ML and AI for Resource Estimation: What Could Possibly Go Wrong? Nothing! Everything!

M Nimmo

Principal Data Scientist, CSA Global, Brisbane, QLD, 4000. Matthew.nimmo@casglobal.com:

Keywords: Machine Learning, ML, Artificial Intelligence, AI, Density Estimation, Mineral Resource Estimation, Pitfalls, Process Model.

# ABSTRACT

Machine Learning (ML), Artificial Intelligence (AI), and Swarm Intelligence (SI) techniques are extremely powerful tools for building predictive and generative models. ML can be used for building highly accurate regression and classification models. But without careful data science and statistical analysis a regression or classification model could be highly biased and completely wrong – and we may not even realise.

An example density estimation task will be used to illustrate the potential gains and possible pitfalls in using ML and AI in Mineral Resource estimation. What could possibly go wrong? Nothing! For most simple regression or classification problems, very little. For density estimation using the global mean may be all that is possible, but the resource classification would need to reflect the uncertainty in global estimation of tonnage. Everything! From data collection and measurement, from selecting ML and AI techniques and designing the analysis workflow to our own cognitive bias. The things that could go wrong is vast and may include but not limited to asking the wrong question, not asking questions, data bias, measurement errors, insufficient data coverage, not collecting the right variables, not collecting enough observations, collecting clustered samples, unbalanced data, filtering out data without reason (outliers), clipping data (outliers), relying on automatic feature selection, including irrelevant variables, excluding relevant variables, geological interpretation, changing context, blindly following best practice, splitting small datasets, sub-setting large datasets, using the wrong tool for the task, insufficient budget to complete the analysis, not allowing enough time for testing and experimentation (the science), focussing too much on building the model (the engineering), not learning from the data, our assumptions, our preconceptions, and our skill.