Hydrometallurgy 181 (2018) 113–122

Contents lists available at ScienceDirect

Hydrometallurgy

journal homepage: www.elsevier.com/locate/hydromet

Decision support system for bioleaching processes

Cecilia Demergasso[a,b,*], Roberto Véliz[a], Pedro Galleguillos[a,b], Sabrina Marín[a], Mauricio Acosta[a], Víctor Zepeda[a], Jaqueline Zeballos[d], Felipe Henríquez[a], Rodrigo Pizarro[a], Juan Bekios-Calfa[c]

[a] *Centro de Biotecnología, Universidad Católica del Norte, Antofagasta, Chile*
[b] *Centro Científico y Tecnológico para la Minería, Antofagasta, Chile*
[c] *Department of Computing & Systems Engineering, Universidad Católica del Norte, Chile*
[d] *APTIC-Management and Engineering, Antofagasta, Chile*

ARTICLE INFO

Keywords:
Copper bioleaching
Machine learning
Decision support system
Expert system

ABSTRACT

The use of information available in the organizations to understand what good performance looks like has been proposed for improving the decreased productivity in mine sector. Detailed monitoring has been performed at the heap bioleaching process in Minera Escondida since the start of the industrial operation in 2006. The huge industrial data recorded represents an opportunity to raise knowledge about complex bioleaching processes for improving the technology. A systematic approach using machine learning tools for the analysis of High Dimensional Feature Space is now being developed to deliver experience-based learning with the aim to serve as the foundation for optimal production planning and operational decision making, in the presence of inherent process variations. The construction of a Decision Support System (DSS) is reported, which considers a Real Time PCR array, a database for data logging and storage, the application of suitable statistical and computational tools for knowledge acquiring and finally the creation of a system of knowledge translation to transform it into action by applying recommendations that come to terms with operational limitations. The user can accurately retrieve data and design similar matches to the historic operation to get, for instance the expected metallurgical performance (such as copper recovery, acid consumption and bacterial activity) and recommendations. The process followed to construct the base of knowledge of the DSS is discussed.

1. Introduction

The application scenario of copper bioleaching is changing; this is due to the variable copper price, the new competitive technologies, the more complex raw materials and the corresponding process requirements (low grade, run of mine, size of the heap, water/solution quality, among others) (Fundación Chile, 2016). In the current scenario of copper bioleaching, the industrial processes have to deal with growing uncertainties.

In addition, a steady decline in the mining sector productivity during the past decade has been reported (Mitchell et al., 2014) and the described features for the global sector have also been evidenced in bioleaching operations. The main reasons recognized in the specific sector of bioleaching are associated to: i) the resources, the sharp decrease in the copper grade of the exploited ores (from average 1.7 to 0.5% in the last 25 years in Chile) and the increasing complexity of the exploitable ores (> 60% of chalcopyrite); ii) the operation, a high turnover of workers means a reduction of the knowledge residence time in the organization and of the operational experience to best optimize resources by controlling the key factors.

The bioleaching process at Escondida Mine has been in operation since 2000 at pilot scale and since 2006 at commercial scale, and is a good example of the observed trends. Escondida Mine is located 170 km South-East of Antofagasta at 3100 m above sea level. The bioleaching heap was built using run-of-mine (ROM) ore and air is supplied through blowers. The ore was initially (year 2006) characterized as low-grade containing approximately 0.60% (*w*/w) total Cu, consisting of chalcocite (40%), covellite (10%) and chalcopyrite (50%). During the last two years, the grade of the ore allocated for the bioheap process has decreased down to 0.5–0.3% (w/w) total Cu and it is mostly constituted by chalcopyrite (reaching up to 80% of total copper content). The heap dimensions are 2000 m wide by 5000 m long, and it is divided into operational units called strips, each of 125 m wide by 2000 m long and 18 m height (Fig. 1). The operational design considers up to 7 levels of 18 m for each strip and, when the heap is operating with more than one level, the irrigation solution passes through the different levels until reaching the bottom of the heap (Pregnant Leach Solution -PLS- sampling points). The heap process operates using one irrigation solution

* Corresponding author at: Centro de Biotecnología, Universidad Católica del Norte, Antofagasta, Chile.
E-mail address: cdemerga@ucn.cl (C. Demergasso).

https://doi.org/10.1016/j.hydromet.2018.08.009
Received 23 March 2018; Received in revised form 6 August 2018; Accepted 19 August 2018
Available online 08 September 2018

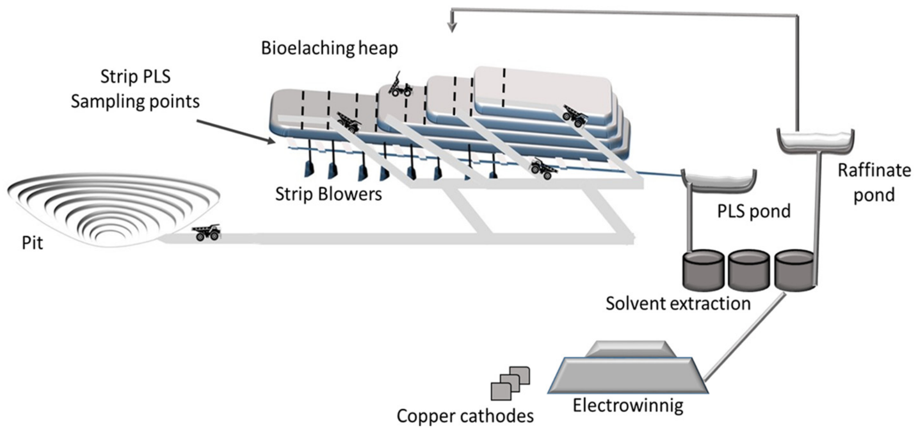


Fig. 1. Flowchart of the copper bioleaching process at Minera Escondida.

(raffinate) directed from the solvent-extraction plant, usually feeding between 12 and 18 ore strips at a steady-state regarding to stacking and irrigation.

Some key solutions have been proposed to face the productivity challenges in the mining sector. Among them are the development of new technologies and the use of information and data available in the organization to understand what good performance looks like. The greatest copper heap bioleaching processes are operating in Chile and there is a big opportunity of obtaining insights from the industrial data (Fundación Chile, 2016) (Mitchell et al., 2014). In this context, the process for creation, systematization, transfer and management of knowledge is relevant to facilitate efficient operation.

Decision Support System (DSS), which consists of a database, model base and knowledge base, is visualized as an integrated approach to bring this understanding from the industrial data for improving the efficiency of mining processes. (Reyes et al., 2014; Tejeda et al., 2013; Zhang et al., 2011).

Mathematical models have been developed to determine the impact of key parameters to facilitate the heap design and control (Watling, 2006). Critical assessments of the significant, public-domain, heap leach models has concluded that most of the efforts have focused in developing predictive models (Watling, 2006) addressed on specific subprocesses in isolation (the chemistry, the microbiology or the hydrodynamics, among others) and have failed to account for the interactions between those processes (Watling, 2006). In addition, the models do not consider the complexity of the current operations. According to our knowledge, models constructed based on statistical industrial data and empirical knowledge of the commercial heap-bioleaching process have not been reported.

