STRATUM AI

Stratum Machine Learning

Farzi Yusufali & Daniel Mogilny

An overview of the technical problem of mining and the artificial intelligence (AI) solution.

www.stratum.ai founders@stratum.ai

Data

There are two kinds of data, historical “production data” and “drillhole data”. Production data is essentially what has already been mined out while drillhole data are discrete samples collected by drilling a narrow rod in the ground in areas that have not been mined. When our neural network is deployed, it only has access to drillhole data for predicting a region. However, during the learning process, it's possible to integrate historical production data to better understand how geology is deposited.

Even when learning with production data, the data is abundant but sparse. If the input data is analogous to a picture (200x200 pixels), in most cases only 100-1000 pixels are filled in with varying density. The good news is that you sometimes have up to millions of potential pictures to learn from. An individual mine contains 200k-800k unique production data samples and mines of similar geology can be bundled together to create larger datasets.

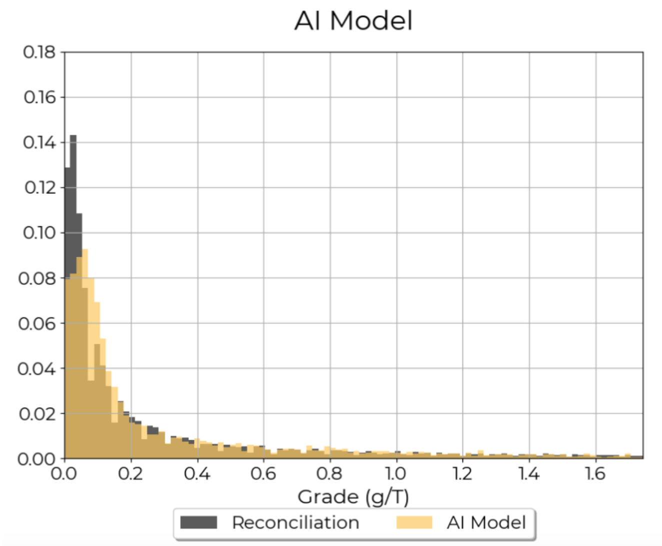
Industry Standard

The industry standard is **kriging**. Snippet from wikipedia:

The basic idea of kriging is to predict the value of a function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point. The method is mathematically closely related to regression analysis. Both theories derive a best linear unbiased estimator, based on assumptions on covariances, make use of Gauss–Markov theorem to prove independence of the estimate and error, and make use of very similar formulae. Even so, they are useful in different frameworks: kriging is made for estimation of a single realization of a random field, while regression models are based on multiple observations of a multivariate data set.

Architecture

The basic premise is that given input drillhole data, predict the gold grade to best fit the individual prediction and the dataset distribution. The dataset distribution loosely looks like a decaying exponential.



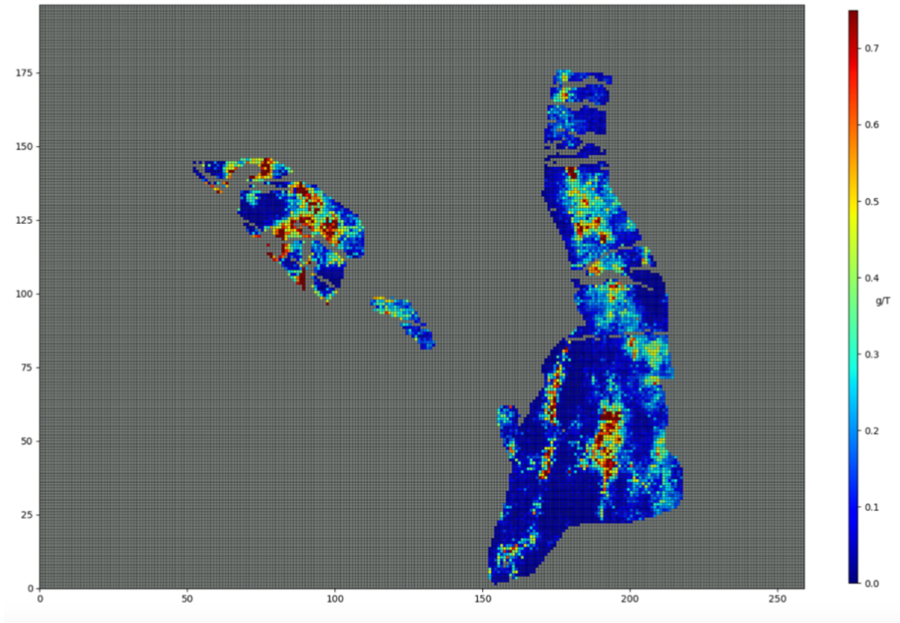
The input into the model is the surrounding drillhole samples of a certain x,y,z point the model is trying to predict. These points are bucketed together into a 3D grid to broadly look like a 3D picture composed of distinct pixels with 0.2-5% density. This has demonstrated to be advantageous to flat encoding (i.e delta(x,y,z), g for all surrounding samples) as it allows the model to aggregate local patterns before making the final prediction.

The varying density and high sparsity means that data augmentation techniques are extremely crucial and a constant point of innovation for the company. Some techniques include exploiting symmetry, sampling, negative sampling, varying encoding techniques (to distinguish no data with 0% grade if relevant), input data noise etc.

The actual neural network structure is somewhat dependent on hyper parameters relevant to mine geology and data availability. Generally, structures designed for large dataset image classification problems (variations of DenseNet, ResNet) have proven to be the most consistent performers on grade prediction tasks.

In some cases (typically extreme variation), the problem may be subdivided and separate neural networks may be tasked with addressing different questions about the geology (i.e is it high grade? How high is the grade?)

Once the neural network is designed, it is tasked with predicting a map of unknown points to create maps that look like this:



What Drives Performance Advantage?

i) Learning from historical data

Patterns from historical data give insight on how minerals are deposited which can be leveraged to create better predictive models. The industry standard does not utilize this information and as an interpolative method has no ability to, particularly as this data sometimes follows somewhat different data distributions.

ii) Learning from cross-deposit data

Stratum aggregates data across different customers to create a higher quality learning environment that the industry standard models do not have access to. This allows it to leverage more broad information about geology to make more accurate predictions, particularly relevant with newer mines with less historical data.

iii) Ability to fit complex patterns

A deep neural network can learn more complex non-linear patterns over a linear interpolative tool like kriging.

iv) Non-gold data for predicting gold

Kriging takes only the same metal into input as the one being predicted. Neural networks can take other information into account like secondary metals (silver in a gold mine), trace metals (mercury traces) and lithology (classification of rock type), which provide access to latent information that can be leveraged for more accurate predictions.