



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**PREDICCIÓN DE DEMANDA DE BOLAS DE MOLIENDA EN  
CONCENTRADORAS DE MINERA ESCONDIDA MEDIANTE  
ALGORITMOS DE DATA MINING**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

JOSÉ IGNACIO NAVARRO RIVERA

PROFESOR GUÍA:  
IGNACIO ALEJANDRO CALISTO LEIVA

MIEMBROS DE LA COMISIÓN:  
EDGARDO SANTIBAÑEZ VIANI  
MANUEL REYES JARA

SANTIAGO DE CHILE

2020

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL  
POR: JOSÉ IGNACIO NAVARRO RIVERA  
FECHA: 2020  
PROF. GUÍA: IGNACIO ALEJANDRO  
CALISTO LEIVA

## **PREDICCIÓN DE DEMANDA DE BOLAS DE MOLIENDA EN CONCENTRADORAS DE MINERA ESCONDIDA MEDIANTE ALGORITMOS DE DATA MINING**

Minera Escondida se convirtió en 2017 en la primera faena minera que opera tres concentradoras de manera simultánea, logrando procesar más de 400.000 toneladas por día. Uno de los insumos críticos de las concentradoras son las bolas de molienda, las cuales se utilizan en todos los procesos de molienda del mineral. Para el año fiscal 19 se destinó un presupuesto de \$111,5 MUSD en este insumo, pero finalmente se gastó \$93,3 MUSD; lo que representó el 56,1% de la dinámica de gastos en insumos de concentradoras para ese año.

El proceso de carguío de bolas de molienda se realiza con una frecuencia diaria y actualmente el área operacional para realizar la reposición de bolas utiliza la tasa de desgaste de bolas; en gramos por toneladas para las bolas de 5,5" en molinos SAG y en kilowatts por toneladas para las bolas 3" en molinos de bolas. Esta metodología de reposición de bolas no considera otros factores en el consumo de bolas como lo son la mineralogía de alimentación, índices de dureza (SPI, SAG) y elementos menores (As, Mo, Au y Ag), los cuales dan origen a la hipótesis de este trabajo: estudiar cómo estas variables impactan en el consumo de bolas de molienda.

Entonces, el objetivo principal del presente trabajo es: Desarrollar modelos predictivos de consumo de bolas de molienda, para proyectar la demanda futura en Minera Escondida, mediante algoritmos de minería de datos. Para cumplir con los objetivos se desarrollaron 3 modelos de machine learning: Bagging Trees, Random Forest y Gradient Boosting Trees para consumo de bolas 3" y 5,5" en las concentradoras: Laguna Seca 1, Laguna Seca 2 y Los Colorados, resultando ser GBT el que mejor ajusta independiente de la concentradora y molino, de acuerdo a métricas RMSE, Rsquared y MAPE.

Los algoritmos de GBT entrenados para explicar el consumo de bolas generaron modelos que de haber sido utilizados para realizar el forecast del año fiscal 2019, se podría haber alcanzado un accuracy de 95,37% y 95,69 para bolas 3" y 5,5" respectivamente. Esto, habría reducido el presupuesto de este insumo estratégico en \$23.082.978 USD, lo que se traduce en un ahorro efectivo de \$8.735.088 USD.

Finalmente, incluir nuevas tecnologías de analytics y data science permiten agregar valor en una industria que constantemente se enfoca netamente en producción, privilegiando la eficacia sobre la eficiencia. Automatizar procesos de análisis de datos aumenta la eficiencia y permite dedicar menos tiempo en análisis y reportabilidad, enfocando recursos en gestión y toma de decisiones. La gestión del cambio es una etapa fundamental en la utilización de estas nuevas tecnologías.

## **Agradecimientos**

Junto con este trabajo que requirió mucho esfuerzo y dedicación, quiero agradecer a todos aquellos que han sido parte de mi formación como persona y profesional, ya que considero que tienen un gran mérito en el desarrollo de esta memoria.

Primero que todo a mi familia, mi núcleo, que me brindo todo su cariño y apoyo en momentos fáciles y difíciles, preocupándose siempre por mi bienestar y salud.

A mis padres, por ser un pilar fundamental en todo este proceso y brindarme amor, valores y herramientas para llegar donde estoy; que sin su elogiado esfuerzo no hubiese podido acceder a este camino.

A mis hermanos, por su apoyo y compañía para seguir adelante desde y hasta siempre.

A mis amigos de la vida, con los cuales puedo distraerme, conversar, disfrutar, compartir y a pesar del tiempo seguimos creciendo juntos.

A mis amigos de la universidad, porque sin ustedes nada hubiese sido igual, gracias por las horas de estudio, carrete, deporte, paseos; sin duda estos años juntos nos ayudaron a ser mejores personas y profesionales.

A mi casa de estudio, por ser la casa de una gran etapa de mi vida y brindarme un ambiente que fue tanto acogedor como exigente, en el cual pude crecer día a día como persona y llegar a ser profesional.

Finalmente, un especial agradecimiento a BHP y Minera Escondida por darme la gran oportunidad de realizar mi memoria, por recibirme y apoyar mi trabajo, sin ustedes este trabajo no habría sido posible.

# Tabla de Contenido

<b>I.</b>	<b>INTRODUCCIÓN</b> .....	<b>1</b>
<b>II.</b>	<b>DESCRIPCIÓN DEL PROYECTO (O PLANTEAMIENTO DEL PROBLEMA) Y JUSTIFICACIÓN</b> .....	<b>6</b>
<b>III.</b>	<b>OBJETIVOS</b> .....	<b>9</b>
	III.I Objetivo General:.....	9
	III.II Objetivos Específicos: .....	9
	III.III Resultados Esperados.....	9
	III.IV Alcances.....	9
<b>IV.</b>	<b>MARCO TEÓRICO Y METODOLOGÍA</b> .....	<b>11</b>
	IV.I Minería de datos .....	11
	IV.I. I Data Warehousing .....	12
	IV.I. II Modelos Predictivos .....	13
	IV.I.III Validación.....	16
	IV.II CRISP-DM .....	17
<b>V.</b>	<b>DESARROLLO DEL PROYECTO DE MEMORIA</b> .....	<b>21</b>
	V.I Macro Proceso de Producción de Cobre: .....	21
	V.I. I Proceso en Concentradora .....	22
	V.I. II Proceso de carguío de Bolas: .....	25
	V.I. II. Carguío en Molino SAG:.....	25
	V.I. II. II Carguío en Molinos de Bolas:.....	25
	V.II Consolidación de la base de datos: .....	26
	V.III Caracterización de los datos históricos:.....	28
	V.IV Estadística Descriptiva:.....	28
	V.IV Modelo predictivo como solución al problema planteado.....	31
	V.IV. I Selección del algoritmo a utilizar: Challenge de modelos de machine learning.....	32
	V.IV. II Resultados teóricos a nivel operacional.....	35
	V.IV. III Resultados teóricos a nivel comercial .....	38
	V.V Dashboard para herramienta predictiva.....	44
	V.VI Visión Estratégica .....	45
<b>VI.</b>	<b>CONCLUSIÓN</b> .....	<b>47</b>
	VI.I Cumplimiento de objetivos propuestos.....	47
	VI.II Resultados Observados .....	48
	VI.III Recomendaciones a Futuro .....	48
<b>VI.</b>	<b>BIBLIOGRAFÍA</b> .....	<b>50</b>
<b>VII.</b>	<b>ANEXOS</b> .....	<b>51</b>

## Índice de Tablas

Tabla 1: Producción 2018 (TM) y Participación de Mercado.....	1
Tabla 2: Grande Cifras 2017. ....	5
Tabla 3: Accuracies Forecasts Bolas 3” y 5,5” en Minera Escondida.....	7
Tabla 4: Algoritmo de Bagging. ....	15
Tabla 5: Algoritmo de Boosting. ....	16
Tabla 6: Volumen y variables de las bases de datos disponibles.....	26
Tabla 7: Variables de Carguío de bolas L1-L2. ....	26
Tabla 8: Variables de Mineral Alimentado. ....	27
Tabla 9 : Estadística descriptiva Laguna Seca 1.....	29
Tabla 10: Estadística descriptiva Laguna Seca 2.....	30
Tabla 11: Estadística Descriptiva Concentradora Los Colorados.....	31
Tabla 12: Variables independientes a utilizar en modelos predictivos. ....	32
Tabla 13: Resultados Modelos de Machine Learning en Molinos SAG.....	33
Tabla 14: Resultados Modelos de Machine Learning en Molinos de Bolas. ....	33
Tabla 15 : Grilla para buscar hiperparámetros óptimos.....	34
Tabla 16: Hiperparámetros óptimos para modelos GBM. ....	34
Tabla 17: Accuracy bolas 5,5” en Laguna Seca 1.....	38
Tabla 18: Accuracy bolas 5,5” en Laguna Seca 2.....	39
Tabla 19: Accuracy bolas 5,5” en Los Colorados. ....	40
Tabla 20: Accuracy bolas 3” en Laguna Seca 1.....	40
Tabla 21: Accuracy bolas 3” en Laguna Seca 2.....	41
Tabla 22: Accuracy bolas 3” en Los Colorados. ....	42
Tabla 23: Accuracy bolas 5,5” en Minera Escondida. ....	42
Tabla 24: Ahorro de bolas 5,5” en Minera Escondida Limitada.....	42
Tabla 25: Accuracy bolas 3” en Minera Escondida Limitada.....	43
Tabla 26: Ahorro de bolas 5,5” en Minera Escondida Limitada.....	43

## Índice de Ilustraciones

Figura 1: BHP Global. ....	2
Figura 2: Forecast Bolas 5,5” Año Fiscal 19.....	7
Figura 3: Forecast Bolas 3” Año Fiscal 19. ....	7
Figura 4: Etapas del modelo CRISP-DM.....	18
Figura 5: Proceso de Molienda.....	23
Figura 6: Esquema del Data Warehouse a Construir. ....	28
Figura 7: Influencia relativa de variables en LS1.....	35
Figura 8: Influencia relativa de variables en LS2.....	35
Figura 9: Influencia relativa de variables en LC.....	36
Figura 10: Influencia relativa de variables en LS1.....	36
Figura 11: Influencia relativa de variables en LS2.....	37
Figura 12: Influencia relativa de variables en LC.....	37
Figura 13: Forecast para año fiscal 2019 en Concentradora Laguna Seca 1. ....	38
Figura 14: Forecast para año fiscal 2019 en Concentradora Laguna Seca 2. ....	39
Figura 15: Forecast para año fiscal 2019 en Concentradora Los Colorados. ....	39
Figura 16: Forecast para año fiscal 2019 en Concetradora Laguna Seca 1. ....	40
Figura 17: Forecast para año fiscal 2019 en Concetradora Laguna Seca 2. ....	41
Figura 18: Forecast para año fiscal 2019 en Concentradora Los Colorados. ....	41
Figura 19: Forecast para año fiscal 2019 de bolas 5,5” en Minera Escondida.....	42
Figura 20: Forecast para año fiscal 2019 de bolas 3” en Minera Escondida Limitada. ..	43
Figura 21: Dashboard para la visualización de la herramienta predictiva. ....	44
Figura 22: Sección KPI’s FY20 - Dashboard.....	44
Figura 23: Evolución de Analytics. Valor del negocio v/s complejidad. ....	46

# I. INTRODUCCIÓN

## Industria Minera del Cobre en Chile

La minería del cobre ha sido la principal exportación de Chile durante más de un siglo. Desde 1970 la producción nacional del cobre ha aumentado marcadamente debido a varios factores: la nacionalización del cobre, la entrada de mayor inversión extranjera directa y el descubrimiento de varias faenas de clase mundial. Al año 2018 Chile es el principal productor de cobre con un 28,3% del total mundial, seguido por Perú (11,6%), China (7,7%) y Estados Unidos (5,8%). Adicionalmente, las reservas del país equivalen a poco menos del 40% de las reservas del planeta.

De manera directa, el año 2018 la participación de la minería del cobre en el PIB alcanzó al 8,9, y su participación en (US\$) respecto de la exportación total del país a un 48,2%. más del 50% de las exportaciones. Representa también uno de cada tres dólares de inversión extranjera directa recibida en el país, y cerca del 15% de los ingresos fiscales de la última década.

### Principales Actores

Dentro de los principales actores de la gran minería del cobre se encuentran Codelco, BHP, Collahuasi, Antofagasta Minerals y Anglo American, cuya participación de mercado se resume en la siguiente tabla:

Empresa Minera	Producción 2018 (Miles TM)	Porcentaje
Codelco	1.677,50	29%
BHP	1485,3	25%
Collahuasi	559,2	10%
Antofagasta Minerals	785,2	13%
Anglo American	505,7	9%
Otros	818,7	14%
Total	5.831,60	100%

*Tabla 1: Producción 2018 (TM) y Participación de Mercado.*

Fuente: Adaptado de [1].