There is intense research in knowledge discovery defined as the non-trivial extraction of implicit, previously unknown, and potentially useful information from data (Small & Medsker, 2014). A recent review of approaches for Information Extraction (IE) (Small & Medsker, 2014) highlighted a group of current automatic tools for knowledge extraction and representation (Machine learning and Data mining). Those techniques are based on the concept of automatic learning from data and they use different kinds of algorithms which match the three types of IE work: 1) Unsupervised learning, the model is constructed from a group of *input data* without any kind of feedback (clustering: identifying which one of a small number of cluster centroids best represents the input), 2) Supervised learning, the model counts on *input data* and *output data* that are useful to build a function, and it learns a parametric map that can directly compute a representation for new points $f(in) = out$ and 3) Reinforcement learning, the model learns from a series of enforcements and penalties (Russell & Norvig, 2003).

Information is the raw material for building mathematical models based on IE techniques. Usually data is described by the volume, the veracity, the variety and the velocity. In bioleaching operations there is an important volume of data obtained directly from the mining plant. The data describe the different processes involved in the Cu extraction. Depending on each process, the information is acquired at different frequency and stored for several years. Then, the available data are characterized because of their high dimensionality (the curse of dimensionality) what have a negative impact on the model learning (Small & Medsker, 2014). Considering those characteristics, the use of Big Data techniques -which couples information extraction methods and large amounts, variety and velocity of the data- were incorporated. In such cases data preprocessing, through representation, transformation and projections is used in order to obtain better results.

Data mining methods have been already applied to get knowledge from the industrial data produced in Escondida Mine (Demergasso et al., 2011b; Demergasso et al., 2011a; Soto et al., 2009; Soto et al., 2013). Some of the knowledge inferred from the industrial data has been validated by lab scale tests (Davis-Belmar et al., 2012).

The construction of the knowledge base of a dynamic DSS by the application of IE tools and the gathered expert knowledge of bioleaching is reported.

2. Material and methods

2.1. Process description and data availability

A summary of the bioleaching process at Escondida Mine is shown in Fig. 1. Actually, strips in the heap are operated on levels 1, 2, 3 and 4.

Daily measurements of physico-chemical parameters (pH, Eh, temperature), of solution chemistry (Cu, Fe, acidity, sulfate, ammonium, among others), of operational parameters (solution flow, volume in ponds and reservoirs of solutions like PLS and raffinate, addition of acid and activity of blowers, among others) were recorded. The information about the initial mineralogy of the resources loaded in each strip of the heap (91 strips) was provided by chemical and SEM-based quantitative mineralogical analysis (QEMSCAN). The data gathered included percentages of total Cu (TCu), total Fe (TFe) and soluble Cu (SC), percentages of Cu bearing ore content, chalcocite (CSP-Cs), covellite (CSP−Cv), chalcopyrite (CSP-Cpy) and Cu oxide ores (CSP-ox, mainly brochantite) and percentage of pyrite (Py). In addition, the data included the content of each mineralization type (M1, M2 and M3) in the material stacked in each strip. The mineralization types were defined taking into account the lithology, the alteration and the acid consumption. M1 enclosed mineralization in porphyries and breccias with quartz–sericitic alteration; M2 came from a body of andesites with chlorite, sericite and clay alteration, and it had the highest acid consumption; M3 included mineralization in andesites, porphyries and breccias with quartz, sericitic and potassic alteration. Samples of PLS (from each ore operating strip), common PLS, and raffinate have been monthly sampled since June 2006 to study the composition of the microbial community by Quantitative Polymerase Chain Reaction (Q-PCR) and the kinetics of ferrous iron consumption in flask tests at different temperatures (25 °C and 40 °C) (Demergasso et al., 2011b; Galleguillos et al., 2008; Remonsellez et al., 2009). In addition, since 2014, RNA samples have been collected to determine genetic expression of the microbial population (Galleguillos et al., 2013; Marín et al., 2017). The recorded data were systematically stored in a database. For long time, empirical knowledge of the process has been drawn up as reports and presentations which were registered. Those documents have been manually computed by experts in the fields of heap construction, inoculation, acid management, heat and fluxes management, solvent extraction and related processes, according to the professional knowledge and scheduling experience they have accumulated over several years.

The architecture defined for the DSS consists of three layers (Fig. 2). Each layer was implemented using the programming language Java[1] and is composed of a specific set of libraries and frameworks that are an important part of the system's integration:

- Human-Computer Interface. This layer is responsible for interacting with administrators, users and analysts (experts) of the system through a graphical interface on a Web server. The selected server was WildFly 11.0.0[2] and the implementation was done in Java language. In addition, an MVC (Model, view and controller) Framework (Bucanek, 2009) based on"Ruby on Rail" 5.1.4[3] and Ruby 2.4.1[4] language was used.
- **Engine Inference.** This layer has the responsibility of receiving user queries and delivering logical responses from a set of rules obtained from the analysis of data and expert knowledge by means of an inference engine. For this, the Java programming language was used for the implementation of the modules, the Drools Framework 7.5[5] (Workbench with KIE Execution Server) and the programming language DRL for the definition of the rules of the expert system.
- **Persistence Layer.** The backup of all the industrial data of the bioleaching heap at Escondida mine is installed in a database Postgres 9.6.[6] This layer has the responsibility of storing and supplying the data generated in the plant as well as storing the results obtained from the inferences of the DSS. One advantage of the software is its capability of saving and updating knowledge accumulation.

The DSS is composed of five reasoning modules: 1) Impact of mineralogy/mineralization in the metallurgical performance, 2) Microbial activity and copper recovery, 3) Estimated temperature inside the heap, 4) Marker genes for critical issues and 5) Inoculation/reinoculation requirements (Fig. 2).

Here, the construction of two of the reasoning modules of the DSS will be described. Researchers, IT enterprises, experts and engineers involved in the decision making process have participated in this development.

2.2. Construction of DSS modules by applying CRISP-DM

For constructing the rules and the knowledge base, expert knowledge and knowledge discovered in industrial databases were applied. CRISP-DM was the methodology, among the ones availables, (CRISP-DM (Chapman et al., 2000) and SEMMA (Inc., S. I, 2013)) for IE from the database (Small & Medsker, 2014) oriented to the building of a DSS (Azevedo & Santos, 2008; Shafique & Qaiser, 2014). The CRISP-DM model defines a clear separation between the requirements stablished by the client, by the expert users (Business Knowledge) and by the personnel associated with the analysis of the data (Knowledge of the data). In addition, this methodology has good documentation support and is constantly updated with new revisions and extensions (IBM, n.d.). This methodology has also been used in the knowledge acquisition process in other mining processes (Azevedo & Santos, 2008; Shafique & Qaiser, 2014).

The main conclusion of the phase of **business requirements**, established by the Crisp-DM methodology, for" Impact of the mineralogy/mineralization and the Microbial activity and copper recovery" modules was that the objective of both is to improve the efficiency of the copper recovery by heap bioleaching (Fig. 3).

In the phase of **data comprehension**, the data required for the construction of each module were identified. For the building of module 1" Impact of mineralogy/Mineralization in the metallurgical performance", the global summary of the mineralogy/mineralization composition of each strips given by MEL was used (Fig. 2). The summary collected the information of 91 strips that have been operated since 2006. More than 50,000 records obtained, also since 2006, which includes daily, biweekly and monthly obtained data systematized and stored in a database, were used for the building of the module" Microbial activity based on oxidation" (Fig. 3).

