

MODELACIÓN PARA MAXIMIZAR RECUPERACIÓN DE AGUA EN MOLIENDA SAG



<https://www.datagrind.com/>



DataGrindCL



DataGrindCL

Lucas Peryra
Jorge Allende
Fernando Guerrero
Gonzalo Herrera



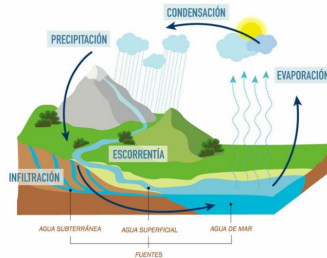
DataGrind

Situación Actual

- ★ El volumen de material que entra al proceso minero puede llegar a ser de miles de toneladas, y por lo mismo, la cantidad de agua necesaria para procesarlos, alcanza una magnitud aún mayor.
- ★ El recurso hídrico al ser escaso en ciertas épocas estacionales junto con el uso ineficaz de ésta en etapas del proceso, puede ocasionar pérdidas sustanciales tanto en procesamiento y ganancias de la minera.
- ★ En la actualidad, las decisiones se basan en controlar el proceso con información de operarios quienes a partir de su experiencia, manejan las variables para obtener un mejor resultado aunque, sin un protocolo definido, no logran asegurar lo eficiente del proceso.

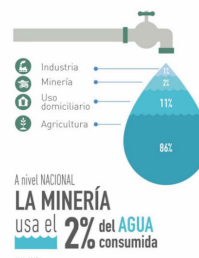
EL AGUA Y LA MINERÍA

CICLO DEL AGUA



El ciclo del agua genera fuentes que son utilizadas por las diversas actividades humanas. La minería formal utiliza algunas de ellas de acuerdo a la ubicación del proyecto minero.

