Using Machine Learning-Based Data Factory to Unlock Mining in Australia for Environmental, Social and Corporate Governance (ESG)

Introduction

The road to net zero requires a lot of raw materials from the mining industry. Renewable energy systems for solar, hydro, and wind need to be built to support the transition. Among the many metals critical to technology and infrastructure necessary for new energy, copper is highly sought after thanks to its conductive efficiency making it an irreplaceable element of any electrical equipment. Therefore, it is projected that by 2050, the demand for copper will reach more than 53 million metric tons. This is "more than all the copper consumed in the world between 1900 and 2021". Given the above, copper price spikes, and copper supply challenges are to be expected (Bonakdarpour & Bailey, 2022). Hence, it is crucial to optimize the way copper is mined in order to meet future demands, accelerate the energy transition and execute the plans of stakeholders to achieve Environmental, Social and Corporate Governance (ESG) targets.

Optimization of copper mining exploration and operations starts with the capability to easily make decisions and gain insights using the organizations' data. However, this data is often unstructured, scattered, and unsearchable. To extract, manage and sustainably utilize all these unstructured data, a digital data factory composed of an orchestration of Machine Learning (ML) pipelines, data tracking, and monitoring services, has been implemented on a subset of data from the Geological Survey of Queensland (GSQ) in Australia. Utilizing the ML-based Data factory approach, this paper highlights how mining information from the GSQ can be analyzed, unlocked, and used in optimizing the various stages of the copper mining operation such as exploration, mining operation, copper ore processing, reclamation and safety, health, and environmental control.

Methodology

Unstructured data from the GSQ contains scientific reports, borehole completion reports, publications, journals, mining datasets, map collections, and mining records. The documents are highly technical and spread over decades of mining operations making manual human interpretation and data extraction challenging. The dataset covers 62 years of mining operation in Queensland and has been ingested in the data factory at a rate of 3,000 pages and 4,000 images per day through its scalable automated ML pipeline and big data capabilities. The steps of the processing include uploading the data to the cloud, audit of the data, text/image extraction, image classification, and table export capabilities as seen in Figure 1. The features of this ML-based data factory transform voluminous unstructured data into structured data that are readily accessible through a cloud-based application for text, image, and knowledge search. The data factory's features have applications that can be extended to all copper mining stages.

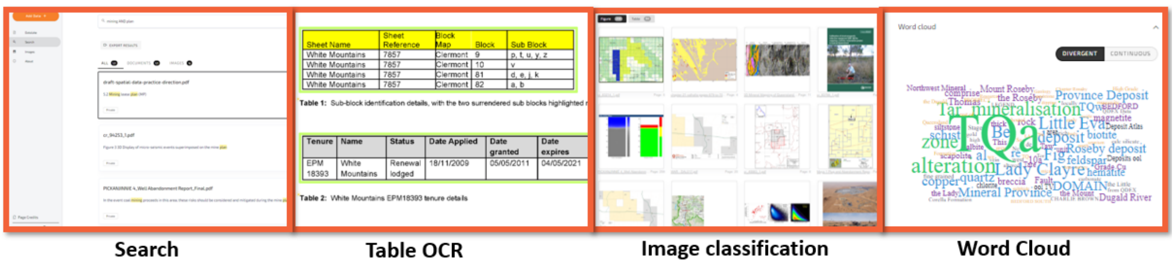

Figure 1 *Transversal corpus search features of the digital data factory.*

Seamless Search Tool

To have a firm grasp of the copper resource information covering the definitions, inferences, indications, and compiled measurements, geologists and mining professionals involved in the exploration stage of the copper mining projects would need to be able to gain new insights and search through their unstructured data (OceanaGold Corporation, 2022). The ingestion and digestion process makes it possible to obtain new knowledge and insights, which is very difficult to achieve from the original unstructured data. The ingested data is run through a Machine Learning-based pipeline that transforms the unstructured data into structured data with its elements made searchable (Mamador et al., 2020). With the seamless search, geological, mineral and deposit information can be found efficiently and fresh insights into the site mineralogy can be gained with relative ease (Maver et al., 2021).

Copper mineral deposit models can be correlated to their appropriate locations on geological maps and supported by visual evidence such as mineralogy descriptions in drill hole cores and geological maps as displayed in Figure 2. Solid geological inferences can be made regarding the characteristics of the copper ore deposit, which can then lead to feasible drilling and productive mining plans supported by owned data.