## BHP Global

BHP es una compañía líder a nivel global en recursos naturales diversificados. Con presencia en 12 países, es uno de los mayores productores de commodities del mundo: incluyen petróleo, mineral de hierro, cobre y carbón, y cotiza en las bolsas de valores tanto de Londres como de Australia.

En 2014 inicia un profundo plan de reestructuración de sus activos, con el objetivo de tener una estructura organizacional más simple, ágil y eficiente. Es así como en marzo

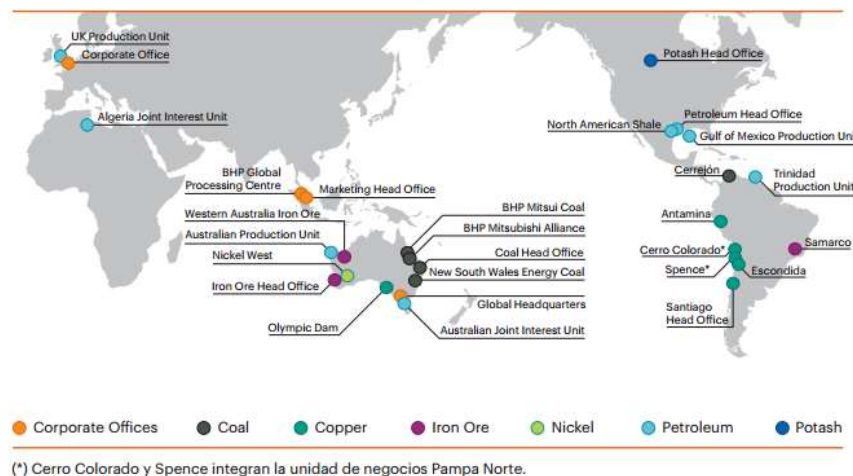
de 2016 anuncia un giro fundamental en su forma de trabajar, que esencialmente consiste en una reagrupación de sus unidades de negocio en tres áreas: Minerals Americas, Minerals Australia y Petroleum.

Junto con reagrupar las operaciones por zonas geográficas, estandariza todos sus procesos a nivel mundial, para luego simplificarlos y globalizar la mayoría de las funciones. A través de este modelo, que implica trabajar como equipo global e integrado, BHP busca concentrar el esfuerzo de sus líderes operacionales en seguridad, cultura, productividad y costos, con el objetivo común de conseguir el mejor resultado posible.

El propósito de BHP es “es crear valor para los accionistas en el largo plazo a través del descubrimiento, adquisición, desarrollo y comercialización de recursos naturales”. Su estrategia es “poseer y operar activos de procesos primarios, de gran tamaño, larga vida, bajo costo y con potencial de crecimiento, diversificados por producto, geografía y mercado.

Los valores de la compañía son:

- Sostenibilidad: Situar a la salud y la seguridad en primer lugar, ser ambientalmente responsables y apoyar a nuestras comunidades.
- Integridad: Hacer lo correcto y cumplir con nuestra palabra.
- Respeto: Valorar la transparencia, la confianza, el trabajo en equipo, la diversidad y las relaciones de beneficio mutuo.
- Desempeño: Alcanzar altos resultados para el negocio, utilizando al máximo nuestras capacidades.
- Simplicidad: Concentrar nuestros esfuerzos en lo más importante.
- Responsabilidad: Definir y aceptar la responsabilidad, y cumplir nuestros compromisos.



**Figura 1: BHP Global.**  
**Fuente: Obtenido de [3].**



## **Minerals Americas:**

Minerals Americas es la unidad de negocios que agrupa las operaciones mineras de BHP localizadas a lo largo del continente americano, cuya sede está en Santiago de Chile. Reúne las faenas de Minera Escondida Ltda. (57,5%) y Pampa Norte (Compañía Minera Cerro Colorado Ltda. y Minera Spence S.A.) en Chile, Potash (Canadá) –integrada el 1 de marzo de 2017–, y los joint ventures no operados Antamina (Perú), Samarco y Fundación Renova (Brasil), Cerrejón (Colombia), Resolution (Estados Unidos) y Nimba (Guinea<sup>1</sup>). Desde el año 2015 el chileno Daniel Malchuk es el presidente de BHP Minerals Americas.

## **Operaciones en Chile:**

BHP llega a Chile en 1984 tras adquirir Utah Corporation, que entonces era el socio principal de Minera Escondida. En el año 2000 Billiton compra Rio Algom, que tenía entre sus activos a Spence y Cerro Colorado. En 2010, estas dos últimas faenas pasan a integrar la unidad de negocios Pampa Norte.

Con estas tres operaciones en su portafolio, BHP se convirtió en el mayor actor privado de la minería en Chile. La actividad minera en Chile contribuye de manera significativa al ingreso nacional, a la recaudación fiscal (US\$ 2.644 millones en 2017), a la generación de oportunidades empresariales y laborales, y al crecimiento del país. El cobre, en particular, tiene un impacto gravitante en la economía nacional.

Chile es el mayor productor de cobre del mundo, con 5.5 millones de toneladas producidas en 2017, equivalentes al 27,2% de la producción mundial. Asimismo, el cobre sigue siendo el principal producto de exportación nacional, representando el 49,1% del valor total exportado en 2017. BHP es el primer actor privado de la industria del cobre en Chile. Sus tres operaciones produjeron 1.190.142 toneladas en este ejercicio, equivalentes al 21,6% del total nacional y cerca del 6% de la producción mundial. Las exportaciones de la compañía sumaron US\$ 6.570,3 millones, representando aproximadamente un 19% del total de exportaciones de cobre del país.

Además de su aporte como principal fuente de divisas para el país, la minería del cobre genera un impacto indirecto en la economía nacional y regional a través de sus diferentes efectos multiplicadores. El aporte a la sociedad se refleja más detalladamente en la distribución del valor agregado<sup>2</sup>, que en 2017 totalizó US\$ 7.816 millones, entre Minera Escondida y Pampa Norte. El pago total a proveedores nacionales fue de US\$ 3.318 millones.

## **Minera Escondida:**

Escondida es una de las operaciones de cobre más importantes de Chile y del mundo. Está situada en el desierto de Atacama, a 170 km al sureste de la ciudad de Antofagasta y a una altura de 3.100 metros sobre el nivel del mar. Produce concentrado y cátodos de

---

<sup>1</sup> Activo ubicado en África que integra la organización de joint ventures no operados dependientes de Minerals Americas.

<sup>2</sup> Valor agregado corresponde al valor económico distribuido que considera, entre otros, costos operacionales, sueldos de trabajadores y beneficios, otros gastos a terceros, impuestos e inversión a la comunidad

cobre a través de la explotación de dos rajos abiertos (Escondida y Escondida Norte). El concentrado de cobre se obtiene a través del proceso de flotación de mineral sulfurado, y los cátodos de cobre, mediante lixiviación de mineral oxidado, biolixiviación de sulfuros de baja ley, extracción por solventes y electro-obtención. La infraestructura actual de la empresa en faena consiste en sistemas de chancado y transporte de mineral, tres plantas concentradoras y dos desalinizadoras, dos pilas de lixiviación, dos plantas de extracción por solventes y una planta de electro-obtención. A ello se suman dos mineroductos que transportan el concentrado hasta sus instalaciones en Puerto Coloso, donde es filtrado y embarcado a los clientes.

Las oficinas corporativas de Minera Escondida están ubicadas en la ciudad de Antofagasta, en el norte del país. Sus propietarios son BHP, empresa operadora, con el 57,5%; Río Tinto (30%), y las japonesas JECO Corporation (10%) y JECO 2 Ltd. (2,5%). Mauro Neves es el Presidente de Minera Escondida desde abril de 2017.

En 2017 la producción de Escondida totalizó 925.414 toneladas, que se desglosa en 687.440 toneladas de cobre contenido en concentrado y 237.974 toneladas de cátodos. Este volumen representó una disminución de 8% en comparación con el año anterior, caída que se explica por la baja en la ley del mineral, la huelga de 44 días que afectó a la operación en febrero y marzo, y a eventos climáticos ocurridos en junio que llevaron a suspender temporalmente las operaciones.

En la segunda mitad del año la producción de Escondida registró un aumento de 29% respecto al mismo periodo de 2016, debido principalmente a la puesta en marcha la extensión de la planta concentradora Los Colorados.

Para resumir las grandes cifras del 2017 se puede ver la siguiente tabla:

Datos Operacionales	Métrica	Pampa Norte	Minera Escondida	Total BHP Chile
Total cobre fino producido	Tmf	264.748	925.414	1.190.142
Participación de producción de cobre en Chile	%	4,8	16,8	21,6
Ventas	MMUS\$	1.689	6.364	8.053
Empleados directos (1)	Personas	2.245	3.578	8.677
Contratistas (2)	FTE (3)	2.428	6.249	8.677
Inversión comunitaria	MMUS\$	3,8	24,2	28

Número de proveedores locales (4)	Proveedores		562
Porcentaje de proveedores locales	%		28,06%

**Tabla 2: Grande Cifras 2017.**  
**Fuente: Obtenido de [3].**

- (1) Incluye personal de oficina de Santiago.
- (2) Incluye valores de contratistas en Santiago.
- (3) La sigla significa Full Time Equivalent.
- (4) Se publica número de proveedores de las tres operaciones a una de las regiones en las que se encuentran las operaciones, para este caso, Tarapacá y Antofagasta.
- (5) Porcentaje calculado sobre el total de proveedores nacionales.

## II. DESCRIPCIÓN DEL PROYECTO (O PLANTEAMIENTO DEL PROBLEMA) Y JUSTIFICACIÓN

Con la puesta en marcha del proyecto de extensión de la planta concentradora Los Colorados, en septiembre de 2017, Minera Escondida se convirtió en la primera faena minera que opera tres concentradoras de manera simultánea (considerando además de las plantas Laguna Seca 1 y 2), logrando una capacidad de procesamiento superior a 400.000 toneladas por día.

Esta es una iniciativa clave para la estrategia de Escondida en el largo plazo, ya que permitirá compensar la caída natural de la ley recuperar los niveles de producción a un promedio de aproximadamente 1,2 millones de toneladas anuales de cobre fino durante la década.

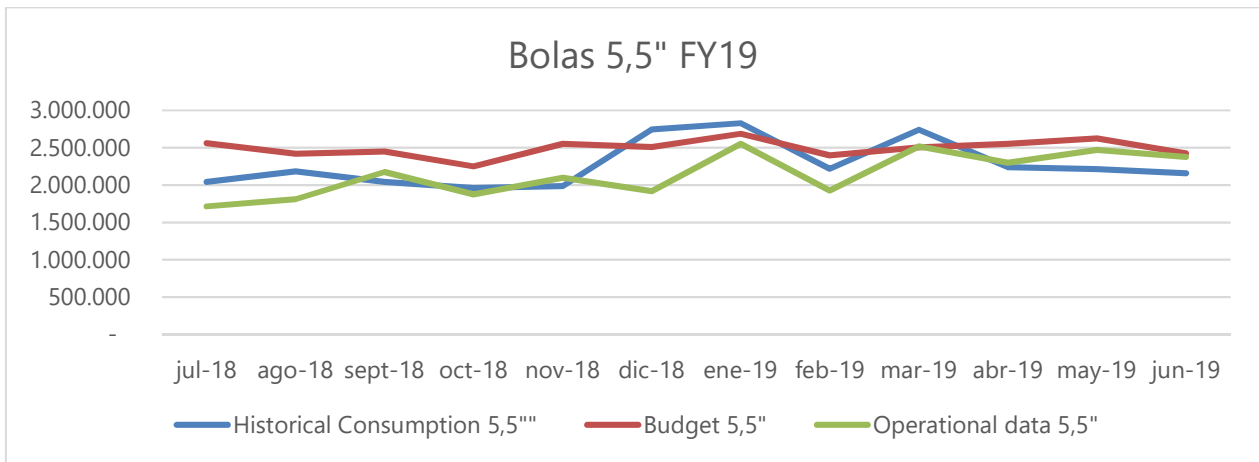
Estas concentradoras utilizan insumos críticos: bolas de molienda, cal, colectores, NaHS; es decir, aquellos cuya situación del abastecimiento es crucial para la operación minera en el mediano y largo plazo. La criticidad está en función del riesgo de desabastecimiento, la alta dependencia de las importaciones, el monopolio de los insumos, como también el peso relativo del insumo en la estructura de costos de la operación.

Uno de los insumos críticos son las bolas de molienda, las cuales se utilizan en todos los procesos de molienda del mineral: convencional, unitaria (con y sin HPGR), SAG y remolienda. Para el año fiscal 19 (FY19), correspondiente entre julio del 2012 y julio del 2019, se utilizaron 57.173 toneladas de bolas 3" y 25.740 toneladas de bolas 5,5", lo que equivale a un gasto de \$62.792.386 USD y \$30.522.814 USD, respectivamente. Es más, este gasto representa el 56,2% de la dinámica de gastos en insumos para el año fiscal.

