Machine Learning – Limit the Bias and Validate for Value

J Coombes[1] and T Tran[2]

1.
FAusIMM, Head Innovation, Snowden Group, Perth WA 6000. Email: Jacqui.coombes@snowdengroup.com

2.
Geostatistician, Snowden Group, Perth WA 6000. Email: Tuyet.Tran@snowdengroup.com

ABSTRACT

Machine Learning has emerged as an apparent panacea for managing, processing and extracting knowledge from big data. The ability to drive geological insight from multi-variate data, potentially collected in real time, has enormous implications for resource estimation and real-time mining decisions.

Several machine learning algorithms have emerged to process the advancing data avalanche and include K-mean clustering, K-mediod clustering, Mean-Shift Clustering, Gaussian Mixed Models (GMM), Hierarchical Clustering, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Agglomerative Hierarchical Clustering.

Not all machine Learning algorithms are equal and nor can a single machine learning algorithm apply universally. Moreover, as Amazon's recruitment system and biases in Google searches demonstrate not all machine learning algorithms perform as expected, overfitting and bias considerations can provide misleading results.

So how does one select the right method for the data or circumstance and ensure outcomes are valid? When is it appropriate to reject a Machine Learning outcome or continue to pursue improvements in the settings?

This paper brings clarity to the discussion by comparing results from several methods as applied to typical multivariate assay data, including analyses and validations from geological, geochemistry and geometallurgical examples. The intent is to help mineral industry professionals frame their questions to quality outcomes from machine learning.

The innovation and speed of advancing technology provides enormous opportunity for our industry. For us to realise true benefit from machine learning, our mining professionals must not be blindsided by technologies' apparent brilliance, and instead appropriately challenge technology to ensure underlying scientific principles are honoured, and that innovation brings its intended value: the ability to extract additional significance without misleading, and by enabling enhanced efficiency and quality in machine learning applications.