USO DEL AGUA POR SECTORES



A nivel NACIONAL
LA MINERÍA
usa el 2% del AGUA
consumida

Julio 2017

AGUA EN EL PROCESO MINERO

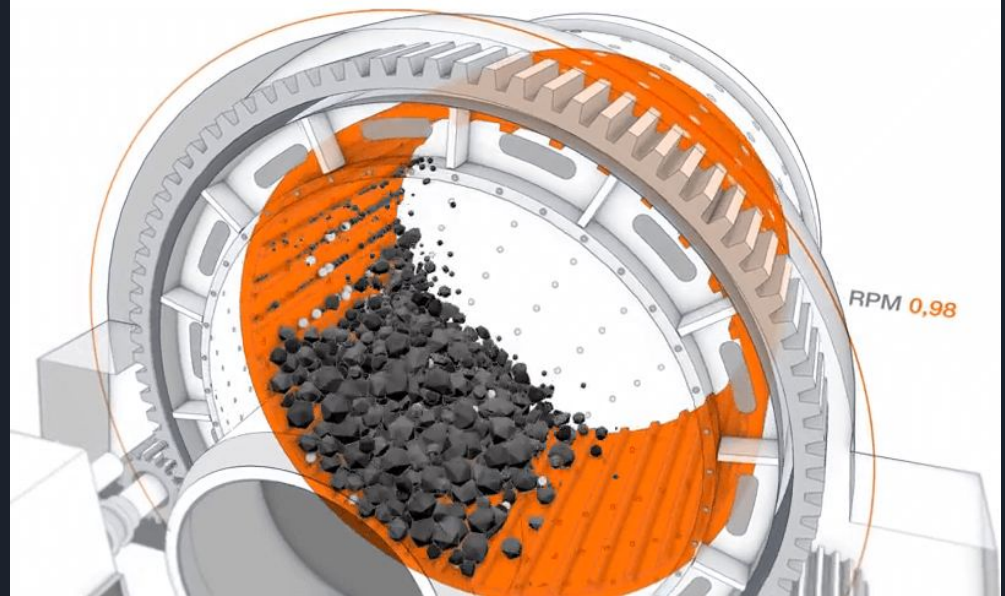


APORTE A LA AGRICULTURA



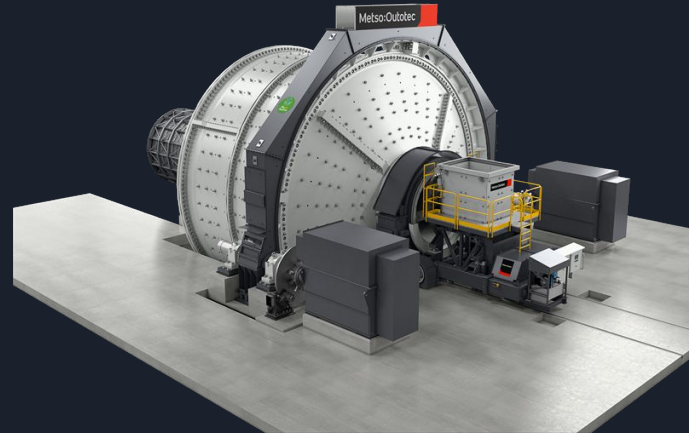
Proceso de reducción con Molino SAG

- ★ Tambor giratorio de gran tamaño (hasta 12 metros de altura) que contiene bolas de acero de varios tamaños (entre 5" a 6,25").
- ★ Ingreso de Materia prima.
- ★ Molienda.
- ★ El transporte.
- ★ Molino SAG5.



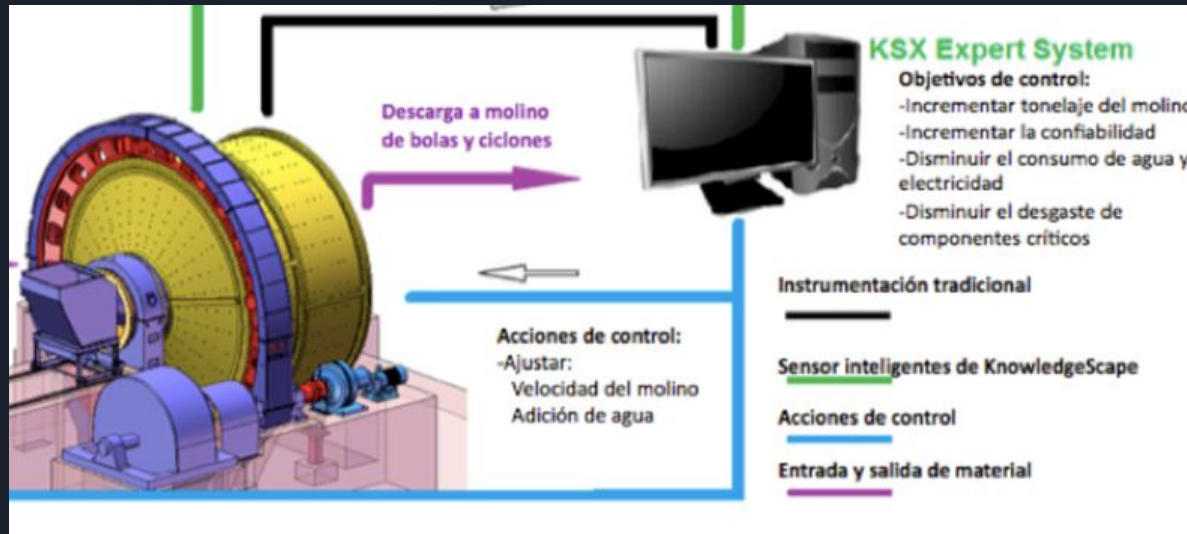
Hipótesis y Objetivo

- ★ Modelar la optimización del uso de los recursos hídricos a partir de un modelo de ML, cuantificando la eficiencia del proceso.
- ★ Modelar la optimización del uso de los recursos hídricos para maximizar la eficiencia de los procesos y establecer las variables más importantes que permitan la elaboración de un procedimiento para el operador.