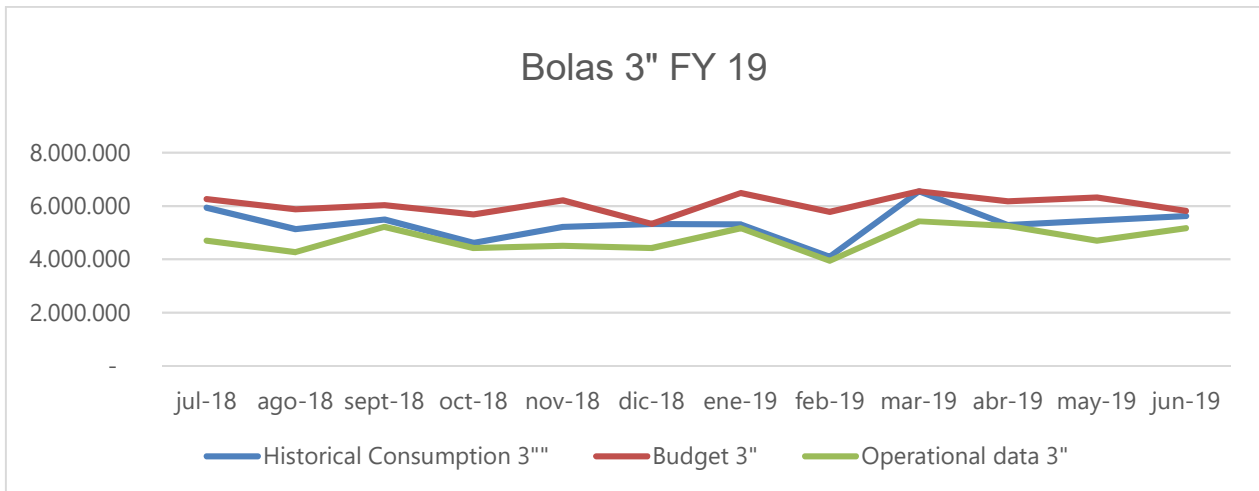
El proceso de carguío de bolas de molienda se realiza con una frecuencia diaria y actualmente la operación de Minera Escondida para realizar la reposición de bolas utiliza la tasa de desgaste de bolas; en gramos por toneladas para las bolas de 5,5" en los molinos SAG y en kilowatts por toneladas para las bolas 3" en los molinos de bolas. Esta metodología de reposición de bolas no considera otros factores en el consumo de bolas como lo son la mineralogía de alimentación, índice de energía SAG (SPI), índice de trabajo de Bond (BWI) y elementos menores (As, Mo, Au y Ag), los cuales pretenden estudiarse en este trabajo de tesis como impactan en el consumo, ya sea en bolas 3" como bolas 5,5".

Por el lado Comercial, el forecast se realiza por años fiscales, los cuales comprenden para el año fiscal X, desde julio de (X-1) hasta junio de X. Este forecast utiliza tasas de desgaste históricas, por lo que el accuracy del forecast no es de los mejores.

A continuación, se presentan los forecast realizados para el año fiscal 19 (FY19), los cuales contemplan el Presupuesto (Budget), lo que se compró (Historical Consumption) y lo que realmente se usó según la data operacional (Operational Data):



**Figura 2: Forecast Bolas 5,5" Año Fiscal 19.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**



**Figura 3: Forecast Bolas 3" Año Fiscal 19.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Luego, los respectivos accuracies son (calculados como  $1-MAPE^3$ ) :

Insumo	Accuracy (1-MAPE)
Budget Bolas 5,5"	81,9%
Historical Consumption 5,5"	87,05%
Budget Bolas 3"	72,39%
Historical Consumption 3"	87,85%

**Tabla 3: Accuracies Forecasts Bolas 3" y 5,5" en Minera Escondida.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Entonces, se observa que la actual predicción de demanda deja una oportunidad de mejora en términos de eficiencia, considerando el volumen del consumo. Por ende, general un modelo de predicción de demanda robusto y confiable para las bolas de molienda combinando la experiencia operacional con analítica avanzada, permitirá facilitar las conversaciones de valor entre la operación y el área comercial alineadas con

<sup>3</sup> Error Porcentual Absoluto Medio

la demanda futura requerida; su cobertura, riesgos y oportunidades comerciales, como, por ejemplo, gestionar proactivamente la compra y entrega de este insumo y también negociar con mayor certeza los volúmenes (y por ende mejores precios) con los proveedores, impactando directamente en la dinámica de gastos de este insumo.

### **III. OBJETIVOS**

#### **III.I Objetivo General:**

“Desarrollar modelos predictivos de consumo de bolas de molienda, para proyectar la demanda futura en Minera Escondida, mediante algoritmos de minería de datos”

#### **III.II Objetivos Específicos:**

Los objetivos específicos que se buscan como parte del estudio son:

- Conocer y estudiar el proceso de molienda para comprender de mejor forma el problema de negocio.
- Realizar análisis exploratorio y descriptivo de los datos para determinar variables que permitan pronosticar el consumo de bolas de molienda.
- Desarrollar y seleccionar modelos de minería de datos apropiados para abordar el problema.
- Utilizar modelos seleccionados para predecir la demanda de un año fiscal, y analizar su desempeño en términos teóricos a nivel operacional y comercial.
- Utilizar una herramienta de visualización como TIBCO Spotfire, para desarrollar un dashboard de uso intuitivo que permita visualizar los resultados de la herramienta predictiva a nivel operacional y comercial.

#### **III.III Resultados Esperados**

El proyecto de tesis busca desarrollar modelos predictivos de consumo de bolas de molienda de 3” y 5,5” para las concentradoras de Minera Escondida Limitada, siendo el entregable los modelos teóricos de minería de datos junto con métricas evaluativas del desempeño de estos, junto con la visualización de los resultados de los modelos en un dashboard mediante el software TIBCO Spotfire.

#### **III.IV Alcances**

Según lo expresado en el capítulo introductorio, en este trabajo se abordará el estudio de la incorporación de nuevas variables para predecir el consumo de bolas 3” y 5,5” en molinos de bolas y SAG respectivamente. Este estudio se enfocará en dos áreas funcionales. A continuación, se presentan los alcances de cada una de estas:

-Área Operacional: corresponde al área de planificación y control de procesos. Esta área es la encargada de la reposición y carguío diario de bolas de molienda. No se cuestionará la validez del carguío actual. En este estudio se trabajará para apuntar a predecir este carguío diario de bolas y entender que variables explican la variabilidad de este proceso.

-Área Comercial: corresponde al área de contratos de concentrado y el área de planta y servicios de procesamiento de Minerals America, las cuales se encargan de comprar y proveer las bolas de molienda a la operación de Minera Escondida. En este estudio se trabajará para apuntar a mejorar la predicción de demanda de bolas de molienda para un año fiscal, mediante algoritmos de minería de datos.

Finalmente, el alcance de esta memoria es hasta la implementación del modelo, entregando resultados teóricos (accuracy, RMSE, Rsquared), junto al ejercicio de proyectar la demanda para el año fiscal 20. No se considera dentro del alcance de esta memoria la gestión del cambio necesaria para la implementación de la herramienta predictiva dentro de las rutinas operacionales ni comerciales.



## IV. MARCO TEÓRICO Y METODOLOGÍA

### IV.1 Minería de datos

La minería de datos es “la integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión” (Molina y otros, 2001).

La minería de datos es una rama del conocimiento científico que afronta el problema de transformar datos de bajo nivel (datos que son muy numerosos para ser interpretados con facilidad) en información que puede cumplir el objetivo de verificar una hipótesis del usuario, o el descubrimiento de nuevos patrones [5]. Este objetivo se logra a través de la utilización de distintas técnicas y algoritmos que en conjunto con la interpretación del usuario pueden generar conocimiento valioso en base a un gran conjunto de datos.

El término ha sido acuñado en varias metodologías, por ejemplo en el proceso de descubrimiento de conocimiento en bases de datos (*Knowledge Discovery in Database*, KDD), que cuenta de varias etapas definidas para lograr el objetivo, pero además existen otros procesos que también estandarizan los pasos a seguir para lograr el objetivo de la minería de datos, como lo son los procesos CRISP-DM (Cross Industry Standard Process for Data Mining) por sus siglas en inglés, o SEMMA (Sample, Explore, Modify, Model and Asses) y si bien tienen distintos enfoques, existen muchas fases o tareas que son homologables [6].

Tomando esto en cuenta, la minería de datos puede ser interpretada como una serie de metodologías interdisciplinarias que permiten encontrar correlaciones significativas, ajuste de modelos, o relaciones entre variables a partir de los datos disponibles. Esta amplia gama de objetivos incluye también la tarea realizar una predicción [5], lo que hace que sea pertinente considerar sus metodologías para llevar a cabo el objetivo de este trabajo.

No obstante, su beneficio para la predicción no sólo recae en las técnicas antes mencionadas, sino que también con una metodología más global que hace referencia a todo el proceso sobre la construcción del modelo de predicción.

Los proyectos de minería de datos, en general, siguen una metodología como las planteadas anteriormente (KDD, SEMMA, CRISP- DM), no obstante, existen pasos de estas metodologías que son homologables con las otras, y por ende es factible describir un proyecto de minería de datos cumple, independiente de la metodología, con las siguientes etapas [6]:

1. Filtrado de Datos
2. Selección de Variables
3. Extracción de Conocimiento
4. Interpretación y Evaluación

## **IV.I. I Data Warehousing**

El término Data Warehousing introducido por primera vez por Bill Inmon, quien lo define como “una colección de datos orientados a temas, integrados, no-volátiles y variables en el tiempo, organizados para soportar necesidades empresariales” [7]. Ralph Kimball, define un Data Warehouse como “una colección de datos en forma de una base de datos, que guarda y ordena información que se extrae directamente de los sistemas operacionales y datos externos” [8].

Ambas definiciones ven en común un Data Warehouse como una colección de data organizada y orientada a un problema o negocio en particular y proveniente de distintos sistemas o fuentes de datos.

En relación a los procesos de data mining, es necesaria la previa construcción de un consolidado de data que permita trabajar de forma ordenada el modelo. Esta forma ordenada puede ser guiada mediante la construcción de un modelo estrella.

Luego de este diseño teórico, es necesario poblar las tablas mediante un proceso conocido como ETL (Extraction, Transformation, Loading).

### **IV.I.I. I Modelo Estrella**

Consiste en una representación multidimensional de datos utilizando la nomenclatura del modelo Entidad-Relación, pero que no involucra los estándares de normalización asociados a la metodología estándar de creación de bases de datos. Está compuesto por una tabla central (Fact) y un conjunto de tablas dimensionales. Cada tupla que compone la tabla Fact posee como identificador el conjunto de llaves primarias de las dimensiones. A su vez, cada tupla posee atributos asociados a las distintas combinaciones posibles de llaves dimensionales. Si alguna de las dimensiones posee entidades asociadas, el modelo posee el nombre de snowflake.

### **IV.I.I. II ETL**

Como fue descrito anteriormente, las fuentes de datos necesarias para un proceso de minería de datos suelen provenir de distintos sistemas o fuentes de datos, cada una con un formato propio y distinto es entre ellos, además, se deben procesar los datos de manera de ser transformados en información útil para ser cargada en un repositorio. Todo este proceso, llamado ETL [9]:

- **Extracción:** Esta etapa consiste en obtener de las distintas fuentes de datos la data necesaria para posteriormente trabajar sobre esta. La etapa de extracción de la data va muy de la mano de las etapas de entendimiento del negocio y entendimiento de la data descritas más adelante en la metodología Crisp-DM [10], puesto que depende de los objetivos del negocio y de la interpretación de la data cuáles serán los sistemas consultados y la data extraída. Debido a esto, es necesario una estrategia bien definida de extracción de data, considerando la

posibilidad de incorporar iterativamente nuevas fuentes de data y una calendarización adecuada.

- Transformación: Esta etapa resulta ser la más compleja del proceso ETL [9]. Esto a causa de diferentes razones, como son la inconsistencia de unidades de medidas entre distintas fuentes de datos, definición de jerarquías entre datos y generación de valores calculados para almacenar en el DW. Estas dificultades pueden ser abordadas mediante un DSA (Data Stagin Area) que se considera una etapa intermedia donde los datos se cargan únicamente para ser transformados, pero sin ser consolidados aún en el DW definitivo. Esto permite mantener intacta la fuente original de datos en caso de cometer algún error.
- Carga: La carga de datos al DW resulta ser la etapa más sencilla [9], dado que generalmente luego de la etapa de transformación, los datos se encuentran en un formato amigable. Consiste en términos prácticos en la consolidación de todas las fuentes (ya incorporadas en el DSA) y es la base de datos con que se trabajará finalmente.