Preprocessing of the data considers ensuring correctness and reducing redundancy and extraneous data. Records with anomalous values were analyzed and discarded (e.g. variables numbers outside the expected ranges, null data, data with text instead of numeric data, among others) in agreement with the research team.

Fig. 3 summarizes the approach used for IE which includes unsupervised learning approaches in a first step (Cluster) and supervised learning methods (Decision tree learning) in a second step. A proper selection of the variables to be used in the analyses is relevant in the next phase for improving the efficiency of the implemented classifiers and for rule generation mainly when high dimensionality data, such as the obtained from the bioleaching heap, are available (Davis & Foo, 2016). Correlated variables were discarded and the selection of variable based on expert knowledge was performed. In addition, the data were separated based on mineralogical/mineralization characteristics for the IE process related to module 2. To select the best predictor variables per group, the wrapper method (Phuong et al., 2005) plus a brute force search (Robinson & Quinn, 2018) was used.

Data mining and machine learning techniques were applied in the **Modelling** phase (Fig. 3). The algorithm for unsupervised learning K-means was used, including the mineralogical and mineralization variables as input ones, to find the patterns for strip classification. In addition, to construct predictive models to describe the oxidation kinetics for each mineralogical group, the data of each group were used to train decision trees (supervised learning technique). In the experiments, the *CRT* or *CART* (Classification and Regression Tree) (Robinson & Quinn, 2018) algorithms were used. At the end of these stages, it is possible to describe the behavior of each mineralogical group (K-means) and the behavior of the target variables (decision trees).

The rules for each module were constructed, tested and validated based on the obtained knowledge about the factors that best explain the variability. The recommendations to tune up those factors come to terms with operational possibilities at industrial level.

3. Results

3.1. Impact of mineralogy/mineralization in the metallurgical performance

In order to get insights from the industrial data, the following strategy was designed: i) clustering of the strips based on the mineralogical and chemical composition (chalcocite, covellite, chalcopyrite, pyrite, soluble copper, copper oxide, Fe content, mineralization types);

[1] https://www.java.com
[2] http://wildfly.org/
[3] http://rubyonrails.org/
[4] https://www.ruby-lang.org/en/
[5] https://www.drools.org
[6] https://www.postgresql.org/

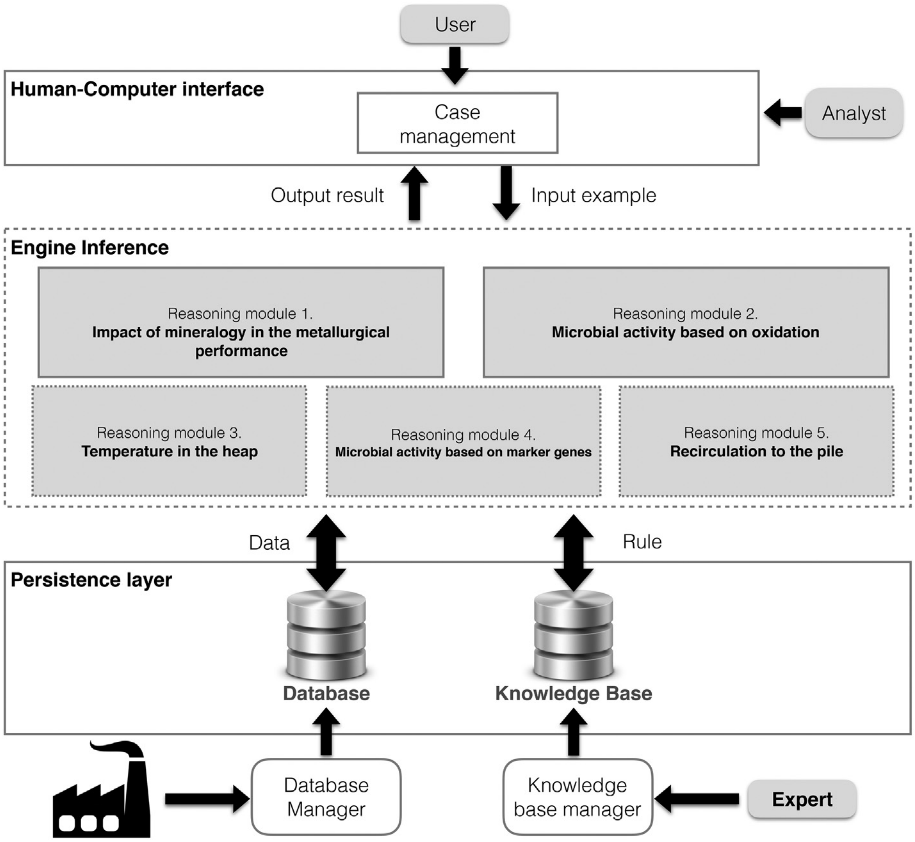


Fig. 2. Architecture of the actual version of the DSS for bioleaching operations. The data base collected the mineralogical composition of each strip (ore and mineralization type composition), the metallurgical parameters (acid consumption, copper recovery, iron recovery), the operational conditions (blowers and flow rate among others), the physicochemical factors (pH, Eh, total Fe and Fe II ion concentrations, Cu concentration, impurity levels, entrained organic solvent) and microbiological data (specific cell number determined by Q-PCR, iron oxidation activity, expression of marker and housekeeping genes). The knowledge base collected the decision rules constructed based on statistical data and empirical knowledge of the process.

ii) exploring the coincidences between the mineralogical groups and the profiles of the metallurgical performance of the strips allocated in each cluster (copper recovery, acid consumption, flow rate, PLS temperature, Eh, total Fe in the PLS); iii) once the groups were improved and validated by the occurrence of those relationships from the point of view of the experts, the rules for strip classification were constructed; iv) the logic of this module was developed to be able to execute the rules once the user inputs the case.

3.1.1. *Strips clustering based on mineralogy/mineralization*

Non-hierarchical clustering classified the leaching strips into five groups (G1-G5) considering the mineralogical/mineralization characteristics of the material stacked. The main variables selected by the clustering tool to form the groups were SC, CSP-ox and M2 content (Fig. 4). Similar clusters were obtained replacing the SC by Py variable (data not shown). Cluster 1 and 2 are the ones with the lowest ($< 3.4\%$) and highest ($> 19.2\%$) M2 content, respectively (Fig. 4 and Supplementary Fig. 1). Group 3 is characterized by its high content of SC and oxide copper ores (CSP-ox ≥ 0.18). Groups 4 and 5 have similar M2 content ($3{,}4 < \text{M2} < 19{,}2$) but differ in SC content (> 0.21 and $< 0.125\%$, respectively). In addition, G5 is the cluster with the highest Py content ($> 2.05\%$, data not shown).

3.1.2. *Correlation between clusters based on mineralogy/mineralization and their metallurgical performance*

To validate the classification obtained, the metallurgical performance of the clusters were analyzed (Cu recovery, acid consumption, temperature, total Fe, irrigation flow, among others). The data obtained from the strip operations were organized into the different groups and the profiles of the metallurgical parameters of each group were analyzed and compared. Features such as the effect of the slope when the strips were located in the border of the heap and significant differences in the irrigation flow produce distortions in the behavior of some strips compared to the cluster behavior. Those strips (21) were removed for further analyses.

