MIT 805: Big Data
Assignment Part 1: Collection and Process

Kelly Langa
Student Number: u13402103

20 September 2021

The mining industry is rich with historical data. Metallurgical plants have decades of data which are generated every second from many process units and stored in databases. The chosen dataset is from a Metallurgical gold recovery plant. The dataset is composed of operational data, reagent consumptions and metal accounting information. The dataset was collected in order to enable data driven plant monitoring, optimization, design and process simulation. The collection of metallurgical plant data comes with many opportunities, including; real-time intelligent response to process changes, process and equipment conditions monitoring, reduction in consumption of reagents, integrated planning, and safety and environmental issues. Many metallurgical plant decision makers believe that the vest amount of data available in their industry is the key to transforming their plants into highly automated processes (Ghorbani, 2020). However, to date metallurgical plant data is still underutilised.

The plant processes ore material sourced from four different shafts and surface rock dump material. The process flow is composed of crushing, milling, thickening, leaching, carbon-in-pulp (CIP), elution, electrowinning and smelting. The dataset is a batch dataset, which was collected on a daily basis over a period of 6 years from 1 January 2016 to 5 September 2021. The plant operates 24 hours and 7 days a week. There are three 8-hour shifts per day. The daily metallurgical assay data represents a time-weighted average from three samples, which are collected at the end of each shift. The operational data and reagent consumption are totalised numbers or daily averages where applicable. Flowmeter, belt weightometer, densitometer, probes, manual titrations, sample cutters, fire assay and atomic adsorption spectrometer are some of the equipment used in measuring the collected data.

The volume of data is comprised of 62 076 records with 2 076 rows and 41 columns. In terms of velocity, the data was summarized to daily averages and daily-totalised numbers for throughputs. The operational data collected by plant monitoring sensors generates a reading every 10-50 seconds, in streams. The data is stored is on a Data Historian from the SCADA in real-time. However, these readings were converted to a daily figure. The metallurgical assay data is calculated using a daily weighted average from three daily samples collected during each shift. Therefore, the data is collected at a velocity of one unit per day of each record. In terms of volume the data is also summarized to a daily figure.

The data in the dataset is comprised of different metallurgical data, namely; operational data, metallurgical accounting data and reagent consumptions. In terms of operational data, the following data was collected; milled and treated tonnage, mill running time, mill kilowatts, availability, utilisation and number of elutions. In terms of metal accounting samples, the following data was collected; residues (solids and solution), recovery, head grade, eluted carbon, carbon activity, above 150micron particles in leach feed and carbon calcium content. The reagent consumption data is composed of the following plant reagents; cyanide, lime, caustic, paraffin, hydrochloric acid, flocculent, steel balls, portable water and compressed air. The dataset is composed of only quantitative data with text headlines and units, and the date, which the data was collected. The dataset is composed of structured data and is stored in a CSV file format. The figures are both machine generated and human generated through manual sample analysis and are static data.

A huge gap exists in the mining industry in terms of standardise data governance (Vererka, 2020). This results in reduced trust and dependency on data analytics and data driven information. The metal accounting samples are collected three times in a day, and sent out for analysis the next day to an external laboratory. Therefore, there is latency in terms of analysing the samples and reporting the results. With this latency comes discrepancy in the samples as certain side chemical reactions are still occurring. The dataset does contain missing entries, unknown and inconsistent number of significant figures. The missing entries is mainly due to the plant or process unit being offline due to maintenance and breakdowns. Due to sampling inconsistencies and data quality issues, the dataset has unknown accuracy and inferred precision. In terms of weightometer, calibrations are conducted weekly and quarterly with the OEM, the same applies to flowmeters and densitometers. However, most people in the plant have access to the dataset, as the data is not stored in protected files and therefore the dataset could have been tempered with.

From the dataset collected, it is expected that there will be a direct proportional relationship between treated tonnage and reagent consumptions. The more the material processed, the more the reagents will be required to treat the processed material. However, a saturation level is reached in the circuit where the addition of more reagents does not result in a better output product. The recovery factor will be influenced by the leach feed particle sizes. The coarser the particle size the more the target metal is unliberated and therefore the lower the recovery. The milling rate is a function of mill utilisation. The higher the utilisation the higher the milling rate since the mill will be fully utilised to its capacity. The dissolved gold losses is inversely proportional to the carbon activity. The lower the carbon activity, the higher the solution losses will be.

References

Ghorbani, Y., Nwaila, G.T., Zhang, S.E., Hay, M.P., Bam, L.C. and Guntoro, P.I., 2020. Repurposing legacy metallurgical data Part I: A move toward dry laboratories and data bank. Minerals Engineering, 159, p.106646.

Verevka, T., Mirolyubov, A. and Makio, J., 2020, October. Opportunities and Barriers to Using Big Data Technologies in the Metallurgical Industry. In International Scientific Conference on Innovations in Digital Economy (pp. 86-102). Springer, Cham.