## **IV.I. II Modelos Predictivos**

Cuando se utilizan métodos y tecnologías para detectar tendencias y patrones en los datos recopilados de diversos recursos, las empresas pueden responder a ellos e identificar nuevas oportunidades. Tanto los datos estructurados como los no estructurados se utilizan en el análisis predictivo.

El análisis predictivo es el nuevo mantra para empresas y expertos en data science porque les brinda una ventaja decisiva al momento de la toma de decisiones, ya que permite en base a patrones repetitivos encontrados en data histórica poder actuar proactivamente y abordar los problemas u oportunidades con un mayor sustento teórico.

La construcción de un modelo predictivo permite también automatizar el análisis de data transaccional, y es esta una de las razones por las que resulta ser una solución pertinente al problema que se trabaja en este trabajo de memoria.

### **IV.I. II. I Árboles de decisión**

Los arboles de decisión son modelos predictivos altamente utilizados en la minería de datos (Vijay, 2015) debido a su facilidad de creación desde el punto de vista de analista y su facilidad de entendimiento desde el punto de vista del negocio. Según lo descrito por (Han, Kamber, & Pei, 2012) estos modelos, son una estructura de nodos y ramas que explican reglas realizadas por el proceso de aprendizaje para la separación de las clases en base a los atributos o variables explicativas entregadas en el set de datos de entrenamiento.

Los nodos son test realizados a algún atributo, donde este test separa los datos en base a una condición. Luego, vienen las ramas, las cuales representan el resultado de la condición entregada por el nodo previo. Por último, están los nodos hojas, los cuales poseen la clase o etiqueta predicha en base a las condiciones realizadas previamente.

El método para la creación y aprendizaje de un árbol inicializa con 3 parámetros: el dataset inicial de entrenamiento, la lista de atributos a analizar y el método de selección de atributos. Se inicia con un nodo, si el nodo posee todos los datos de una misma clase, el nodo se transforma en hoja, si no, se utiliza el criterio de separación el cual indica la mejor forma de separar las ramas en base al atributo del nodo, generando cortes en la variable. Dependiendo del tipo de partición, el nodo se etiqueta en base al criterio y se generan las respectivas ramas con su resultado, donde para cada uno de estos, se vuelven a probar los atributos y se repite el algoritmo, el cual se detiene cuando ocurren 3 tipos de condiciones:

- Los datos dentro de un nodo son todos de la misma clase.
- No existen más atributos en los cuales se pueda separar los datos, por lo cual se utiliza voto mayoritario para etiquetar el nodo con la clase más representante
- No existen datos para una rama, por lo cual se crea una hoja en el nodo del cual nace la rama y se etiqueta con la clase mayoritaria

El resultado del algoritmo genera una serie de reglas que permiten clasificar los datos de acuerdo a la variable supervisada deseada, donde esto permite, de una manera gráfica, entender el comportamiento que presentan los datos.

Dado que el problema de negocio radica en pronosticar una variable de tipo continua, para este proyecto se utilizarán árboles de regresión y algoritmos de ensamble.

## **IV.I. II. II Algoritmos de ensamble**

Los algoritmos de ensamble son un método de combinación de predictores débiles para crear un predictor más robusto. La idea radica en combinar múltiples modelos supervisados para lograr un “modelo final”.

### **IV.I. II. II.I Bagging**

Bagging es un método para generar múltiples versiones de un predictor. La agregación promedia las predicciones de cada predictor generado. Estas múltiples versiones de algoritmos débiles se crean en base a un bootstrapping del dataset de entrenamiento [11].

Entonces, supongamos  $h(d, x)$  es un clasificador, por ejemplo un árbol, basado en el set de data  $D$ , produciendo una predicción dado el input  $x$ . Para hacer bagging sobre  $d$ , se generan submuestras mediante bootstrapping de  $D (d_1, \dots \dots d_M)$  cada uno de tamaño  $N$ , entonces:

$$\hat{H}_{bag}(x) = \text{VotoMayoritario}\{h(d_m, x)\}_{m=1}^M$$

Este método puede reducir la varianza de predictores inestables como los árboles, pero se pierden las propiedades de interpretabilidad de estos.

Algoritmo Bagging	
	M=Número de clasificadores débiles
	$h = \{h_1...h_M\}$ Conjunto de clasificadores débiles.
Input	N Número de registros del set de datos utilizados para ajuste (entrenamiento)
	$N' < N$ , Número de ejemplos generados por bootstrapping
	$d \subset D$ Conjunto de entrenamiento
	$g()$ Función de ensamble
Output	Modelo de ensamble $H_{bag}(X)$ .
	<b>para</b> $m = 1...M$
	Generar conjunto $d_m$ seleccionando $N'$ muestras por reemplazo desde $d$
	Entrenar clasificador $h_m$ sobre el subset $d_m$
	Agregar $h_m$ al ensamble
	<b>fin</b>
	devolver $\hat{H}_{bag}(x) = g(h_1(x), \dots, h_M(x))$

**Tabla 4: Algoritmo de Bagging.**  
Fuente: Adaptado de [11].

## IV.I. II. II.II Random Forest

Random Forest es un algoritmo de ensamble creado por Leo Breiman en 1994, que mejora la exactitud de en la clasificación mediante la incorporación de aleatoriedad en la construcción de los clasificadores débiles. Donde la diferencia con el algoritmo de Bagging radica en que cada split el algoritmo escoge solo un subconjunto de predictores, esto implica que la aleatoriedad está presente tanto en la selección del conjunto de entrenamiento como en el conjunto de atributos (variables explicativas del modelo) [11].

## IV.I. II. II.III Boosting

El problema del aprendizaje predictivo puede ser caracterizado por un vector de variables explicativas o predictoras de un fenómeno  $x = \{x_1, \dots, x_m\}$  y un vector de variable independiente  $y$ .

Dada  $M$  instancias  $\{y_i, x_i\}$  con  $i = 1, \dots, M$  observadas, el objetivo es usar esta función para predecir instancias donde solo valores de  $x$  son observados. Formalmente, se intenta estimar la función  $\hat{f}(x): x \rightarrow y$  que minimiza una función de pérdida  $L(y, f)$  sobre la distribución conjunta de  $W$  de  $(y, x)$  :

$$\hat{f}(x) = \underset{f(x)}{\operatorname{argmin}} E_{y,x} L(y, f(x))$$

Algoritmo Boosting	
<b>Input</b>	$M$ = Número de clasificadores débiles $h = \{h_1 \dots h_M\}$ Conjunto de clasificadores débiles. $N$ Número de registros del set de datos utilizados para entrenamiento $d \subset D$ Conjunto de entrenamiento $g()$ Función de ensamble
<b>Output</b>	Modelo de ensamble $H_{boost}(X)$ . inicializar la distribución $W = \{w_1, \dots, w_N\}$ de las muestras <b>para</b> $m = 1 \dots M$ Entrenar modelo $h_m$ sobre $d_m$ y su distribución $w_m$ Evaluar el error $\varepsilon_m \leftarrow P_{x \sim W}(\mathbb{I}(h_m(x) \neq y_i))$ Actualizar la distribución $W$ a partir del error $\varepsilon_m$ . <b>fin</b> devolver $H_{boost}(x) = g(h_1(x), \dots, h_M(x))$

**Tabla 5: Algoritmo de Boosting.**

**Fuente: Adaptado de [13].**

Intuitivamente, la idea detrás de los métodos de boosting es que la combinación de predictores débiles produce un mejor modelo en términos de capacidad predictiva. Un predictor débil es un algoritmo cuya capacidad predictiva es levemente mejor que un resultado aleatorio, y con baja capacidad de generalización.

Cada predictor débil permite caracterizar reglas locales relativas a pocas variables, que, si bien no es suficiente para lograr una buena capacidad predictiva, la combinación de muchos de estos predictores permite generalizar y construir un modelo más robusto. Esta idea fue propuesta por Schapire el 1990, y es conocida como “The strength of weak learnability” [13].

### IV.I.III Validación

Esta etapa es crucial para cualquier proyecto de minería de datos, y es la que permite evaluar si el modelo que estamos construyendo cumple o no con el objetivo que buscamos.

Para llevar a cabo este procedimiento, se dividen los datos disponibles para generar dos conjuntos que tienen distintas funciones. Estos conjuntos son llamados:

- Conjunto de Entrenamiento
- Conjunto de Validación

Sobre el primero de estos conjuntos se aplica el algoritmo de aprendizaje con el objetivo de estimar una función. Una vez que se ha estimado esta función con el conjunto de entrenamiento, se pasa a utilizar el conjunto de validación.

En el caso de este conjunto se aplica la función estimada, utilizando como argumento de la función los datos de las variables explicativas que se poseen. El resultado de esto es que se obtiene para cada observación del conjunto de validación un valor estimado de la variable de interés. Luego, como sabemos cuál es valor verdadero al cual debería haber

Llegado la función, calculamos un error el cual permite evaluar el rendimiento de la función en un caso de la vida real.

En otras palabras, el objetivo del conjunto de validación es poner a prueba a la función estimada en el de entrenamiento. En general, el conjunto de entrenamiento posee más observaciones que el conjunto de validación, siendo alrededor del 70% de los datos totales disponibles, mientras que el 30% correspondería al de validación.

Dado que en este trabajo se busca realizar un pronóstico, se expondrán algunas de las medidas clásicas de error utilizadas en este caso.

Suponiendo que  $f_i$  represente el valor pronosticado para una observación,  $y_i$  el valor real a obtener en esa observación, y  $n$  el número total de observaciones a evaluar, los errores más utilizados son:

- MAE (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$

- MAPE (Mean Absolute Percentage Error)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i - y_i}{y_i} \right|$$

- MSE (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2$$

- RMSE (Root-Mean Squared Error)

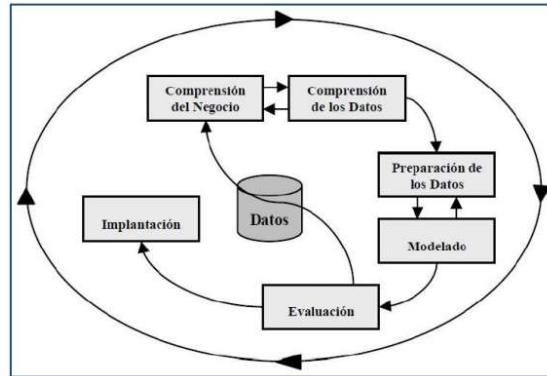
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2}$$

## IV.II CRISP-DM

Los proyectos de minería de datos, en general, siguen una metodología como las planteadas anteriormente (KDD, SEMMA, CRISP- DM), sin embargo, tras revisarlas en detalle, se utilizará la metodología CRISP- DM, cuyo enfoque es obtener el mejor provecho del uso de Data Mining al entender de la manera más completa posible el negocio y el problema que se desea resolver. Lo anterior, permite hacer una correcta recolección de datos e interpretar bien los resultados de los análisis, alcanzando los objetivos propuestos.

CRISP-DM organiza el desarrollo de un proyecto de Data Mining en una serie de etapas, con tareas generales y específicas que permitirán cumplir con los objetivos de la memoria.

Estas etapas funcionan de manera cíclica e iterativa, pudiendo regresar desde alguna fase a otra anterior. En la *figura 4* se presenta un diagrama con las fases propuestas por esta metodología:



**Figura 4: Etapas del modelo CRISP-DM.**  
**Fuente: Obtenido de [10].**

Entonces se plantea la siguiente metodología de trabajo, la cual tiene el fin de cumplir los objetivos propuestos:

## **1. Comprensión del negocio.**

Lo primero es comprender o definir el problema de negocio, lo cual puede ser el paso más importante de la metodología pues permite entender los objetivos y requisitos que tendrá el proyecto. Si el problema no se comprende bien, no se sabrá que algoritmo de Data Mining utilizar para realizar una proyección de demanda óptima.

Para esto, se realizaron visitas a la Faena de Minera Escondida, específicamente a las concentradoras Los Colorados, Laguna Seca 1 y 2 para entender cómo funciona en detalle el proceso de molienda, es decir, cuales son los inputs que influyen en la cantidad de bolas de molienda a utilizar en los molinos SAG y molinos de bolas.

## **2. Comprensión de los datos.**

Teniendo definido el problema y el proceso de negocio, se realizó un levantamiento de los datos y de las variables que se utilizarán para pronosticar la demanda futura de las bolas de molienda.

Las principales tareas que se realizaron en esta etapa son [14]:

- a) Recolección de datos, clarificando desde que lugar fueron obtenidos.



- b) Descripción de los datos, estableciendo los volúmenes de información con que se trabajó y los significados de cada variable.
- c) Exploración de los datos, estableciendo una estructura general de la información, comprobando frecuencia y distribución de estos.
- d) Verificación de la calidad de los datos, determinando la consistencia de los valores, comprobando la existencia de datos nulos y fuera de rango, identificando irregularidades para asegurar la completitud y exactitud de los datos.