The highest copper recovery was obtained in G1 and G2, and the recovery increased at upper lifts. The lowest copper recovery was observed in G3 the one with the highest content of SC (Fig. 5). The highest temperature was observed in G5 and G2, the groups with the highest Py and M2 content (Soto et al., 2013), respectively (data not shown). The highest Fe supply came mainly from G1 and G5, the groups with the lowest M2 and the highest Py content, respectively. However this tendency is observed when those strips were stacked at the first or second lifts of the heap. The decrease in the iron supply in the upper lifts is supposed to be caused by the precipitation processes and the solution retention processes (Demergasso et al., 2010) occurring at the depth profile of the heap (high volume of rocks and low free acid available).

The lowest acid consumption was observed in strips from the G4 with the highest M1 ($\geq 25\%$) content. The M2 strips consume more acid (Net acid consumption = Total acid consumption – acid recovered in solvent extraction process, SX) when those were stacked in the second lift where the acid strength in the irrigation was higher (Supplementary Fig. 2), resembling the observed for silicate gangue minerals (Ghorbani et al., 2016). It is known that the increased gangue dissolution results in increased solution pH and ionic strength, in decreased iron availability and potentially in heap permeability (Watling, 2006). That kind of information was made explicit as an expert recommendations -which consider the process design- that were implemented in the operation of the DSS.

From the microbiological point of view the comparison of the bacterial activity in the obtained groups also shows differences. An interesting relation between Cu recovery (Fig. 5) and the oxidation activity (Table 1) of the different clusters was observed, the higher the oxidation activity the higher the Cu recovery (Table 1, Fig. 5). The lower Cu recovery from G5 compared to G1, in spite of the similar oxidation activity, can be attributed to the different content of CSP-Cpy (G1 and G5 median values 0.10 and 0.20%, respectively) in both groups. The lack of iron oxidation at environmental temperature in the strips of

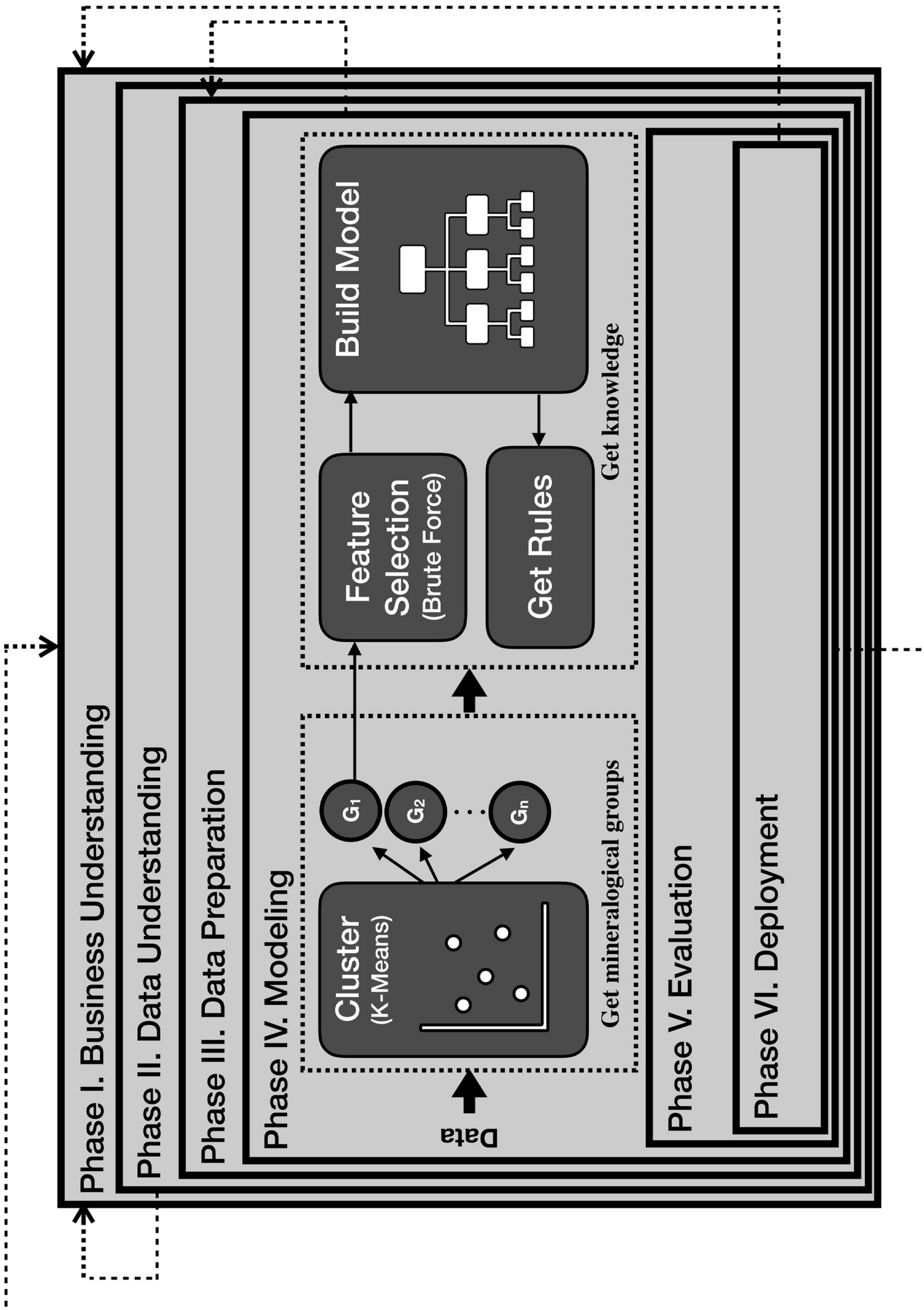


Fig. 3. CRISP-DM consists in six phases: (I) Business Understanding, (II) Data Understanding, (III) Data Preparation, (IV) Modelling, (V) Evaluation y (VI) Deployment.

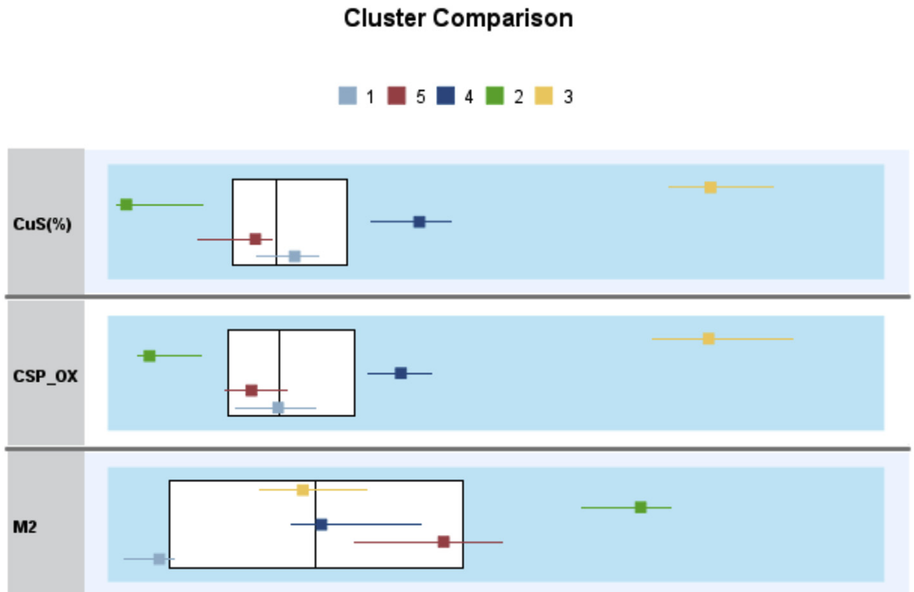


Fig. 4. Grid-style layout, with features selected to make up the clusters in the rows and selected clusters in the columns. This view helps to see differences between clusters not only as compared with the overall data, but also with each other.

