear predictions along the core;
  - Allows better interpretations.
- But tedious and slow if done by hand.

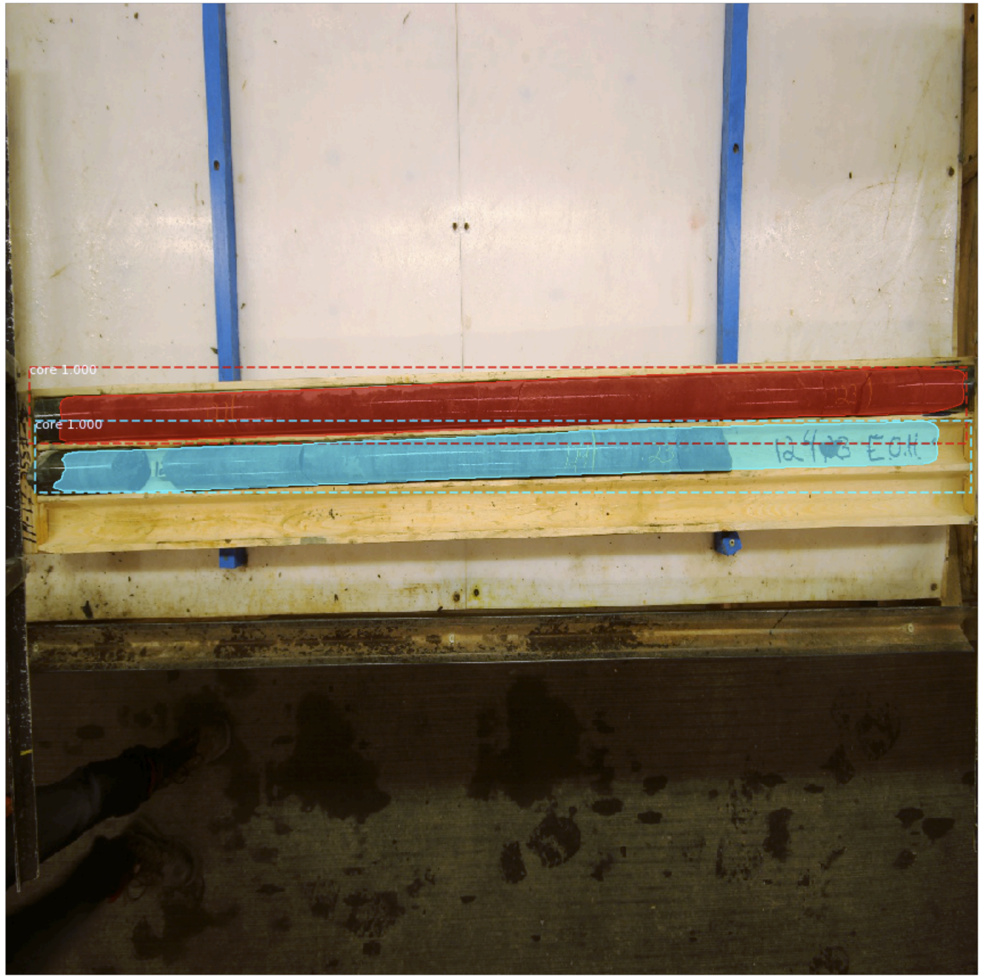
✓ Object detection with deep learning has come a long way.