### **3. Preparación de los datos.**

Una vez que se recopiló la información y datos necesarios, se adaptaron para luego utilizar los algoritmos de Data Mining correspondientes.

Los pasos a seguir en esta etapa fueron los siguientes [14]:

- a) Selección de datos, escogiendo un subconjunto de los datos recopilados en la etapa anterior.
- b) Limpieza de los datos, preparándolos para la fase de modelación, aplicando técnicas de normalización, tratamiento de valores nulos.
- c) Estructuración de los datos, generando nuevos atributos a partir de los existentes.
- d) Integración de los datos, lo que permitió agrupar tablas o campos relacionados, definiendo una estructura que los contenga.
- e) Formateo de los datos, que consistió en transformar los datos sin modificar su significado, para que se ajustaran a las técnicas de Data Mining utilizadas.

### **4. Modelamiento y evaluación.**

En esta fase se escogieron las técnicas de modelado más apropiadas para realizar el pronóstico de demanda. Elegir una técnica de Data Mining para aplicar al problema no fue sencillo, pues debe cumplir los requisitos del problema en estudio junto a criterios tales como: bondad de ajuste, complejidad computacional, supuestos detrás (forma funcional o distribución del error) y propiedades del modelo (interpretabilidad, tendencia a sobreajuste).

Sin embargo, se realizaron modelos de árboles de regresión y algoritmos de ensamble para cada Concentradora, agregando el consumo de bolas de molinos SAG y bolas en las respectivas Concentradoras.

Luego de modelar el problema, se evaluó el desempeño predictivo de los modelos construidos. Los mejores modelos fueron seleccionados de acuerdo con el menor RMSE.

En resumen, para el modelamiento y evaluación se realizaron las siguientes tareas [14]:

- a) Selección de la técnica de modelado, escogiendo la más apropiada para el tipo de problema, en base al objetivo principal del proyecto.

- b) Generación del plan de prueba, diseñando un procedimiento para probar y validar el modelo a obtener. En general, se separa el conjunto de datos en dos: una parte de los datos destinada a entrenamiento del modelo y otra parte que será utilizada para las pruebas
- c) Construcción del modelo, a partir de la técnica de modelado escogida, se aplica sobre el conjunto de datos para generar uno o más modelos. En este punto se van ajustando los parámetros de la técnica seleccionada de forma iterativa para obtener mejores resultados.
- d) Evaluación del modelo, interpretando los modelos en base al conocimiento existente y los criterios de éxito ya establecidos.
- e) Evaluación de los resultados, en base a los objetivos del negocio y basándose en el juicio de la contraparte por parte de la empresa.
- f) Proceso de revisión, que consiste en calificar el proceso completo de minería de datos realizado y ver si existen puntos a mejorar.

## **5. Implementación.**

Esta etapa no está contemplada dentro del alcance del trabajo de título, sin embargo, desde las etapas anteriores se obtendrá conocimiento a partir de los datos. Con el modelo ya construido y validado, este conocimiento pretende transformarse en recomendaciones a realizar dentro del proceso de negocio de BHP, específicamente para la realización del próximo forecast de bolas de molienda, correspondiente al año fiscal 21.

## V. DESARROLLO DEL PROYECTO DE MEMORIA

### V.I Macro Proceso de Producción de Cobre:

El negocio central de Minera Escondida Limitada (de ahora en adelante MEL) radica en la extracción, concentración y posterior embarque de concentrado y cátodos de cobre. El macro proceso se ilustra en la siguiente figura:



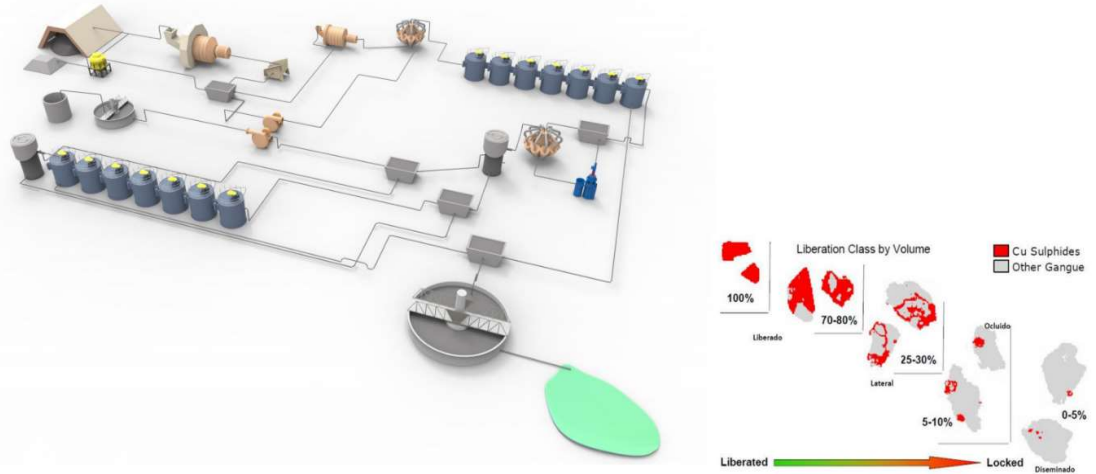
**Figura 4: Macro Proceso de Producción de Cobre en Minera Escondida.**  
**Fuente: Obtenido de [15].**

Los minerales sulfuros de alta ley (calcopirita, calcosita, bornita, covelina, pirita) son aquellos que siguen el proceso de la parte arriba de la figura, los cuales pasan por el subproceso de Molienda en las Concentradoras de Los Colorados, Laguna Seca 1 y 2, transformándose finalmente en Concentrado de Cobre. Entonces, el proceso de elaboración de Concentrado de Cobre secuencialmente contempla:

- Perforación, tronadura y transporte desde Rajo Escondida Principal y Rajo Escondida Norte a concentradoras Laguna Seca 1, 2 y Los Colorados.
- Chancado primario: Reducción de tamaño del mineral.
- Molienda: Liberación de material útil
- Flotación: Separación / Concentración.
- Manejo de relaves: Recuperación de Agua.
- Manejo de concentrado: Ajuste de sólido para transporte.
- Transporte desde MEL hacia Puerto Coloso por Mineroducto de 165 kilómetros.
- Filtrado, secado y embarque en Puerto Coloso.

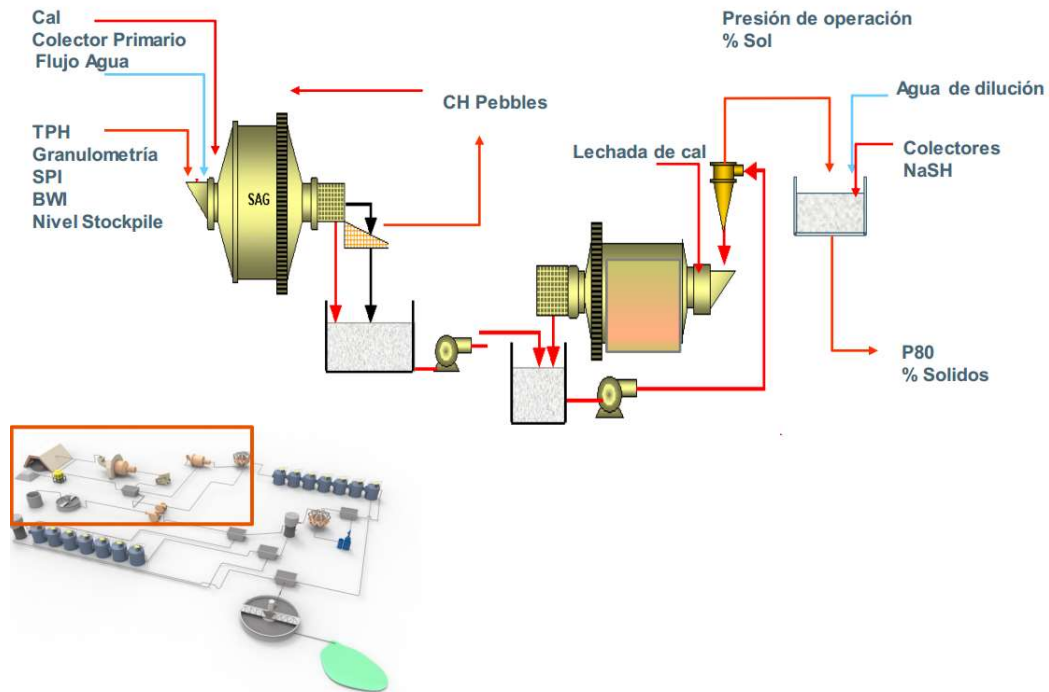
## V.I. I Proceso en Concentradora

MEL cuenta con 3 concentradoras: Laguna Seca 1, Lagunas Seca 2 y Los Colorados. El proceso de concentración de los sulfuros de cobre se resume en la siguiente figura:



**Figura 5: Proceso de Concentración de Minerales Sulfurados.**  
**Fuente: Obtenido de [15].**

Este proceso tiene como objetivo transformar los sulfuros de cobre en concentrado de cobre, en donde la molienda se encarga de liberar el mineral útil y luego la flotación separa la especie valiosa liberada desde la ganga. Ahora hay que enfocarse en el subproceso que se enmarca este trabajo de título, llamado molienda, el cual se observa en detalle en la siguiente figura:



**Figura 5: Proceso de Molienda.**  
**Fuente: Obtenido de [15].**

El proceso de molienda cuenta básicamente de 2 etapas: Molienda Primaria que se lleva a cabo en los Molinos SAG y Molienda Secundaria que se lleva a cabo en los Molinos de Bolas. La Filosofía de operación es la siguiente:

Alimentación al circuito de molienda:

Desde el acopio de mineral grueso, y a través del conjunto de alimentadores se abastece la totalidad del flujo demandado por molienda (según la capacidad de cada planta concentradora) , flujo que corresponde a la alimentación del molino SAG.

El variador de frecuencia instalado en la correa transportadora que alimenta al molino SAG tiene dos objetivos fundamentales, primero actuar como arrancador-partidor- suave y segundo como herramienta de mantenimiento. Además, la velocidad variable de la correa de alimentación al molino SAG permite reducir el tiempo de respuesta para alcanzar el régimen del circuito, ante un requerimiento de modificación del tratamiento de la planta, controlada y visualizada por el operador de molienda vía un pesómetro ubicado antes del retorno de los pebbles, en la correa de alimentación al molino .

La alimentación de mineral al molino SAG será controlada por medio de la variación de la velocidad de los alimentadores de cinta actuando en conjunto con el pesómetro.

Operación Molino SAG:

Al momento de tener carga en el molino se procede a la adición de reactivos cal y colector en modo automático. Se destaca en este punto que la adición de colector primario y la lechada de cal, se controla para ambos casos mediante un control de la razón asociada

al tonelaje alimentado. También se agrega agua con el objetivo de obtener un porcentaje de sólidos adecuado para la operación del molino. Esta condición mínima recomendada no debe ser inferior a un 57% en condición nominal, condiciones menores de sólidos se pueden ajustar a través de una reducción del tonelaje a tratar, manteniendo un control de sólido en las canaletas de alimentación a los molinos de bolas.

#### Carguío de bolas a los molinos SAG y Bolas:

Los sistemas de almacenamiento y carguío de bolas contemplan tolvas de almacenamiento, de 400 toneladas de capacidad cada una, provistas de sistemas de descarga y conteo, dos tolvas molino SAG y dos tolvas molinos de bolas. Para el caso del molino SAG cada una de estas tolvas utiliza un alimentador rotatorio el cual descarga las bolas de 5" de diámetro, sobre la correa de alimentación SAG. Para las tolvas de los molinos de bolas, estas también tienen una capacidad de 400 toneladas, para bolas de 3" de diámetro. Cada una de ellas descarga mediante un alimentador rotatorio a la correa alimentadora, la cual dispone de compuertas de desviación neumáticas que reparten las bolas hacia cada uno de los molinos de bolas. El concepto de diseño indica que una operación de alrededor 1.0 a 1.5 horas por día permite realizar la recarga de bolas a los molinos.

El control de la adición de bolas al molino SAG se realiza mediante sistema óptico de conteo, en tanto para los molinos de bolas la correa cuenta con un pesómetro que realiza este control.

La recarga de bolas a los molinos de bolas sólo se podrá realizar en aquellos equipos que estén operando, ya que las compuertas que permiten la alimentación del mismo no se abrirán si el molino está detenido.