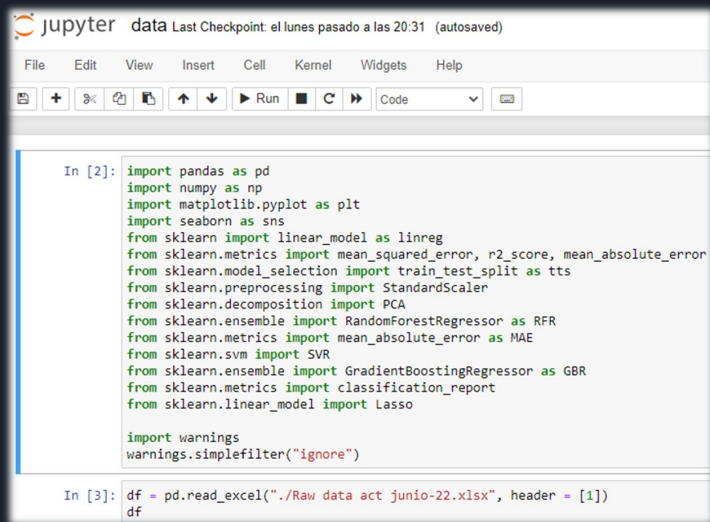
Origen de Datos

- ★ Sensores -> Sistema core transaccional.
- ★ Sistema experto PI con descarga directa de un archivo excel (.xlsx) abordando 3 meses de operaciones.
- ★ Tipos de datos : Punto Flotante.



Ambiente de Trabajo y Preprocesamiento

- ★ Comprende alrededor 20.000 registros, 281 variables continuas.
- ★ Jupyter Notebook, aplicación cliente-servidor dedicada para el análisis de datos (Python).
- ★ Pandas, Numpy, Matplotlib, Seaborn, Sklearn.
- ★ Reducción de variables acotados al propósito del proyecto.
- ★ Variable objetivo : Variable que contabiliza la cantidad de agua utilizada en el proceso. (I)

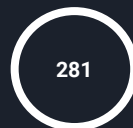


```
jupyter data Last Checkpoint: el lunes pasado a las 20:31 (autosaved)
File Edit View Insert Cell Kernel Widgets Help
+ -> Run Code
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import linear_model as linreg
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.model_selection import train_test_split as tts
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.ensemble import RandomForestRegressor as RFR
from sklearn.metrics import mean_absolute_error as MAE
from sklearn.svm import SVR
from sklearn.ensemble import GradientBoostingRegressor as GBR
from sklearn.metrics import classification_report
from sklearn.linear_model import Lasso

import warnings
warnings.simplefilter("ignore")

In [3]: df = pd.read_excel("../Raw data act junio-22.xlsx", header = [1])
df
```

Reducción de variables



Inicial



Pre



Final

Vector Objetivo / Atributos

Nombre columna	Variable	Unidad
Pot_SAG5	Potencia	[kW]
Pres_SAG5	Presion	[psi]
rpm_SAG5	Velocidad	[RPM]
Cp_SAG5	Porcentaje de solidos	[%]
Tph_SAG5	Alimentacion	[ton/h]
Peb_SAG5	Pebbles	[ton/h]
Peb_Rec_SAG5	Pebbles recirculantes	[ton/h]
Fino_SAG5	Porcentaje de finos	[%]
Intermedio_SAG5	Porcentaje de intermedios	[%]
Grueso_SAG5	Porcentaje de gruesos	[%]
Impactos_Criticos_SAG5	Impactos criticos	[un.]
Impactos_stand_SAG5	Impactos estándar	[un.]
Impac_crit_SAG5_Impfinder	Impactos criticos Impfinder	[un.]
Impac_std_SAG5_Impfinder	Impactos estándar Impfinder	[un.]
Agua_SAG5	Agua	[l.]