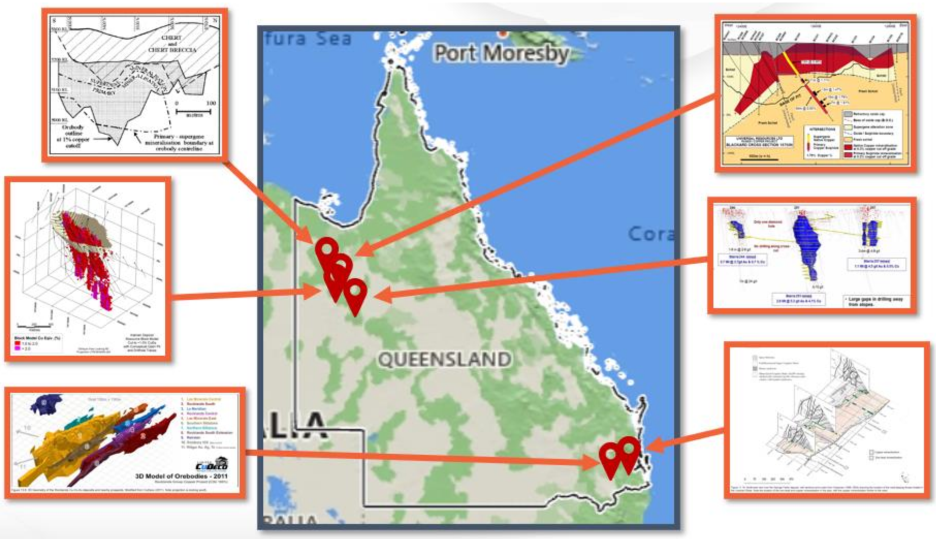

Figure 2 *Findings from deep corpus search of copper resource models and new geological insights across Queensland.*

File, Domain and Image Tagging

File, domain, and image tagging are performed through the data factory. This allows for the consolidation of unstructured data, breaking data silos across documents and disciplines, and making all the relevant data during the copper mining operation accessible across organizations and contractors, hence streamlining mining workflows and supporting cross-company collaboration (Maver et al., 2021).

An ongoing copper mining operation would continuously produce various figurative and imagery data such as resources and machinery management, schedules, rainfall, land survey information, engineering solutions, geology, and more. Some, if not all of these will be integrated or considered in the mining plan or model (OceanaGold Corporation, 2022).

The continuous aggregation by domain experts and ingestion of unstructured data that happens during this stage of operation are improved via machine learning processes and scaled suitably. This is particularly useful to track, reassess, visualize, and evaluate the mining plans or mining models at various copper mining stages, such as tracking the progress of specific sub-blocks at a particular time. Information from various sources would be integrated or correlated with these mining plans or models to support decision-making associated with the site as shown in Figure 3. This process facilitates not only efficient and easy problem-solving conditions but supports planning processes and governance structures to be data-based and able to respond to ESG opportunities, risks and challenges (Maver et al., 2021).

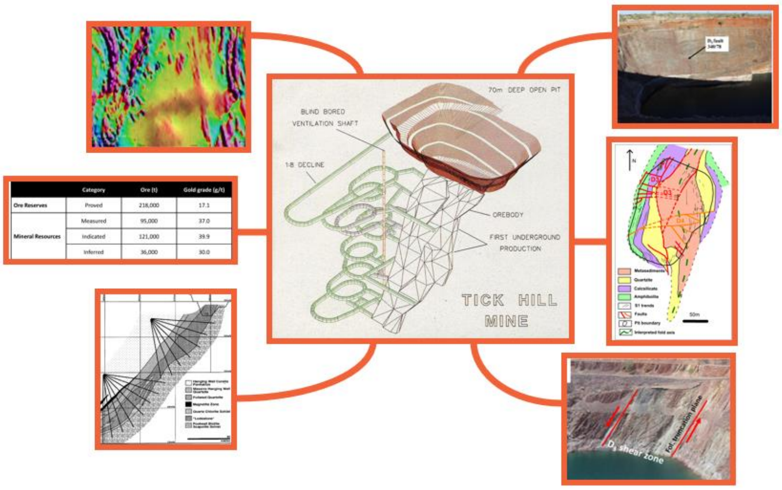

Figure 3 *Important documents and varying information concerning the mining model at different scales are accessible by various roles (mining engineers, operators, surveyors, geologists, etc.) across the organization involved.*

Table extraction

Engineering, geoscience, and even metallurgic processes at the copper processing plant produce a vast wealth of tables and numerical data (OceanaGold Corporation, 2022). Classifying tables to a particular image group is also tracked by the data factory. Optical character recognition (OCR) is used to identify and locate individual and specific information which can be used in further analyses. By automatically converting each table image to a .csv file, valuable information becomes easily available, searchable, and aggregated across various mining operation stages or copper processing plant processes as shown on Figure 4. Manual translation of the table to a .csv file can therefore be avoided, and information tracked to the original location in the report.