#### Descarga del Molino SAG:

El molino SAG cuenta, para la clasificación de su descarga, con un Trommel 6.2 x 5.2 m con una apertura de 25 x 60 mm y harnero de pebbles de 12 x 24' (ft) doble deck con apertura de 35x60 mm en deck superior y 25 x 60 mm en deck inferior. Ambos equipos requieren de la adición de agua de procesos de modo de efectuar una eficiente clasificación de los pebbles. Se ha conceptualizado dicha adición en forma manual, de modo que el operador de terreno debe ajustar el caudal en cada punto, siendo el objetivo la obtención de pebbles lo más libre, que sea posible, de finos (fuente de barro y derrames en las correas y de ineficiencias en el chancado).

#### Alimentación y Descarga de los Molinos de Bolas:

El underflow de cada batería de ciclones descarga de forma gravitacional en los chutes de alimentación de los molinos de bolas. Además por estos chutes se realiza la adición de agua de procesos, y medios de molienda (bolas 3") en el circuito de molienda.

Cada molino de bolas, cuenta con un sistema de velocidad variable y operará con un nivel de llenado cercano a un 30%, para alcanzar la distribución de tamaño para la alimentación a flotación, de un  $P_{80}$  de 145 micrones.

La descarga de molinos de bolas es a través de una parrilla que permite retener bolas y material grueso con el objeto de mantenerlos al interior del molino antes de alimentar de forma gravitacional el cajón de alimentación de ciclones de molienda.

## **V.I. II Proceso de carguío de Bolas:**

El carguío de bolas se realiza diariamente para cada molino SAG y molino de bolas presente en MEL, el cual tiene por objetivo mantener la carga a un determinado nivel de bolas, el cual depende del tipo de molino.

### **V.I. II. I Carguío en Molino SAG:**

El carguío diario en molinos SAG se realiza de acuerdo a una tasa de desgaste (gr/ton), la que tiene un valor referencial según consumos históricos entre 190-220 (gr/ton). Este carguío se realiza con el fin de alcanzar un nivel de llenado objetivo, correspondiente al 19,5%.

Luego, para un día de operación, la carga solicitada corresponde al producto entre la tasa de desgaste y el tonelaje de alimentación para el respectivo día. Esta carga solicitada, tras ser revisada por el área Planificación y Control de Procesos puede sufrir variaciones de acuerdo a problemas operacionales, mantenciones de planta o ajustes de nivel de llenado.

Finalmente, no se puede abrir y observar el nivel de llenado de los molinos todos los días, esto se realiza cada 3 meses para obtener el nivel de llenado real. Luego según los gramos de bolas cargadas y tonelaje tratado en ese período, se busca ajustar la tasa de desgaste para los próximos 3 meses con el fin de alcanzar el nivel de llenado objetivo.

### **V.I. II. II Carguío en Molinos de Bolas:**

El carguío diario en molinos de bolas se realiza de acuerdo a una tasa de desgaste (gr/kW) la que tiene un rango normal según consumos históricos entre (55 – 65 gr/kW). Este carguío se realiza con el fin de reponer el desgaste de bolas y mantener un nivel de llenado objetivo, correspondiente al 33%.

Luego, para un día de operación la carga solicitada corresponde al producto entre la tasa de desgaste (gr/kW) y la potencia consumida el día anterior (kW). Esta carga solicitada también es revisada por el área de Planificación y Control de Procesos y puede variar de acuerdo de acuerdo a las situaciones ya mencionadas.

Al igual que en los molinos SAG, la apertura de los molinos de bolas no se puede realizar todos los días, sino que se realiza cada 6 meses. Entonces, para llevar un control del nivel de llenado, este se relaciona con una variable conocida; la potencia del molino. Para esto, se calculan modelos de regresión lineal que permiten establecer una potencia objetivo para el nivel de bolas ya mencionado.

Luego, cada 6 meses y la respectiva apertura de cada molino, se busca ajustar la tasa de desgaste para el próximo período en base a los gramos de bolas cargadas y la potencia consumida durante ese período, con el fin de alcanzar el nivel de llenado objetivo.

## V.II Consolidación de la base de datos:

Para desarrollar el modelo predictivo es necesario primero construir un Data Warehouse con todas las bases de datos necesarias consolidadas. Las bases de datos disponibles son:

Database	Volumen	Cantidad de Variables
Carguío de bolas L1-L2	968	7
Carguío de bolas LC	610	6
Mineral Alimentado (MD)	1017	15
Consumo de Potencia Molinos	1012	3

**Tabla 6: Volumen y variables de las bases de datos disponibles.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

La base de datos “Carguío de bolas L1-L2” contiene información acerca del carguío de bolas desde 16 de septiembre del 2016 hasta el 30 de junio del 2019 de las Concentradoras Laguna Seca 1 y 2, respectivamente. A continuación, se detallan las variables consideradas en este reporte:

Variable	Descripción
Fecha	Día de operación
TPD	Toneladas de mineral cargado por día
Carga realizada (ton)	Carga de bolas realizadas por Molino, ya sea SAG o de bolas
Tasa de desgaste (g/ton, real cargado)	Tasa de desgaste real para bolas de cada molino en gramos por toneladas
Nivel de bolas estimado (%)	Nivel de bolas estimado para cada molino
BWI (kWh/t)	Bond Work Index
SPI (min.)	SAG Power Index

**Tabla 7: Variables de Carguío de bolas L1-L2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Para el caso de “Carguío de bolas LC” contiene la misma información, sin embargo, esta se encuentra disponible desde el 9 de agosto de 2017 hasta el 30 de junio de 2019.



Luego, para la base de datos “Mineral Alimentado (MD)”, contiene información de la mineralogía de alimentación de las Concentradoras, la cual cubre el mismo periodo ya descrito para Los Colorados y Laguna Seca. Las variables consideradas en este reporte son:

<b>Variable</b>	<b>Descripción</b>
Fecha	Día de operación
TPD	Toneladas de mineral cargado por día
CuT (%)	Ley de Cobre en Porcentaje
CuS (%)	Ley de Cobre Oxidado en Porcentaje
Fe (%)	Cantidad de Hierro en Porcentaje
As (ppm)	Cantidad de Arsénico en partes por millón
Mo (ppm)	Cantidad de Molibdeno en partes por millón
Au (ppm)	Cantidad de Oro en partes por millón
Ag (ppm)	Cantidad de Plata en partes por millón
BWI (kWh/t)	Bond Work Index
SPI (min.)	SAG Power Index
Py (%)	Cantidad de Pirita en Porcentaje
Cpy (%)	Cantidad de Calcopirita en Porcentaje
Cc (%)	Cantidad de Calcosita en Porcentaje
Cv (%)	Cantidad de Covelina en Porcentaje

**Tabla 8: Variables de Mineral Alimentado.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**

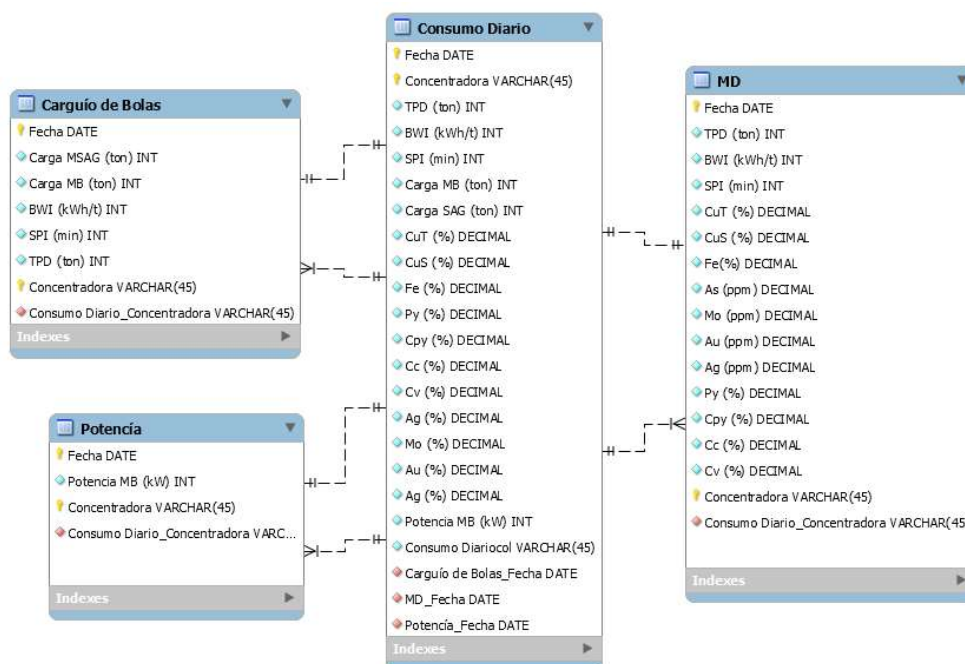
Finalmente, la base de datos “Consumo de Potencia Molinos” contiene el consumo de potencia diario de los molinos de bolas en las Concentradoras de Minera Escondida. Las variables consideradas en este reporte son:

<b>Variable</b>	<b>Descripción</b>
Fecha	Día de operación
Potencia MB (kWh)	Potencia diaria consumida en molinos de bolas

**Tabla 9: Variables de Potencia Consumida.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**

La base de datos principal que se utilizará más adelante para entrenar un modelo, se construye a partir del siguiente esquema:



**Figura 6: Esquema del Data Warehouse a Construir.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

### V.III Caracterización de los datos históricos:

Una vez consolidada la base de datos recientemente mencionada, el repositorio final cuenta con la siguiente estructura:

Database	Volumen	Variables
Laguna Seca 1	802	18
Laguna Seca 2	749	18
Los Colorados	494	19

**Tabla 10: Repositorio final de datos.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Entonces, el volumen de datos final queda definido por la disponibilidad de la data, es decir, la Concentradora en cuestión.

### V.IV Estadística Descriptiva:

Se realizará un análisis descriptivo para cada concentradora, centrándose principalmente en los consumos históricos de bolas 5,5” en molinos SAG y bolas 3” en molinos de bolas.

**Laguna Seca:**

<b>Statistic</b>	<b>N</b>	<b>Mean</b>	<b>St. Dev.</b>	<b>Min</b>	<b>Pctl(25)</b>	<b>Pctl(75)</b>	<b>Max</b>
TPD	802	119,972.3	18,964.7	246.0	114,453.5	131,369.7	167,258.1
BWI	802	12.1	0.8	7.0	11.6	12.6	16.5
SPI	802	51.2	7.5	29.4	46.2	55.6	86.4
MSAG	802	21.4	9.3	0.0	18.5	23.0	84.0
MB	802	60.1	24.1	0.0	47.0	72.1	208.3
CuT	802	1.0	0.2	0.6	0.8	1.0	2.2
CuS	802	0.05	0.03	0.01	0.03	0.1	0.3
Fe	802	2.4	0.5	1.3	2.1	2.7	6.5
As	802	62.6	28.0	0.0	44.5	72.3	279.3
Mo	802	77.7	39.3	0.0	51.3	100.2	333.8
Au	802	0.1	0.6	0.0	0.1	0.1	18.4
Ag	802	2.4	1.5	0.0	1.7	3.0	33.6
Pirita	802	4.6	15.7	0.3	2.2	3.3	215.2
Calcopirita	802	63.5	1,780.9	0.0	0.3	0.5	50,467.0
Calcosita	802	22.4	613.9	0.0	0.2	0.6	17,396.3
Covelina	802	19.2	538.5	0.0	0.1	0.1	15,258.6
PotenciaMB	802	1,123,640	197,196	14,411	1,062,967	1,250,147	1,331,760

**Tabla 9 : Estadística descriptiva Laguna Seca 1.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

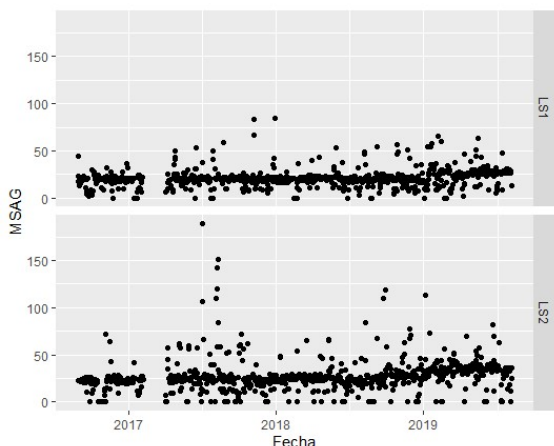
<b>Statistic</b>	<b>N</b>	<b>Mean</b>	<b>St. Dev.</b>	<b>Min</b>	<b>Pctl(25)</b>	<b>Pctl(75)</b>	<b>Max</b>
TPD	749	140,415.5	29,074.8	0.0	132,006.1	157,074.0	182,620.0
BWI	749	12.4	0.7	10.4	11.9	12.9	14.7
SPI	749	52.1	6.4	34.5	47.8	55.9	75.2
MSAG	749	28.1	16.6	0	22	32.4	189
MB	749	61.5	29.7	0	44	79	204
CuT	749	0.9	0.2	0.5	0.8	1.0	2.0
CuS	749	0.04	0.02	0.01	0.03	0.05	0.2
Fe	749	2.2	0.3	1.4	2.0	2.4	4.2
As	749	54.3	25.7	0.0	37.4	63.2	205.5
Mo	749	58.3	39.1	0.0	32.7	78.9	254.2
Au	749	0.1	0.04	0.0	0.1	0.1	0.3
Ag	749	3.0	1.4	0.0	2.3	3.8	7.2
Pirita	749	4.3	18.7	0.1	1.9	2.8	237.1
Calcopirita	749	0.7	2.8	0.0	0.4	0.5	39.0