El vector objetivo corresponde a la variable Agua_SAG5

Análisis Descriptivo

Data Perdida

- ★ No se encontraron datos perdidos

Medidas de Tendencia Central

- ★ Promedio
- ★ Moda

Medidas de dispersión

- ★ Desviación estándar
- ★ Varianza
- ★ Gráficos de dispersión

Distribución de probabilidad

- ★ Generación de histogramas

Matriz de correlaciones

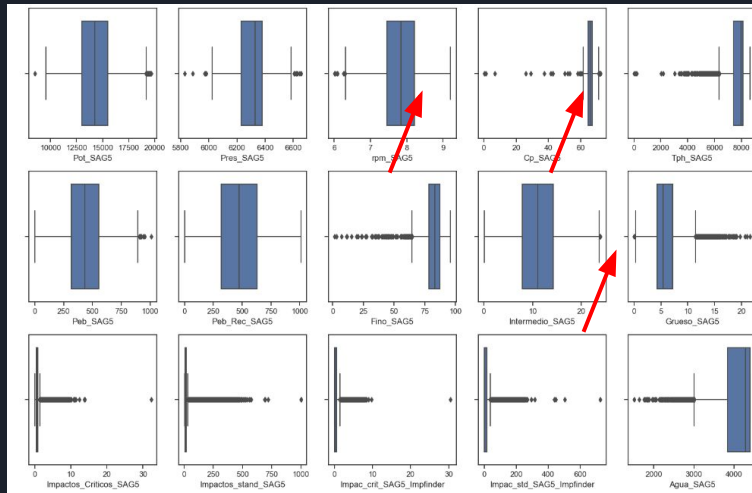
- ★ Generación de Heatmap

	Pot_SAG5	Pres_SAG5	rpm_SAG5	Cp_SAG5	Tph_SAG5	Peb_SAG5	Peb_Rec_SA	Fino_SAG5	Intermedio_Grueso_SAG	Impactos_Cr	Impactos_std	Impac_crit_S	Impac_std_S	Agua_SAG5
count	20016	20016	20016	20016	20016	20016	20016	20016	20016	20016	20016	20016	20016	20016
mean	13603.9881	6291.9836	7.441231	64.933106	7232.42607	408.900449	436.476021	79.791012	10.95583	5.783121	0.810738	25.237982	0.449215	17.945942
std	3616.32795	243.847816	1.787685	7.750451	1876.24077	185.308819	225.557621	13.966453	4.537049	2.87807	1.260305	56.605538	0.936142	40.5768
min	0	13.99	0	0	2.23	0.31	0	0	0	0	0	0	0	0
25%	12723.235	6226.9475	7.4	64.47	7031.275	293.5775	293.3625	77.15	7.48	4.11	0.39	9.52	0	0.94
50%	14083.675	6325.475	7.81	65.66	7947.09	421.335	452.135	83.04	10.73	5.3	0.58	11.45	0.06	4.625
75%	15396.515	6377.7525	8.2	67.1725	8113.4325	544.8075	623.7625	87.23	14.05	7.02	0.8	17.91	0.5	16.52
max	19655.1	6841.78	9.21	99	8655.04	1352.29	1010.67	96.12	23.89	21.6	78.11	1000	30.48	786.32

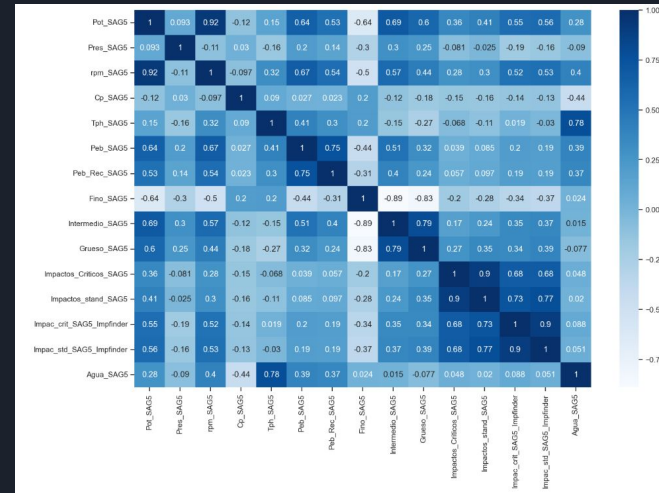
- ★ Se realiza un análisis descriptivo de los datos para extraer totales, promedios, mínimos, máximos, entre otros