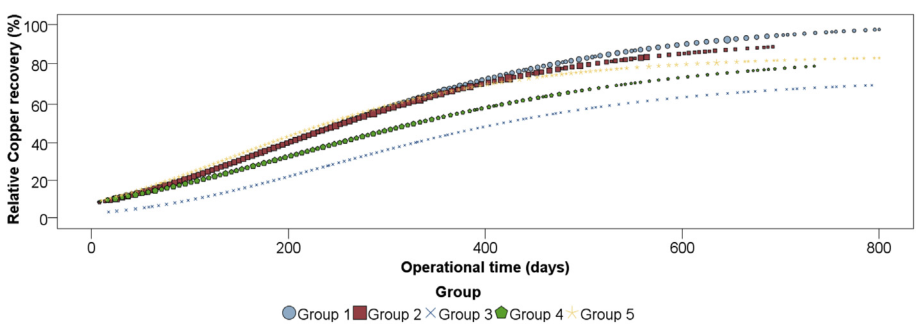


Fig. 5. Smoothed fit line computed by means of locally weighted iterative robust least squares regression (Gompertz equation) obtained for each group. This method computes a series of regressions, each focused on a small area within the plot, and produces a series of local regression lines that are then joined to create a smooth curve.

group 3 was attributed to the low sulfide content (pyrite and chalcocite) and to the high content of oxide ores (Demergasso et al., 2017).

3.1.3. Construction of rules/recommendations/operation

Rules for strip classification. The clustering analysis, the assessment of the metallurgical performance and the expert knowledge allowed defining the rules to be set in the knowledge base of the DSS (Fig. 6, Algorithm 1).

Boundaries. In addition, the boundaries were allowed to be determined by the information gathered in the industrial process. Some of the provided variables are:

$\overline{CSP-ox}$: upper boundary of CSP-ox suitable to be stacked in a strip (percent of the copper bearing ores in the heap).
$\overline{H}$: upper boundary of acid strength for irrigating G2 strips.
$\underline{Py}$, $\overline{Py}$: Lower and upper boundaries of pyrite content for a strip to be considered as a source of iron for the system (percent of pyrite content in the heap) together with M2 content.
$\overline{M2}$: upper boundary of M2 for a strip to be considered as a good source of iron for the system (percent of M2 mineralization content in the heap) together with the pyrite content.

Recommendations. Based on the classification rules and the boundary definitions, recommendations for the decision making process were constructed, like:

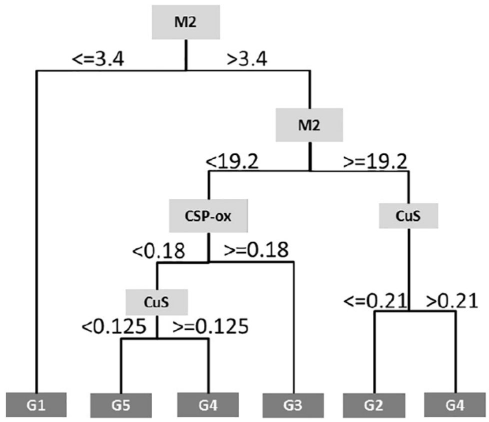
The oxide (CSP-ox) ore content suitable to be stacked in the heap is less that 0.2%. Material with higher CSP-ox must not be stacked in the heap or should be mixed in a proper proportion with materials of lower CSP-ox content.

The acid concentration in the irrigation solution (raffinate) for the heap should have a strength lower than 7.5 g L^{-1} to irrigate the strips of G2 (with predominant M2 mineralization), if it is possible, in order to avoid unnecessary acid consumption and sulfate accumulation in the industrial solutions. The same boundary had been established during

Table 1
Percent of the kinetics oxidation types in each of the mineralogical groups (G1-G5).

T: 25 °C	Groups					T: 40 °C	Groups				
Type[a]	1	2	3	4	5	Type[a]	1	2	3	4	5
Fast	37%	24%	0%	11%	33%	Fast	66%	72%	61%	52%	76%
Normal	53%	62%	38%	52%	58%	Normal	23%	23%	13%	10%	17%
Non- Oxidation	11%	14%	62%	37%	9%	Non- Oxidation	11%	5%	26%	38%	7%

[a] The oxidation tests were classified as Fast, Normal, Non-oxidation when it takes up to 10, 20 and more than 30 days to consume 4 g/L of Fe II, respectively.



ALGORITHM 1: Get mineralogical group from mineralogy/mineralization

```
Data: x = {M2, CPS−ox, CuS}
Result: Return Mineralogical Group
        (G_i, i = {1, 2, 3, 4, 5})
if M2 ≤ 3.4 then
  | return G_1
end
if M2 > 3.4 then
  | if M2 < 19.2 then
  |   | if CSP−ox < 0.18 then
  |   |   | if CuS < 0.125 then
  |   |   |   | return G_4
  |   |   | end
  |   |   | if CuS ≥ 0.125 then
  |   |   |   | return G_3
  |   |   | end
  |   | end
  |   | if CSP−ox ≥ 0.18 then
  |   |   | return Cluster 5
  |   | end
  | end
  | if M2 ≥ 19.2 then
  |   | if CuS ≤ 19.2 then
  |   |   | return G_2
  |   | end
  |   | if CuS > 19.2 then
  |   |   | return G_3
  |   | end
  | end
end
```

Fig. 6. Classification tree model showing the predicted mineralogy/mineralization types of strips.

the design of the plant.

The strips labeled as the best to supply iron (Py ≥ 1.5%, M2 ≤ 15%) are relevant to maintain its level in the system. Those strips are recommended to be stacked in the first two lift of the heap to avoid iron precipitation and trapping at the deeper zones of the heap (see Supplementary Fig. 3).

Operations. The user can, for example, manually input the mineralogical parameters selected of a strip under irrigation to obtain the group in which it is associated and to be able to compare the physicochemical parameters, the metallurgical performance and the bacterial activity of the present strip with other strips of the same group. In addition, the user can also input the mineralogical parameters of future strips, regarding mine exploitation schedule, and the DSS chooses the appropriate model to predict its behavior. The computing results can be stored in a database, which can serve as a historical decision making scheme. A knowledge base management system provides storage, query, modification, adding, deleting and other operations for the knowledge. The expert judges how reasonable the computing results are for the different cases and the updating necessity.