Calcosita	749	0.6	3.0	0.0	0.2	0.4	45.5
Covelina	749	0.2	0.8	0.0	0.1	0.1	13.3
PotenciaMB	749	1,357,666	233,655	3,349	1,335,721	1,471,825	1,621,927

**Tabla 10: Estadística descriptiva Laguna Seca 2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

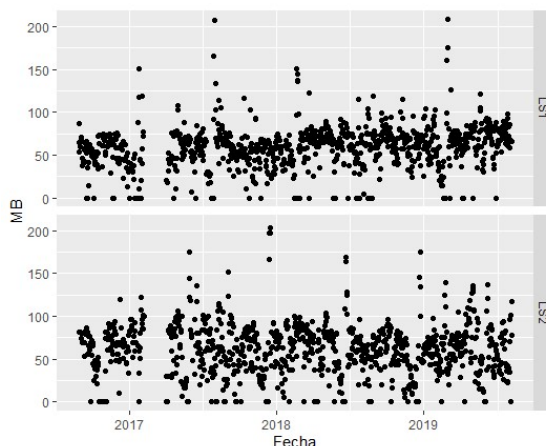
### Consumo de bolas:

La data histórica del consumo de bolas 3” y 5,5” se obtiene directamente de las planillas de carguío de bolas, donde el consumo de bolas de molino se agrega en los molinos por cada Concentradora para observar el consumo en el período considerado en los siguientes gráficos:



**Figura 9: Consumo diario de bolas 5,5”.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**



**Figura 10: Consumo diario de bolas 3”.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**

Se puede observar que el consumo de bolas 3” presenta mayor variabilidad que el de bolas 5,5” y que ambos presentan “outliers” dentro del período de análisis. Además, no hay una correlación observable entre la fecha y el consumo, por lo que posiblemente hay otras variables y/o características que explican la variable dependiente, que serán estudiadas en los modelos.

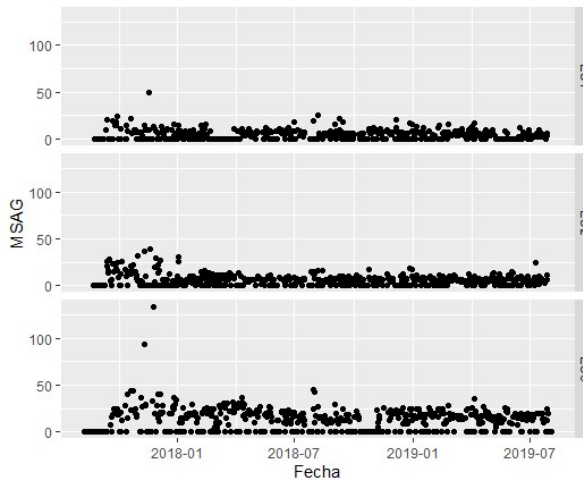
### Los Colorados:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
TPDSAG	492	86,330.1	31,945.8	0.0	64,676.5	112,200.7	131,599.8
TPDMB	492	86,252.0	31,942.3	0.0	64,676.5	112,046.4	131,599.8
MSAG	492	20.8	16.0	0	9.3	29.9	148
MB	492	40.6	30.1	0	17.5	58.2	220
CuT	492	0.9	0.2	0.6	0.8	1.0	2.2
CuS	492	0.04	0.03	0.01	0.03	0.1	0.3

Fe	492	2.5	0.5	1.2	2.1	2.8	5.5
As	492	65.2	25.9	0.0	48.6	76.3	252.4
Mo	492	91.3	46.6	0.0	64.8	114.5	636.2
Au	492	0.2	1.0	0.0	0.1	0.1	21.3
Ag	492	2.5	2.0	0.0	1.7	3.1	38.8
BWI	492	11.9	0.9	9.2	11.3	12.4	17.1
SPI	492	48.9	8.2	23.5	43.4	54.3	77.3
Pirita	492	4.6	17.7	0.05	2.2	3.3	256.4
Calcopirita	492	37.4	873.1	0.0	0.4	0.5	20,790.9
Calcosita	492	13.3	301.0	0.0	0.2	0.5	7,166.6
Covelina	492	11.3	264.0	0.0	0.1	0.2	6,286.0
PotenciaMB	492	675,612	172,497	130,974	574,420	821,482	910,087

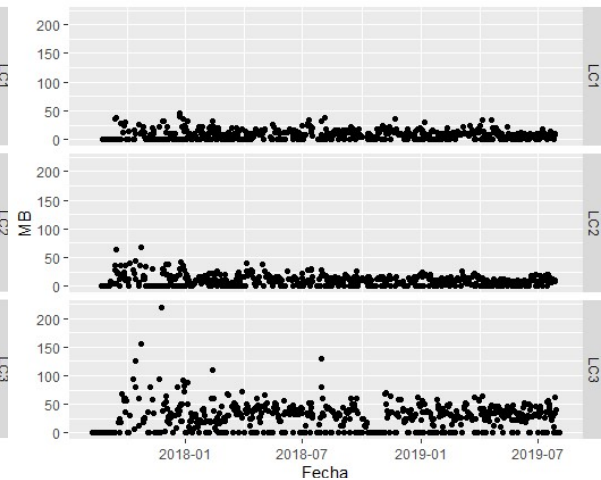
**Tabla 11: Estadística Descriptiva Concentradora Los Colorados.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

### Consumo de bolas:



**Figura 11: Consumo de Bolas 5,5''.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**



**Figura 12: Consumos de Bolas 3''.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**

Se puede observar menor variabilidad en el consumo de ambos tipos de bolas con respecto a Laguna Seca. Por otro lado, hay una leve variabilidad superior en el tratamiento de la línea 3 de Los Colorados. Tampoco se observa una correlación entre la fecha y el consumo, por lo que posiblemente hay otras variables que explican el consumo de las bolas.

## V.IV Modelo predictivo como solución al problema planteado

A continuación, se propone una solución al problema planteado en esta tesis. Para ello se aplican los pasos de la metodología explicada en el capítulo anterior. Si bien es evidente que un modelo predictivo basado en machine learning en teoría es capaz de predecir, en base a datos históricos el consumo de bolas de molienda, habrá que analizar si esto tiene una repercusión tanto a nivel operacional aportando día a día información valiosa sobre que variables están influyendo de manera significativa en el consumo, como a nivel de planificación comercial para aumentar la eficiencia del forecast en algún año fiscal próximo.

Luego, la obtención del modelo predictivo de bolas de molienda para cada molino y Concentradora se realizará utilizando algoritmos de data mining, donde la variable dependiente a predecir con el modelo es el consumo de bolas de molienda de 3" y 5,5" en toneladas. Las variables independientes dependen del tipo de molino y buscaran explicar la tasa de desgaste de las bolas, es decir, como impacta la mineralogía de alimentación, índice de SAG, índice de Bond, en conjunto al tonelaje y la potencia en la reposición de bolas diarias.

<b>Molinos SAG</b>	<b>Molinos de bolas</b>
Toneladas de mineral cargado por día (ton)	Potencia consumida el día anterior (kW)
Índice de SAG (min.)	Índice de Bond (kWh/ton)
Ley de cobre (%)	Ley de cobre (%)
Ley de cobre oxidado (%)	Ley de cobre oxidado (%)
Ley de fierro (%)	Ley de fierro (%)
Cantidad de arsénico (ppm)	Cantidad de arsénico (ppm)
Cantidad de molibdeno (ppm)	Cantidad de molibdeno (ppm)
Cantidad de oro (ppm)	Cantidad de oro (ppm)
Cantidad de plata (ppm)	Cantidad de plata (ppm)
Cantidad de pirita (%)	Cantidad de pirita (%)
Cantidad de calcopirita (%)	Cantidad de calcopirita (%)
Cantidad de covelina (%)	Cantidad de covelina (%)
Cantidad de calcosita (%)	Cantidad de calcosita (%)

**Tabla 12: Variables independientes a utilizar en modelos predictivos.**  
*Fuente: Elaborado por el autor en base a datos de la empresa.*

#### **V.IV. I Selección del algoritmo a utilizar: Challenge de modelos de machine learning**

En esta sección se comparan 3 modelos para verificar cuál de ellos provee mejores resultados en términos de distintas métricas. Se entrenan 3 distintos modelos de machine learning; Bagging Trees (BT), Random Forest (RF) y Gradient Boosting Trees (GBT) para molinos SAG y de bolas en las respectiva Concentradoras. Los resultados se muestran a continuación:

##### **Molinos SAG:**

Algoritmo	RMSE	Rsquared	MAE
LS1 BT	7,51	0,072	4,71
LS1 RF	7,19	0,168	4,47
LS1 GBT	7,17	0,204	4,52
LS2 BT	10,31	0,384	6,43
LS2 RF	10,29	0,36	6,34
LS2 GBT	10,17	0,432	6,35
LC BT	11,45	0,408	8,27
LC RF	11,28	0,456	8,06
LC GBT	11,21	0,48	8,20

**Tabla 13: Resultados Modelos de Machine Learning en Molinos SAG.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

### **Molinos de Bolas:**

Algoritmo	RMSE	Rsquared	MAE
LS1 BT	19,75	0,368	14,15
LS1 RF	19,68	0,39	13,88
LS1 GBT	19,68	0,386	13,66
LS2 BT	25,60	0,166	19,69
LS2 RF	24,91	0,24	19,11
LS2 GBT	24,95	0,256	19,44
LC BT	22,45	0,50	16,25
LC RF	22,60	0,492	16,22
LC GBT	22,17	0,54	16,02

**Tabla 14: Resultados Modelos de Machine Learning en Molinos de Bolas.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Cabe mencionar que, dado que los modelos basados en árboles incorporan en su algoritmo la selección de variables, se consideran todas las variables mencionadas anteriormente para el estudio y respectivo entrenamiento de los modelos.

Los resultados presentados en las tablas anteriores surgen de un proceso de validación cruzada (K-Folds Cross Validation) con K = 5. Es decir, se realiza una partición aleatoria de los datos para entrenamiento (K-1 sets de datos) y se deja el resto de los datos para testeo, proceso iterativo que se realiza 5 veces con distintas particiones para finalmente entregar el promedio de las métricas de desempeño de los modelos.

Se puede notar que independiente de las concentradora y molino, los modelos que entregan mejor desempeño son los Gradient Boosting Trees (GBT), por ende, el algoritmo a utilizar será este.

## **V.IV. II Construcción de los modelos finales**

El modelo a utilizar es el Gradient Boosting Trees (GBT), implementado vía el software RStudio, utilizando el paquete de funciones gbm y caret [16,17].

Se asume una distribución del tipo Gaussiana ya que se pretende predecir una variable continua.