Automatic core linearization

Predictions

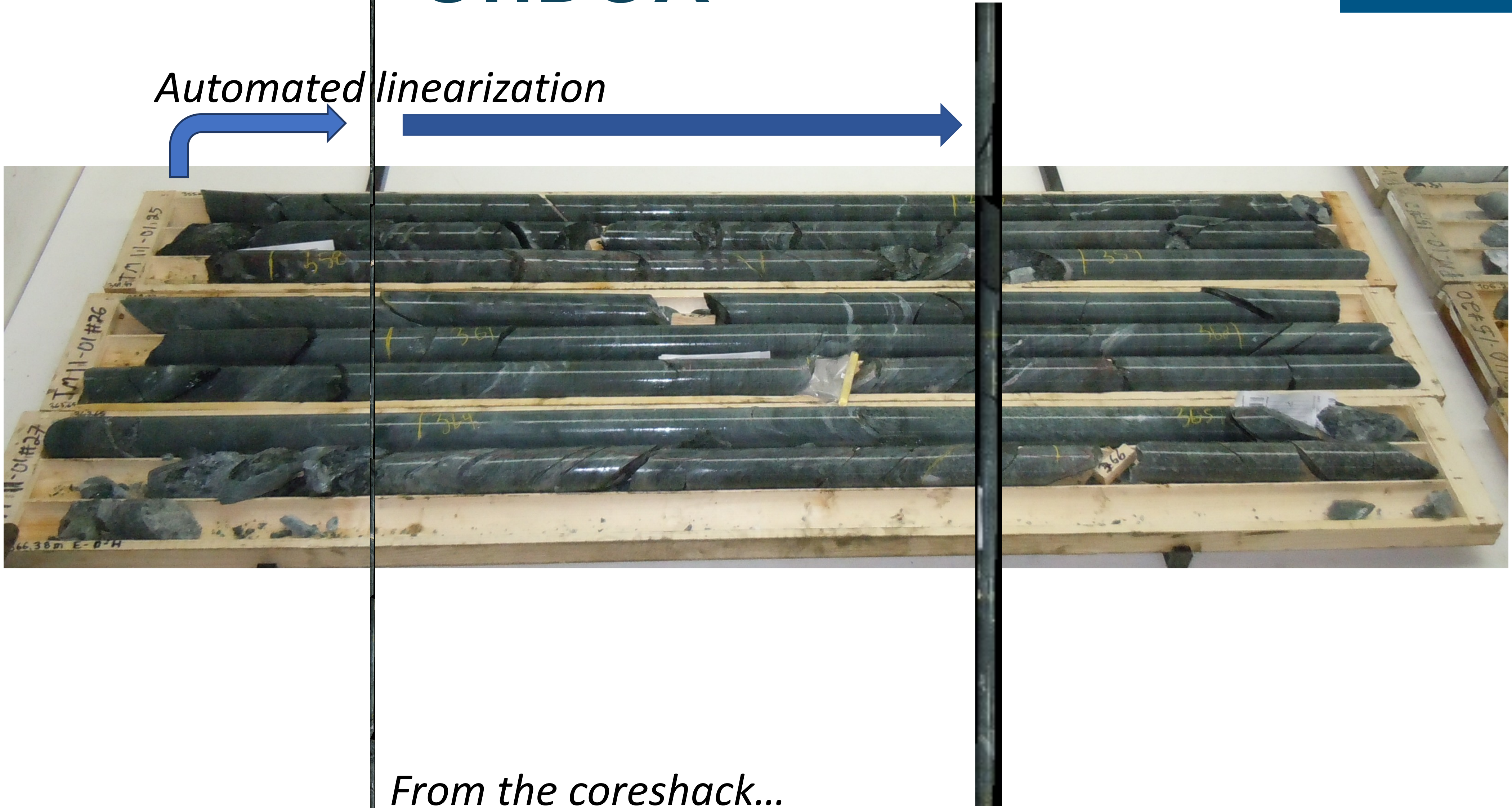


Predictions



UnBOX

Automated linearization

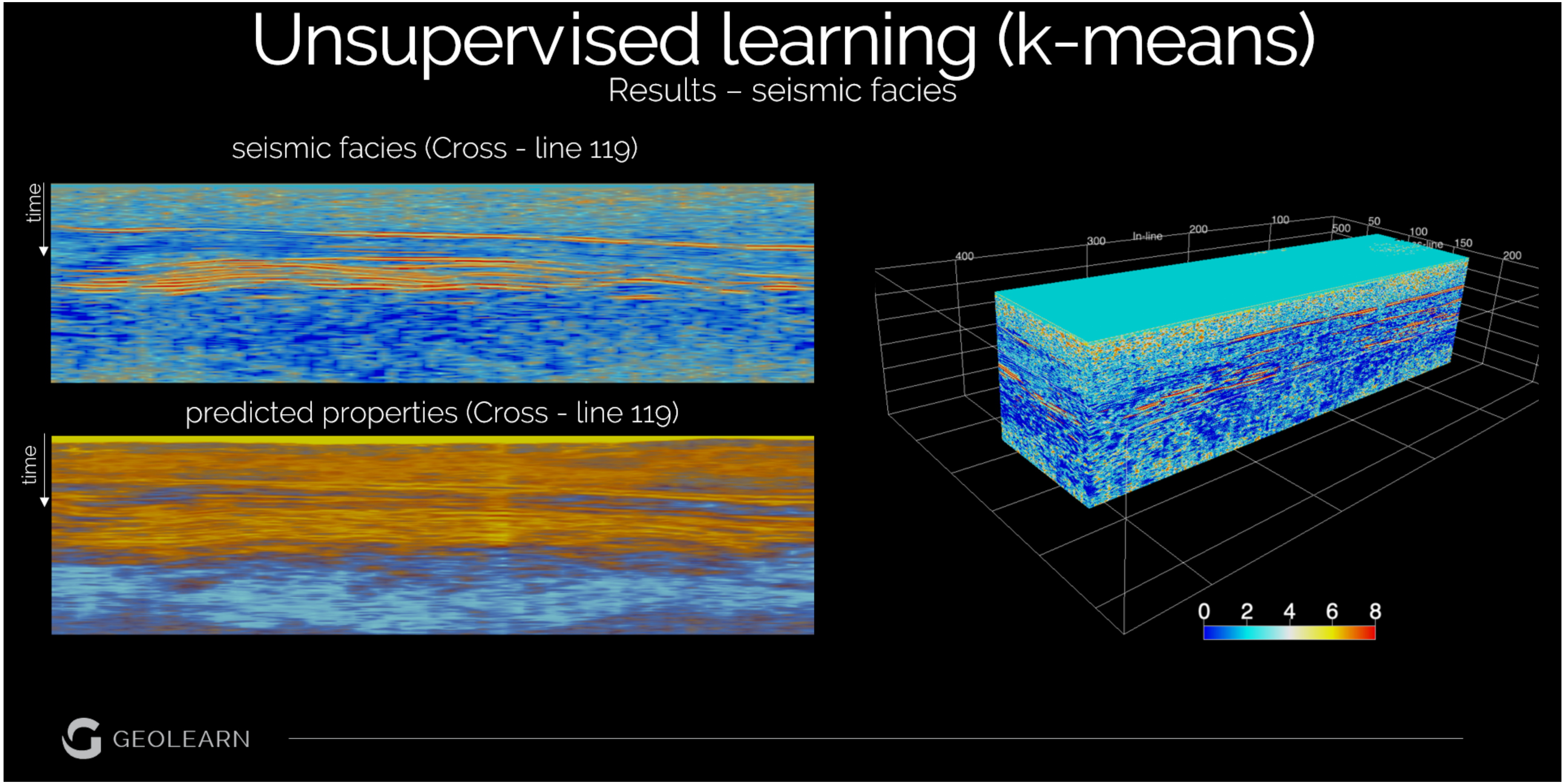


From the coreshack…

…to a geolocalized linear image of your reservoir rock… in 3 seconds.

GEOLEARN

Seismic Advanced Processing



Our services

AI-DRIVEN DATA ANALYTICS

Geolearn uses deep and machine learning as well geostatistics to retrieve more information from your data.

DATA VISUALIZATION

At Geolearn we do not simply plot data, we tell a story with it!

SAAS SOLUTIONS

Geolearn delivers tailor-made applications that can handle basic data processing or more elaborate cognitive computations and deliver insightful visualization.