3.2. Microbial oxidation activity of the clusters based on mineralogy/mineralization

Differences in microbial oxidation activity were found among strips of mineralogical groups (Table 1). The groups 1 and 5 included an elevated proportion of fast oxidation tests at 25 and 40 °C (37 and 33%, respectively, Table 1), while group 3 included most of the tests classified as non-oxidation at 25 °C (62%) and group 4 at 40 °C (38%). Previously, low microbial activity has been reported when oxide ore content was increased in the heap (Demergasso et al., 2017). In this regard, group 3 contained the highest level of copper oxide ore and also showed the highest percentage of non-oxidation tests (Table 1).

The highest number of fast oxidation tests, together with the highest copper recovery (75%) was observed in group 1. Group 2 showed the second highest copper recovery and included the highest percentage of normal oxidation tests. Group 5 also contained a high percentage of fast oxidation tests and its recovery profile was similar to groups 1 and 2 during most of the operation time (400 days).

In spite of the similar proportion of fast oxidation type tests in strips classified into mineralogical groups 1 and 5, a different distribution of this type tests was observed in those groups during the progress of the process. In group 1, most of the fast type oxidation tests occurred during the first stage before operation day 100, (Fig. 7a), while in group 5, the majority of fast oxidation tests occurred after operation day 100 (Fig. 7b).

In order to extract new knowledge regarding oxidation type test recorded at 25 °C for mineralogical groups 1 and 5, CRT algorithm was used. The main parameters influencing the oxidation test type in group 1 (highest copper recovery) were copper recovery and the concentrations of total mesophiles and *Acidithiobacillus ferrooxidans*. The Fig. 8 (Algorithm 2) shows that non-oxidation tests recorded at 25 °C in strips belonging to group 1 were completely separated depending on the percentage of copper recovery, all fast oxidation tests occurred when copper recovery was higher than 23%. Similarly, most of normal oxidation tests were classified depending on the percentage of copper recovery (72%, when copper recovery is lower than 23%) and then the separation between non-oxidation and normal oxidation tests is determined by the concentration of *Acidithiobacillus ferrooxidans* (10^4 cells/mL). Confusion matrix built for the Algorithm 2 showed between 87 and 100 assertiveness percent (Supplementary Table 1).

In this regard, the distribution of oxidation type tests depending on *Acidithiobacillus ferrooxidans* concentration is shown in Fig. 9, where most of normal oxidation occurred at cell concentrations between 10^3 and 10^6 cells/mL which contrast with the observed in group 5 (See Fig. 10).

In group 5, acid concentration, total Fe in irrigation solution and the cells concentration of *L. ferriphilum* were relevant parameters to classify most of normal oxidation tests performed at 25 °C. The decision tree for oxidation test type was constructed with data of group 5 (86 oxidation tests, Supplementary Fig. 4). Oxidation tests recorded at acid levels higher than 2.55 g/L (54 out of 86) were twice separated by total Fe levels in the irrigation solution (nodes 3,4,7 and 8) reaching 19 normal oxidation tests at node 8. Then most of the tests of node 8 were recorded when *L. ferriphilum* was higher than magnitude order 6 (2.0×10^6 cells/mL), finally sorting out 30% of normal oxidation tests. Normal oxidation tests recorded at acid levels higher than 2.55 g/L (26 out of 50) were separated by total Fe in the irrigation solution (lower than 1.5 g/L), to finally sort out 48% of normal oxidation test (Supplementary Fig. 5, Algorithm 3). Interestingly, more than 70% of the instances recorded in node 2 (Supplementary Table 2) registered *L. ferriphilum* concentration lower than 10^6 cells/mL.

Therefore *L. ferriphilum* concentration combined with acid level and total iron in the irrigation solution can be used to predict approximately 80% of the normal oxidation tests. The content of pyrite in strips of group 5 was higher than in groups 1 and 2. The higher content of pyrite can influence the Fe availability for the microbial community. Differences in Fe availability in bioleaching processes can determine the predominant species in the process (Boon et al., 1998; Demergasso et al., 2010; Rawlings et al., 1999). Thus, the higher pyrite content in strips of group 5 could have favored the fact that *L. ferrpihilum*, instead

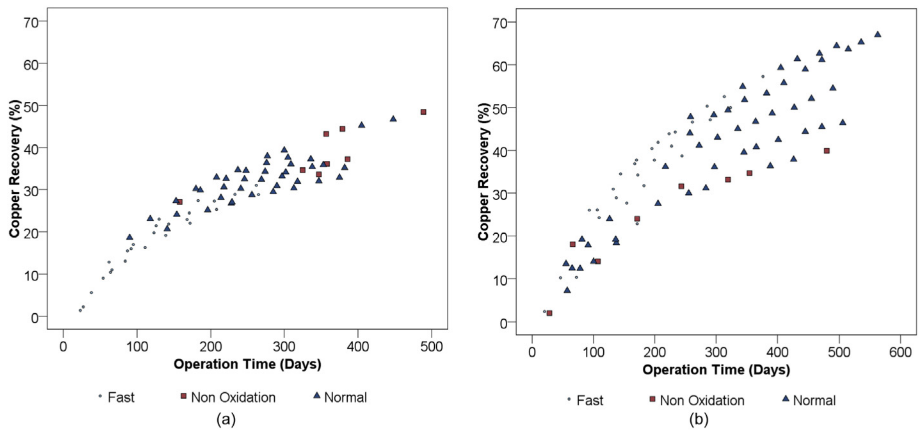
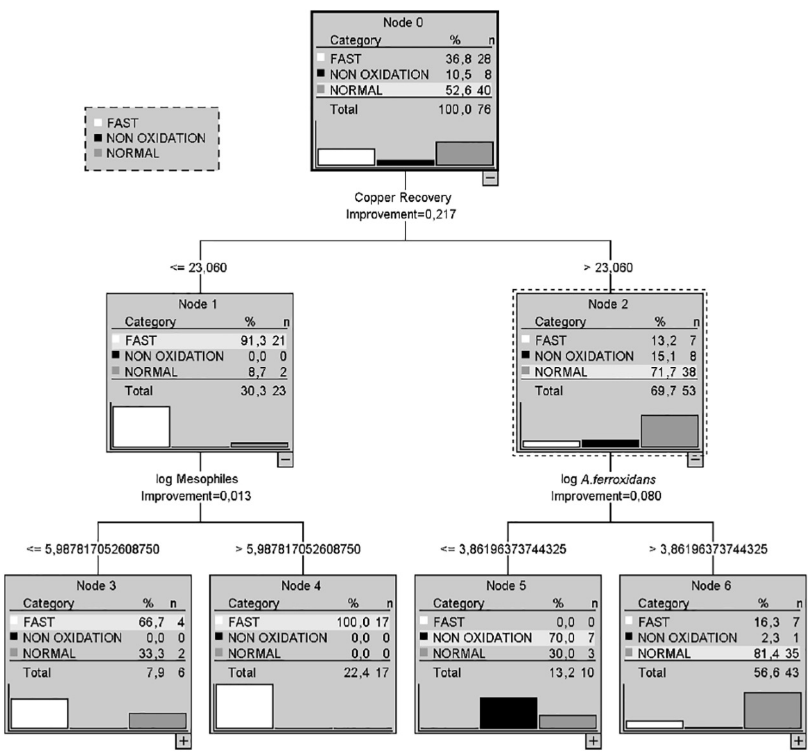


Fig. 7. Distribution of iron oxidation type tests (25 °C) in mineralogical groups 1 (a) and 5 (b).