Análisis Descriptivo y Preselección de Atributos

- ★ Se eliminan todas las filas donde los valores de velocidad sean menores a 6 [rpm] y el agua menor a 3000 litros, puesto que se considera una parada por mantención del equipo y no es representativo de la operación de molienda



- ★ Se observan bajas en las correlaciones a excepción de Tph, por lo cual se evaluarán modelos con y sin esta variable, de manera de poder comparar sus métricas.



Modelos Predictivos Escogidos y Métricas de Evaluación

1

Regresión Lineal

Este modelo es una primera aproximación del comportamiento de las variables independientes en relación a las variables dependientes y cómo están interrelacionadas.

R²

2

Random Forest

Este modelo otorga una gran robustez en la exploración de diferentes casos

MSE

3

Gradient Boosting Regression

Este modelo permite suavizar los errores a medida que se avanza en su proceso iterativo.

4

Máquinas de Soporte Vectorial

A diferencia de otros modelos basados en árboles de decisión, la interpretabilidad de este modelo es mayor.

RMSE

5

Lasso

Es una técnica de regresión lineal regularizada que ayuda a predecir variables de respuesta al reducir el número de variables predictoras mediante la adición de una penalización en la suma de los valores absolutos de los coeficientes de las variables predictoras.

MAE

Modelos evaluados

Model	MSE	RMSE	MAE	R2
gb_model	1.51E+03	38.87544	27.590168	0.989519
gb_model_lowcorr	3.20E+04	178.999944	122.016876	0.777779
gb_model_lowcorr_std	3.20E+04	179.020756	122.024608	0.777738
gb_model_std	1.51E+03	38.920403	27.595504	0.989495
lasso	1.65E+04	128.261472	93.626045	0.885909
lasso_lowcorr	6.85E+04	261.746569	192.033106	0.524862
lasso_lowcorr_std	6.85E+04	261.746569	192.033106	0.524862
lasso_std	1.64E+04	128.219914	93.664403	0.885983
lr_cn_int	1.65E+04	128.275787	93.610957	0.885884
lr_cn_int_lowcorr	6.85E+04	261.731807	192.036521	0.524916
lr_cn_int_lowcorr_std	6.85E+04	261.731807	192.036521	0.524916
lr_cn_int_std	1.65E+04	128.275787	93.610957	0.885884
lr_sn_int	1.78E+04	133.30008	101.374112	0.876769
lr_sn_int_lowcorr	6.96E+04	263.820266	199.328739	0.517304
lr_sn_int_lowcorr_std	1.69E+07	4112.88719	4099.96889	-116.31436
lr_sn_int_std	1.69E+07	4106.36437	4098.86539	-115.942546
rf_model	5.14E+02	22.67598	12.931299	0.996434
rf_model_lowcorr	1.47E+04	121.414212	62.134866	0.897766
rf_model_lowcorr_std	1.48E+04	121.476672	62.131963	0.897766
rf_model_std	6.75E+05	821.463614	750.599322	-3.679875
svm_model	1.56E+05	395.03636	265.68664	-0.08226
svm_model_lowcorr	1.70E+05	412.167304	278.874398	-0.178161
svm_model_lowcorr_std	8.99E+04	299.796228	197.823432	0.376682
svm_model_std	1.44E+05	379.977061	311.425629	-0.001319

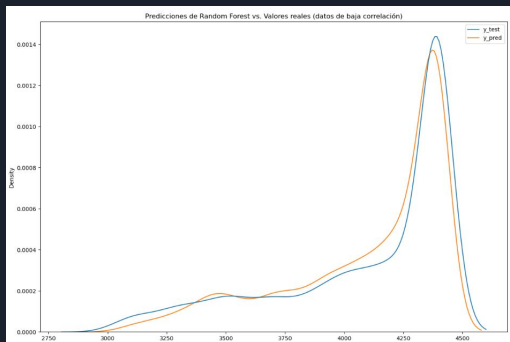
- ★ Se realiza el Train Test Split con un 33% de base Test
- ★ Se trabajan los datos para crear 4 conjuntos de datos a entrenar:
 - datos normales
 - datos escalados
 - datos normales sin las variables correlacionadas
 - datos escalados con las variables correlacionadas
- ★ En función a los resultados obtenidos, se decide conservar los modelos con un R2 entre 0,9 - 0,7.