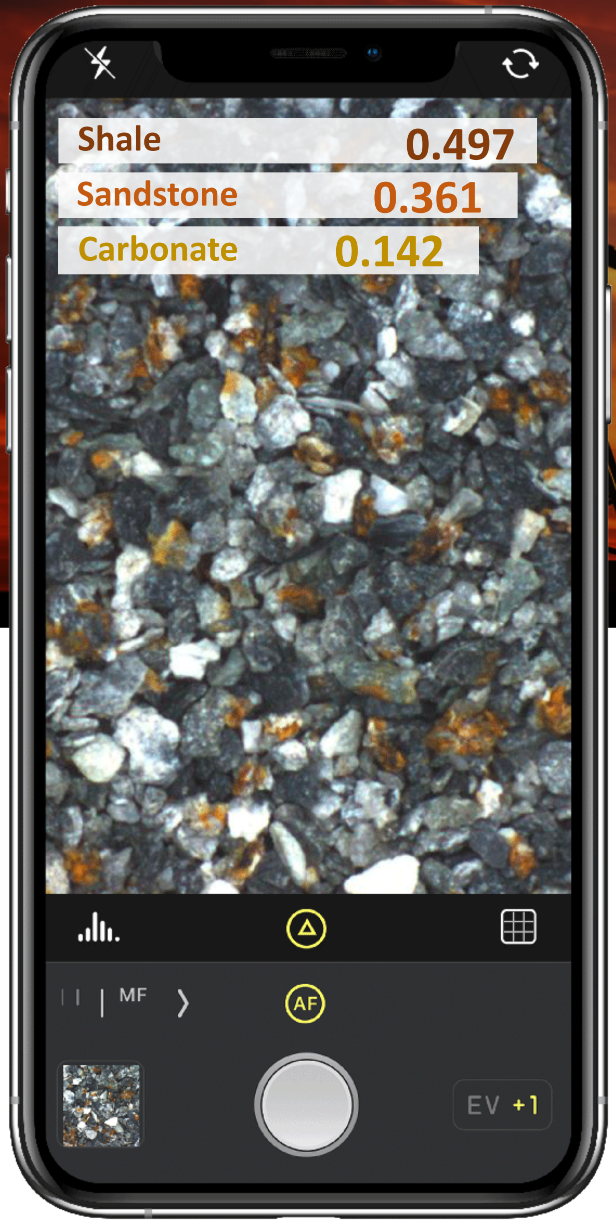
KNOWLEDGE SHARING

We offer training in machine learning, geostatistics and python applied to the geoscience industry. We also organize "meet-up style" events to bring together experts from all field and get the discussion going.

We are proud to be among the qualified suppliers to provide Canada with responsible and effective AI services, solutions and products

Training offered by Geolearn

Introduction to geoscientific computing 2 day	**Machine Learning in geoscience: from theory to practice** 3 day
Introduction to Python	Introduction to machine learning;
Data manipulation in Python	Database and input data
IO data from multiple format (.xlsx, csv, ascii, …)	Feature engineering.
Beautiful plotting in Python	Unsupervised methods: applications and exercises for geoscience;
Introduction to Python libraries for geocientific applications	Supervised learning: from linear regression to ensemble methods.
End-to-end exercises applied to geology and geophysics.	From data to models: end-to-end custom exercise (possible BYOD: Bring your own data)



Borehole Enhanced and Automated Realtime Description

BEARD

database training

Machine learning algorithms are trained on a database of core libraries photos

deployment on the field

BEARD automatically describe rock chips from photos

UPITCH INVENT AND INVEST

Timeline

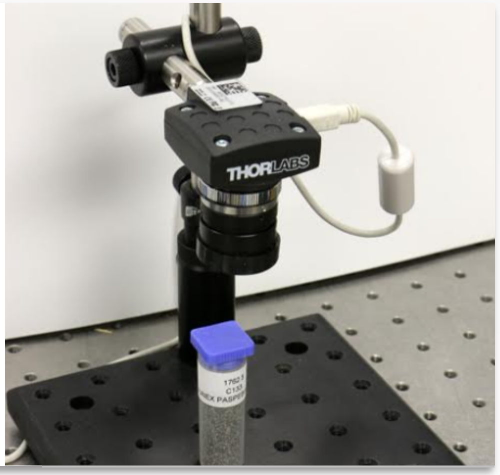
GEOLEARN

2019

July

Image capture suite ready

Fast sampling of core shack database



September

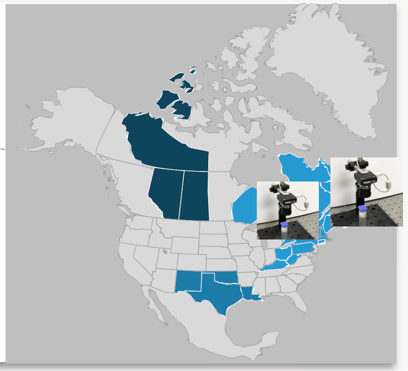
First public core shack test completed

Compilation of log and image for Utica Shales and other formations of Quebec's St. Lawrence Lowlands

October 2019

Replicating the acquisition tool

Getting a robust workflow to capture more data in Nova Scotia and Ontario



2020

November 2019

ML algorithms parametrization

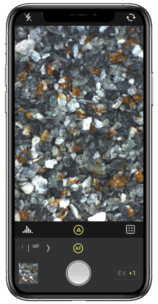
Tune and train ML algorithms to create predictive models