ALGORITHM 2: Get NORMAL oxidation test (25 °C) in group 1.

```
Data: x = {Copper recovery, log
      A.ferroxidans}
Result: Return Get oxidation test
        type
if Copper recovery ≤ 23.060 then
  | return FAST
end
if Copper recovery > 23.060 then
  | if log Mesophiles ≤ 5.98 then
  |   | return NON OXIDATION
  | end
  | if log Mesophiles ≤ 5.98 then
  |   | return NORMAL
  | end
end
```

Fig. 8. Decision tree generated for oxidation test types (25 °C), based on data from mineralogical group 1. At the first level, copper recovery percentage classifies most of fast tests. At the second level, concentrations of *A. ferrooxidans* and mesophiles are able to classify most of normal and non-oxidation tests.

of *A. ferrooxidans*, was a more important microbiological parameter in order to predict the oxidation type tests.

3.2.1. Construction of rules

The rules were inferred from CRT Analyses in order to predict the microbial oxidation activity in strips classified in mineralogical groups 1 and 5 (with elevated copper recovery), and the main factors to determine the occurrence of fast oxidation tests in groups 1 and 5 are cell concentrations of *A. ferrooxidans* and *L. ferriphilum*, respectively.

3.2.2. Boundaries

In addition, the boundaries were allowed to be determined by the information gathered in the industrial process. Some of the provided variables are: Atf lower boundary for fast oxidation activity in G1. Lf lower boundary for fast oxidation activity in G5.

3.2.3. Recommendations

The recommended *A. ferrooxidans* concentrations in PLS for getting a fast Fe II oxidation is higher than 10^4 cells/mL at the starting phase of the bioleaching process (up to 100 operation days) in strips belonging to groups 1. *Re*-inoculation can improve the cell density when it is lower that this boundary. The advantage of forced inoculation has been proved in bioleaching tests (Tupikina et al., 2014). In the MEL bioleaching system it was proved that the time to get high Eh and a maximum rate of Cu recovery was decreased after forced inoculation (Supplementary Fig. 6). The recommended *L. ferriphilum* concentration in PLS for getting a fast Fe II oxidation is higher than 10^3 cells/mL in strips belonging to G5. A practical approach to increase microbial concentration in the process is to drive PLS directly to raffinate pond. To support this action, the DSS possesses a module considering several parameters to advice what PLS strip would be more suitable for re-circulating and for how long (module not described in details here).

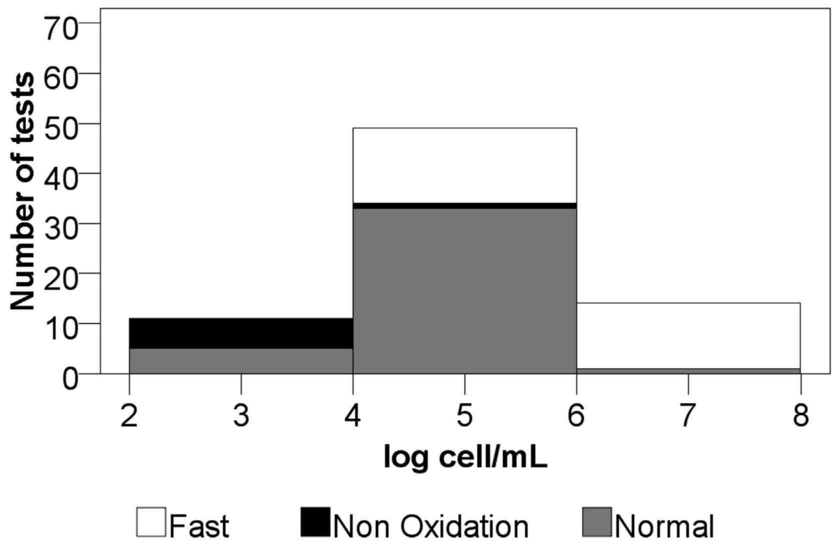


Fig. 9. Distribution of oxidation test types in mineralogical group 1 at different concentration of *A. ferrooxidans*.

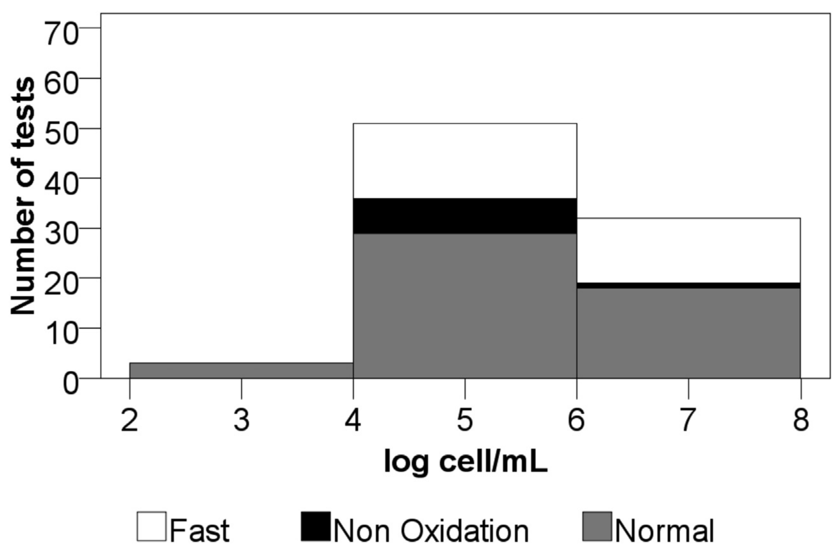


Fig. 10. Distribution of oxidation test types in mineralogical group 5 at different concentration of A. ferrooxidans.

4. Conclusions

Obtaining insights from the industrial data had given the opportunity to better address the research to improve the technology. In this work, machine learning techniques are shown as important supporting tools to explain the raw data obtained directly from complex industrial processes. The results showed that it is possible to obtain rules from the data and that they can be used by experts for the construction of a knowledge base. Also a novel method is defined for improving the efficiency in the generation of rules using unsupervised learning techniques (mineralogical groups), supervised learning (decision trees) and selection of variables to reduce the complexity of the models generated.

The development of a decision making support system offers the frame to transfer the obtained knowledge and transform it into recommendations to assist the plant metallurgist and leaching operator with the control of the heap to maintain it at optimum bioleaching conditions and therefore maximizing leaching kinetics and copper recovery. The DSS developed with its operational recommendations is actually under assessment at industrial level in the bioheap process at Escondida mine. Because of the evolution of the bioleaching process, the DSS must be continuously updated and improved.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.hydromet.2018.08.009.

Acknowledgment

The authors gratefully acknowledge funding from the FONDEF projects IT13I20042 and IT16M10045, and the support of Minera Escondida Ltda.

