

SEG 2022 Conference: Minerals For Our Future

Using AI to Extract Information from Legacy Documents

Amit Juneja, Henri Blondelle, Oriane Nguyen-Thuyet
Agile Data Decisions, Houston, TX, USA

For ore deposit exploration and modeling, the data stored in legacy documents is a key piece of information to make decisions on basins, claims, or leases.

Up to now, extracting this information from documents has been a very labor-intensive task that has rarely been done exhaustively on large amounts of documents. As a result, some key decisions are not taken using all the available data.

The development of artificial intelligence, machine learning, and natural language processing has introduced some new possibilities to get value from legacy documents at scale. Using two case studies we have developed for two major mining companies, we will present different types of methods and algorithms used to capture information in documents like:

- The document classification according to the document text and graphical content
- The capture of collar data such as the Drill Hole name, the Coordinates, the TD, Dip, Bearing, etc.
- The capture of values in assay tables and certificates
- The capture of figures in documents, like maps, logs, cross-sections or diagrams
- The detection of specific geological intervals in composite logs
- The detection of positive recommendations or negative conclusions in reports

The presentation will focus on the role of subject matter experts while building information detection models with machine learning or deep learning tools. The methodologies to measure the accuracy of the models will be detailed with examples to expose what can be achieved with the current technology to get trustable and actionable data.