UPITCH
INVENT AND INVEST

February 2020

Hand-held prototype
Ready for pre-launch

Getting the predictive tool to our beta tester



April 2020

BEARD

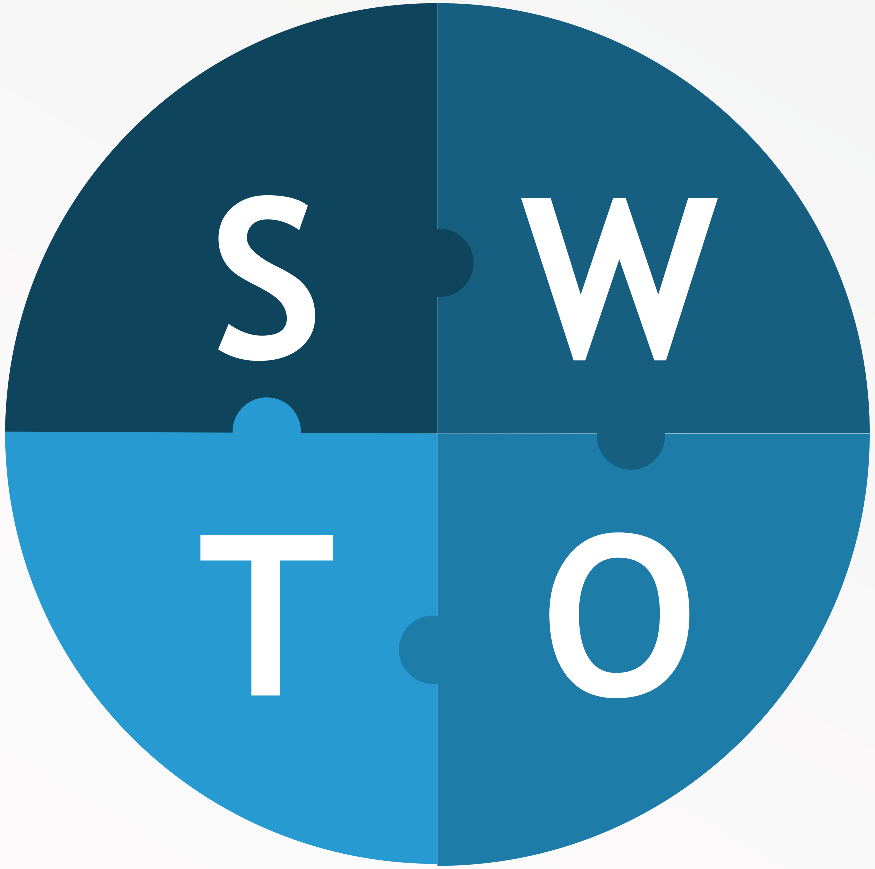
Commercial launch

Active research of public and private partners with core shacks

SWOT Analysis

Strengths

- Sound experience in AI applied to geosciences
- In-house workflows already adapted for this product



Weakness

- Resource limitations
- Geographic location

Opportunities

- Underserved market
- Few (or no) competitors
- Emerging need for the product
- Rapid growth

Threats

- Emerging competitors
- Rapid saturation of the market?
- Big players with massive resources

BEARD - Action Steps

PROOF OF CONCEPT

Demonstrate the technical validity of the approach

GATHER MORE DATA

Get deals with public and private coreshack owners

BUILD A COOL TOOL

Get the predictive model in a hand-held format that will facilitate and optimize field work

BEARD

EXPAND THE MARKET

Grow geographically but also target new applications like geotechnical drilling

www.geolearn.ai
info@geolearn.ai

Geolearn
490, rue de la Couronne,
4th floor
Quebec City, Quebec
CANADA

GEOLEARN.AI

AI Applications developed for geoscience

Martin Blouin, Ph.D., CEO and Lead AI Developer
martin.blouin@geolearn.ai

Lorenzo Perozzi, Ph.D., VP and Operations Lead
lorenzo.perozzi@geolearn.ai

Jean-Sebastien Marcil, eng., M.Sc., Product Development Manager
jsmarcil@geolearn.ai

Jerome Simon, B.Sc., Technical Director
jerome.simon@geolearn.ai

www.geolearn.ai