References

Azevedo, A., Santos, M.F., 2008. KDD, SEMMA and CRISP-DM: a parallel overview. In: In Proceedings of the IADIS European Conference Data Mining.
Boon, M., Heijnen, J.J., Hansford, G.S., 1998. The mechanism and kinetics of bioleaching sulphide minerals. Miner. Process. Extr. Metall. Rev. 19 (1), 107–115.
Bucanek, J., 2009. Model-view-controller pattern. In: Bucanek, J. (Ed.), Learn Objective-C for Java Developers. Apress, Berkeley, CA, pp. 353–402.
Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., 2000. CRISP-DM 1.0 Step-by-Step Data Mining Guide. Technical report. The CRISP-DM consortium.
Davis, J.J., Foo, E., 2016. Automated feature engineering for HTTP tunnel detection. Comput. Security 59, 166–185.
Davis-Belmar, C., Gallardo, I., Demergasso, C., Rautenbach, G., 2012. Effect of organic extractant LIX 84IC, pH and temperature changes on bioleaching microorganisms during SX treatment. Hydrometallurgy 129 (130), 135–139.
Demergasso, C., Galleguillos, F., Soto, P., Serón, M., Iturriaga, V., 2010. Microbial

succession during a heap bioleaching cycle of low grade copper sulfides: does this knowledge mean a real input for industrial process design and control? Hydrometallurgy 104 (3), 382–390.

Demergasso, C., Iturriaga, V., Palacios, F., and Valdebenito, R. (2011a). Effect of SX on microbial populations inhabiting in solutions of a heap bioleaching process. XIX Int. Symp. Solvent Extract. Santiago, Chile.

Demergasso, C., P, S., Zepeda, V., P, G., 2011b. Scientific monitoring of industrial bioleaching process. In: Sobral, L.G.S., Oliveira, D.M.D., de Souza, C.E.G. (Eds.), Biohydrometallurgical Processes: A Practical Approach, pp. 84–94.

Demergasso, C., Véliz, R., Galleguillos, P.A., Marín, S., Acosta, M., Zepeda, V.J., Bekios, J., Zeballos, J., 2017. From knowledge to best practices in bioleaching. Solid State Phenom. 262, 285–289.

Fundación Chile, 2016. From Copper to Innovation: Technological Roadmap 2015–2035. A Impresores, Santiago, Chile.

Galleguillos, P., Remonsellez, F., Galleguillos, F., Guiliani, N., Castillo, D., Demergasso, C., 2008. Identification of differentially expressed genes in an industrial bioleaching heap processing low-grade copper sulphide ore elucidated by RNA arbitrarily primed polymerase chain reaction. Hydrometallurgy 94 (1), 148–154.

Galleguillos, P., Music, V., Acosta, M., Salazar, C., Quatrini, R., Shmaryahu, A., Holmes, D., Velásquez, A., Espoz, C., Pinilla, C., Demergasso, C., 2013. Temporal dynamics of genes involved in metabolic pathways of C and N of L. ferriphilum, in the industrial bioleaching process of Escondida mine, Chile. Adv. Mater. Res. 825, 162–165.

Ghorbani, Y., Franzidis, J.-P., Petersen, J., 2016. Heap leaching technology—Current state, innovations, and future directions: a review. Miner. Process. Extr. Metall. Rev. 37 (2), 73–119.

IBM. IBM Analytics Solutions Uni_ed Method (ASUM). IBM Corporation. New Orchard Road Armonk, NY 10504. Produced in the United States of America. https://www-01.ibm.com/marketing/iwm/iwm/web/pick.do?source=swerpba-basimext&lang=en_US.

Inc., S. I, 2013. Data Mining Using SAS Enterprise Miner: A Case Study Approach. SAS Institute Inc., Cary, NC. http://documentation.sas.com.

Marín, S., Acosta, M., Galleguillos, P., Chibwana, C., Strauss, H., Demergasso, C., 2017. Is the growth of microorganisms limited by carbon availability during chalcopyrite bioleaching? Hydrometallurgy 168, 13–20.

Mitchell, P., Steen, J., Moran, C., Bradbrook, M., Henderson, C., MacAulay, S., Higgins, L., Kastelle, T., and Kunz, N. (2014). Productivity in mining: now comes the hard part. A global survey. EY - Building a better working world. http://www.ey.com/gl/en/industries/mining—metals.

Phuong, T.M., Lin, Z., Altman, R.B., 2005. Choosing SNPs using feature selection. In: 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05), pp. 301–309.

Rawlings, D.E., Tributsch, H., Hansford, G.S., 1999. Reasons why *'Leptospirillum'*-like species rather than *Thiobacillus ferrooxidans* are the dominant iron-oxidizing bacteria in many commercial processes for the biooxidation of pyrite and related ores. Microbiology 145 (1), 5–13.

Remonsellez, F., Galleguillos, F., Moreno-Paz, M., Parro, V., Acosta, M., Demergasso, C., 2009. Dynamic of active microorganisms inhabiting a bioleaching industrial heap of low-grade copper sulfide ore monitored by real-time PCR and oligonucleotide prokaryotic acidophile microarray. Microb. Biotechnol. 2 (6), 613–624.

Reyes, F., Tejeda, G., Herrera, M., Romero, F., Cipriano, A., 2014. Model based decision support system for the heap leaching process. In: In IFAC Proceedings Volumes, pp. 10664–10669.

Robinson, A.C., Quinn, S.D., 2018. A brute force method for spatially-enhanced multivariate facet analysis. Comput. Environ. Urban. Syst. 69, 28–38.

Russell, S.J., Norvig, P., 2003. Artificial Intelligence: A Modern Approach, 2 ed. Pearson Education.

Shafique, U., Qaiser, H., 2014. A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). Int. J. Innov. Sci. Res. 12 (1), 217–222.

Small, S.G., Medsker, L., 2014. Review of information extraction technologies and applications. Neural Comput. & Applic. 25 (3), 533–548.

Soto, P., Meneses, C., Contador, Y., Galleguillos, P.A., Demergasso, C., Serón, M., 2009. Characterization of oxidizing activity of a microbial community in an industrial bioleaching heap. Adv. Mater. Res. 71, 59–62.

Soto, P.E., Galleguillos, P.A., Serón, M.A., Zepeda, V.J., Demergasso, C.S., Pinilla, C., 2013. Parameters influencing the microbial oxidation activity in the industrial bioleaching heap at escondida mine, Chile. Hydrometallurgy 133, 51–57.

Tejeda, G., Reyes, F., Herrera, M., Romero, F., Cipriano, A., 2013. Decision support system for hydrometallurgical processing. In: In IFAC Proceedings Volumes, pp. 243–247.

Tupikina, O.V., Minnaar, S.H., Rautenbach, G.F., Dew, D.W., Harrison, S.T., 2014. Effect of inoculum size on the rates of whole ore colonisation of mesophilic, moderate thermophilic and thermophilic acidophiles. Hydrometallurgy 149, 244–251.

Watling, H., 2006. The bioleaching of sulphide minerals with emphasis on copper sulphides – a review. Hydrometallurgy 84 (1), 81–108.

Zhang, R., Lu, J., Zhang, G., 2011. A knowledge-based multi-role decision support system for ore blending cost optimization of blast furnaces. Eur. J. Oper. Res. 215 (1), 194–203.