Modelos con Mejor Resultado

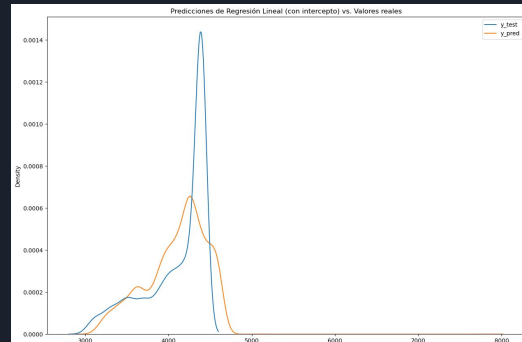
{desafío}
latam_

- ★ Finalmente se seleccionaron 2 modelos para nuestro proyecto. Éstos son "rf_model_lowcorr" y "lr_cn_int".
- ★ En general todos los modelos presentados en la tabla anterior son sumamente parecidos, no obstante decidimos quedarnos con uno de Regresión Logística basándonos en tener un modelo menos costoso al momento de procesarlo, y un modelo de Random Forest ya que es de los que presentaron mejor resultado; también se evitaron modelos que necesiten la estandarización de la data previa, para ahorrar un paso adicional en el procesamiento.

Model	MSE	RMSE	MAE	R2
rf_model_lowcorr	14741.4108	121.414212	62.134866	0.897766
rf_model_lowcorr_std	14756.5819	121.476672	62.131963	0.89766
lasso_std	16440.3465	128.219914	93.664403	0.885983
lasso	16451.0052	128.261472	93.626045	0.885909
lr_cn_int	16454.6776	128.275787	93.610957	0.885884
lr_cn_int_std	16454.6776	128.275787	93.610957	0.885884
lr_sn_int	17768.9113	133.30008	101.374112	0.876769
gb_model_lowcorr	32040.9801	178.999944	122.016876	0.77779
gb_model_lowcorr_std	32048.4312	179.020756	122.024608	0.777738



Random Forest



Regresión Lineal

GUI versión beta

- ★ La interfaz permite al cliente ingresar los valores de entrada con los que desea operar su molino SAG, obteniendo una predicción del agua en litros que deberá añadir al proceso, para lograr dichos resultados.
- ★ Se logra concluir que es factible generar una solución capaz de predecir el agua necesaria para operar un circuito de molienda SAG, introduciendo como variables de entrada, los parámetros operacionales básicos de este tipo de equipos, con los que se desea operar.

The screenshot shows a software window titled 'Predicción d...'. It contains a list of input parameters and their corresponding values, followed by a section for predictions and a 'Predecir' button.

Potencia [kWh]	16450
Presion [psi]	6400
Velocidad [RPM]	8.62
Cp [%]	68.5
Alimentación [t/h]	8199.9
Pebbles [t/h]	581.75
Pebbles Recirculante [t/h]	770.81
Fino [%]	76
Intermedio [%]	17
Grueso [%]	7.2
Impactos Criticos [un.]	0.6
Impactos standar [un.]	11
Impactos criticos Impfinder [un.]	0.4
Impactos sandar Impfinder [un.]	30

Predicciones:

lr_cn_int: 4142.59 litros

rf_model_lowcorr: 4084.53 litros

Predecir



DataGrind

¡Gracias por su atención!

{desafío}
latam_



<https://www.datagrind.com/>



DataGrindCL



DataGrindCL