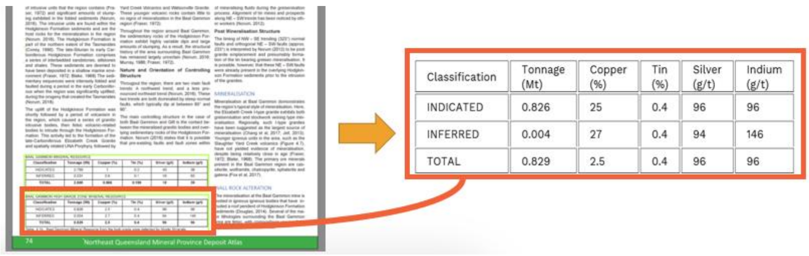

Figure 4 *Table image with numerical values such as tonnage and ore grades identified by the data factory's OCR and extracted for external documents.*

Being able to extract numerical values from tables is valuable to help track mining information such as ore grades, tonnage, production values, coordinates, rainfall, work hours, processing plant or laboratory parameters.

Mapping of Similar Files and Word Cloud

A data factory contains multiple tools that facilitate rapid comprehension and understanding of the context of the data. For example, the searchability of the elements in each document allows transverse documents within the vast database with similar keywords to be grouped together through the search results. The word cloud associated with each document also allows rapid comprehension of the whole document briefly. Finding analogues, similarities, and historical issues is a problem that now has a solution. In the case of safety, health, and environmental control, incidents, lost time injury (LTI), and reports of investigation (RI) can be traced to specific documents or even specific pages in the original reports.

Environmental, Social and Corporate Governance Targets in Mining

The data factory certainly allows the risks and opportunities related to sustainability to be recognized, evaluated and managed under a holistic framework pertaining to environmental, social and governance aspects. The data factory approach is an incentive that can add value and align the mining operation with broader Environmental, Social and Corporate Governance (ESG) goals to limit environment impact. The data-supported holistic ecosystem from this approach is positioned to enhance cost-benefit assessment of the ESG throughout the mining cycle.

Conclusion

Managing unstructured data into structured data embedded with end-to-end ML/AI advanced technology made it possible to explore, analyze and make fast decisions using big data. With this, organizations involved in copper mining projects and more generally in the mining industry are able to process and present new mining information and knowledge from the dataset. Tools within the data factory such as deep search module, word cloud, table extraction, and image identification contribute significantly to providing a comprehensive understanding across the full value chain for the copper deposit models, mining plans, copper treatment processes, rehabilitation plans, and safety, health and environmental trends. Hence, maximizing the capability of unstructured data has proved impactful in terms of significantly reducing the consumption of research time and costs (Maver et al., 2021) and align mining operations with global ESG limiting the impact on the environment.

Acknowledgment

This paper utilizes the data from the GSQ Open Data Portal database (geoscience.data.qld.gov.au). This database is owned by the Queensland Government and is open for public access. However, the interpretation and conclusion contained in this report are those of the authors alone.

References

Bonakdarpour, M. & Bailey, T.M. (2022) The future of copper – Will the looming supply gap short-circuit the energy transition? S&P Global. *IHS Markit*.

Mamador, C., Aranda, J. O., Arif, N. E., Hernandez, N. M., & Baillard, F. (2020, September). A geological regional case study for pressure, temperature, and salinity for the GoM using machine learning technology on unstructured data. In *EAGE/AAPG digital subsurface for Asia Pacific Conference* (Vol. 2020, No. 1, pp. 1-4). European Association of Geoscientists & Engineers.

Maver, K. G., Baillard, F., & Hernandez, N. M. (2021, May). Accelerating E&P Decisions by Applying Artificial Intelligence and Big Data Analytics to Unstructured Data. In *Digital Subsurface Conference in Latin America* (Vol. 2021, No. 1, pp. 1-5). European Association of Geoscientists & Engineers.

Maver, K. G., Hernandez, N. M., Baillard, F., & Cooper, R. (2020). Processing of unstructured geoscience and engineering information for instant access and extraction of new knowledge. *First Break*, *38*(6), 59-64.

OceanaGold Corporation (2022). The mining process. *OceanaGold*. Retrieved from https://oceanagold.com/operation/macraes/the-mining-process/