El paquete caret permite utilizar una grilla para buscar los hiperparámetros óptimos de un modelo GBM mediante la validación cruzada e iteración de distintos valores apropiados a los parámetros. La grilla utilizada y los hiperparámetros seleccionados de acuerdo al menor RMSE son:

Hiperparámetro	Vector de valores
Shrinkage	[1, 3 , 6, 9 , 12]
Interaction depth	[0,005 , 0,001]
Número de árboles	[100, 250, 500, 1.000, 5.000 , 10.000]
Mínimo de observaciones por nodo	[10]

**Tabla 15 : Grilla para buscar hiperparámetros óptimos.**  
**Fuente: Elaborado por el autor.**

Hiperparámetro	LS1 MSAG	LS2 MSAG	LC MSAG	LS1 MB	LS2 MB	LC MB
Shrinkage	0.005	0.005	0.005	0.005	0.005	0.001
Interaction depth	9	12	3	9	12	9
Número de árboles	250	250	500	500	500	1000
Mínimo de observaciones por nodo	10	10	10	10	10	10

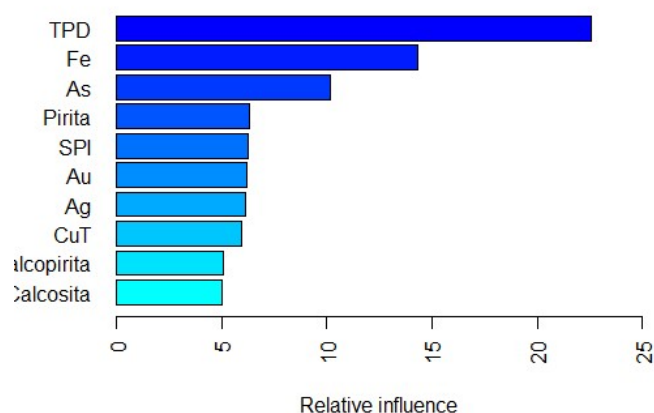
**Tabla 16: Hiperparámetros óptimos para modelos GBM.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**



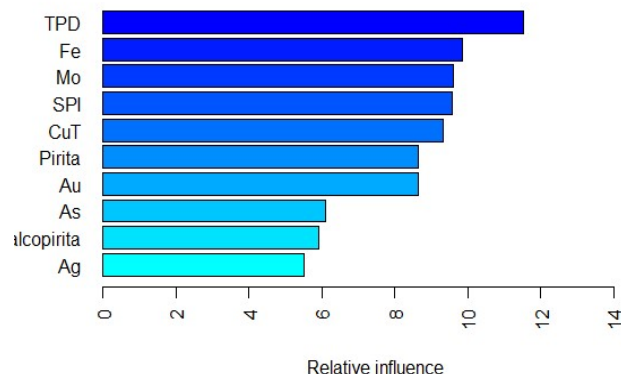
## V.IV. II Resultados teóricos a nivel operacional

Los árboles de decisión incorporan en su algoritmo la selección de variables, atributo que se extiende a los modelos de ensamble, particularmente al modelo utilizado, es decir, GBM (Gradient Boosting Trees). Una vez desarrollados los modelos como se detalló anteriormente, se tienen los siguientes resultados en relación a la importancia relativa de las variables en estudio:

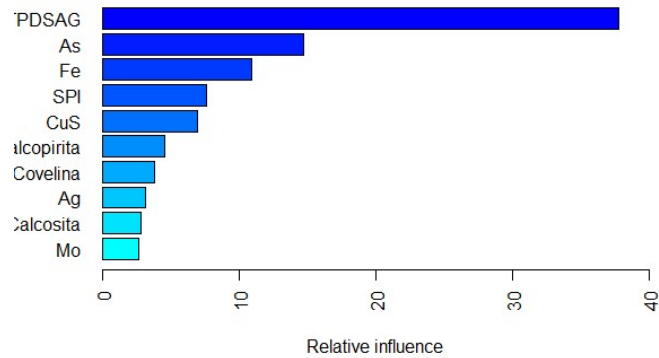
### Molinos SAG:



**Figura 7: Influencia relativa de variables en LS1.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**



**Figura 8: Influencia relativa de variables en LS2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

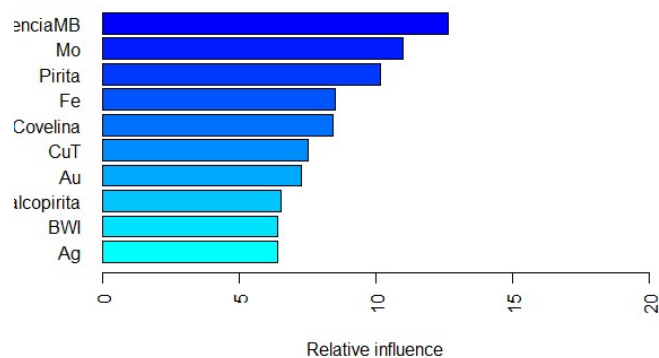


**Figura 9: Influencia relativa de variables en LC.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Se puede observar que la variable que explica la mayor parte de la variación del consumo de bolas 5,5" es las toneladas cargadas. Esto tiene sentido ya que el tonelaje diario es un actual input y tiene una relación lineal con el carguío diario de bolas. Luego aparece el grado férrico (Fe) con una importancia relativa promedio al 15%, lo cual coincide con hipótesis obtenidas a lo largo del trabajo por la contraparte operacional, donde se estipulaba que al sobrepasar el 2-2,5% de grado férrico, aumenta considerablemente el consumo de bolas. También aparece el índice de SAG con una importancia relativa promedio del 10%, lo cual demuestra que la dureza de la roca impacta en un consumo y fractura mayor de bolas.

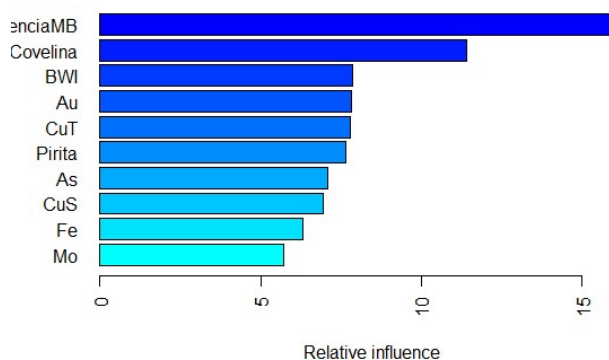
Finalmente, aparece una influencia perceptible de elementos menores en la alimentación a los molinos SAG, donde la cantidad de arsénico y molibdeno tienen una relación directa e indirecta en el consumo, respectivamente. El resto de los modelos se explica por la mineralogía de alimentación, la cual varía dependiendo de la concentradora dada la fuente de alimentación del mineral que estas tienen por diferentes sectores mineros asociados a los rajos de Minera Escondida.

### **Molinos de Bolas:**



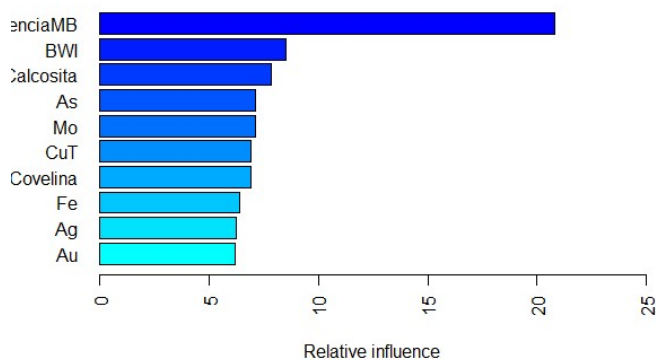
**Figura 10: Influencia relativa de variables en LS1.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**



**Figura 11: Influencia relativa de variables en LS2.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**



**Figura 12: Influencia relativa de variables en LC.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**

Se puede observar que la variable que mayor explica el consumo de bolas 3" es la potencia total consumida el día anterior. Al igual que en el caso de las bolas 5,5" tiene sentido, ya que este es un input actual en el proceso de carguío diario. Luego aparece el índice de Bond, el cual tiene una relación directa dado que explica la resistencia del mineral en kilowatts por tonelada necesarios para reducir el mineral bajo los 100  $\mu m$ .

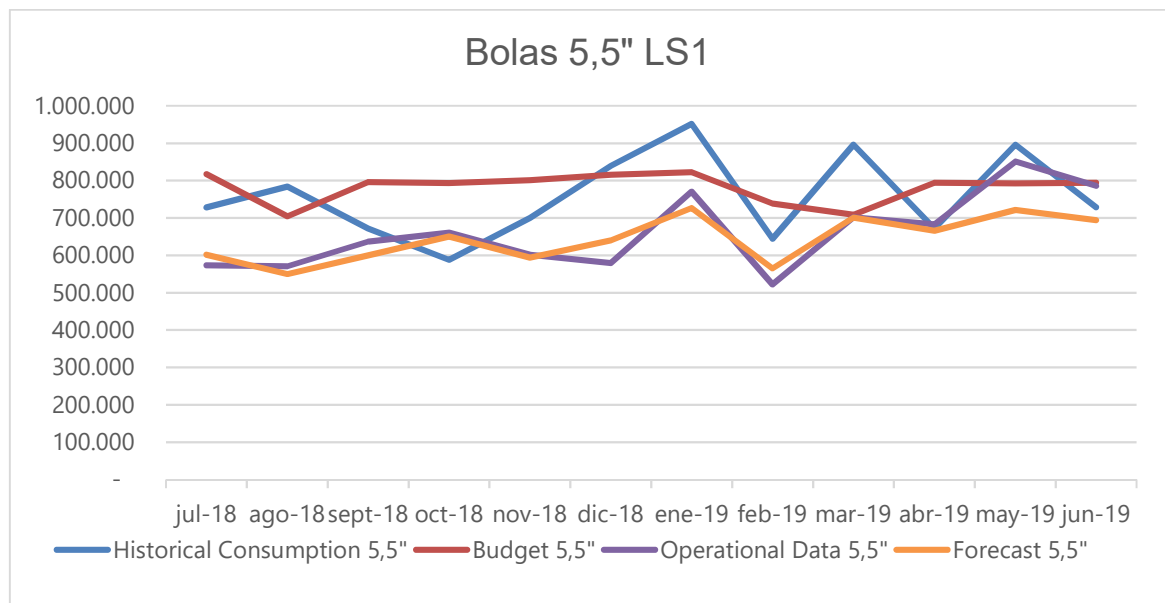
Finalmente, la presencia de elementos menores no tiene una tendencia clara, solo el molibdeno logra superar el 10% en Laguna Seca 1. El resto de los modelos se explica por la caracterización mineralógica, la cual tampoco tiene una tendencia clara, lo cual se puede atribuir nuevamente a la diferencia en la fuente de alimentación de las concentradoras.

## V.IV. III Resultados teóricos a nivel comercial

Tal como se explicó anteriormente, el forecast se realiza por años fiscales, los cuales comprenden para el año fiscal X, desde julio de (X-1) hasta junio de X. Este forecast utiliza tasas de desgaste históricas en gramos por toneladas para ambos molinos, por lo que el accuracy del forecast no es de los mejores y deja una oportunidad de mejora.

Entonces, para analizar el desempeño teórico de los modelos, estos se testearán en el último año fiscal (FY19), entrenando con toda la data histórica disponible anteriormente a esta fecha. La métrica para medir el accuracy será (1-MAPE), la cual se calculará para el Budget, Historical Consumption, Forecast; los que corresponden al presupuesto por parte de finanzas para el insumo estratégico, lo que realmente se compró por parte del área comercial y finalmente el forecast realizado con los modelos ya presentados.

### Molinos SAG:

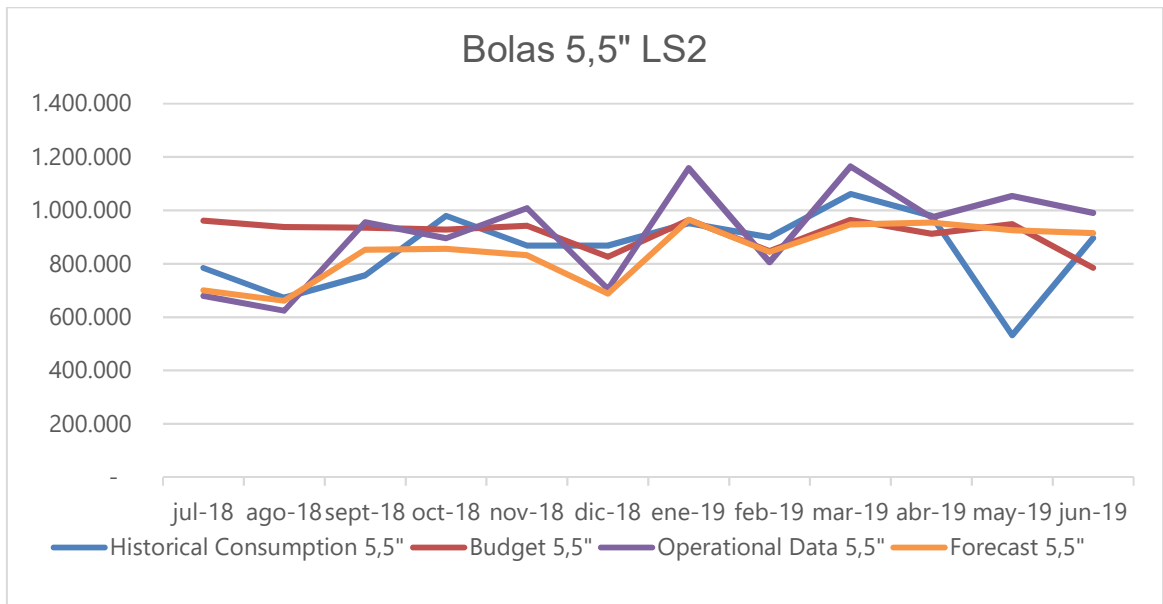


**Figura 13: Forecast año fiscal 2019 en Concentradora Laguna Seca 1.**  
Fuente: Elaborado por el autor en base a datos de la empresa.

Donde el accuracy para bolas 5,5" en Laguna Seca 1 es:

<b>Molinos SAG LS1</b>	<b>1-MAPE</b>
Budget	78,44%
Historical Consumption	80,73%
Forecast	94,05%

**Tabla 17: Accuracy bolas 5,5" en Laguna Seca 1.**  
Fuente: Elaborado por el autor en base a datos de la empresa.

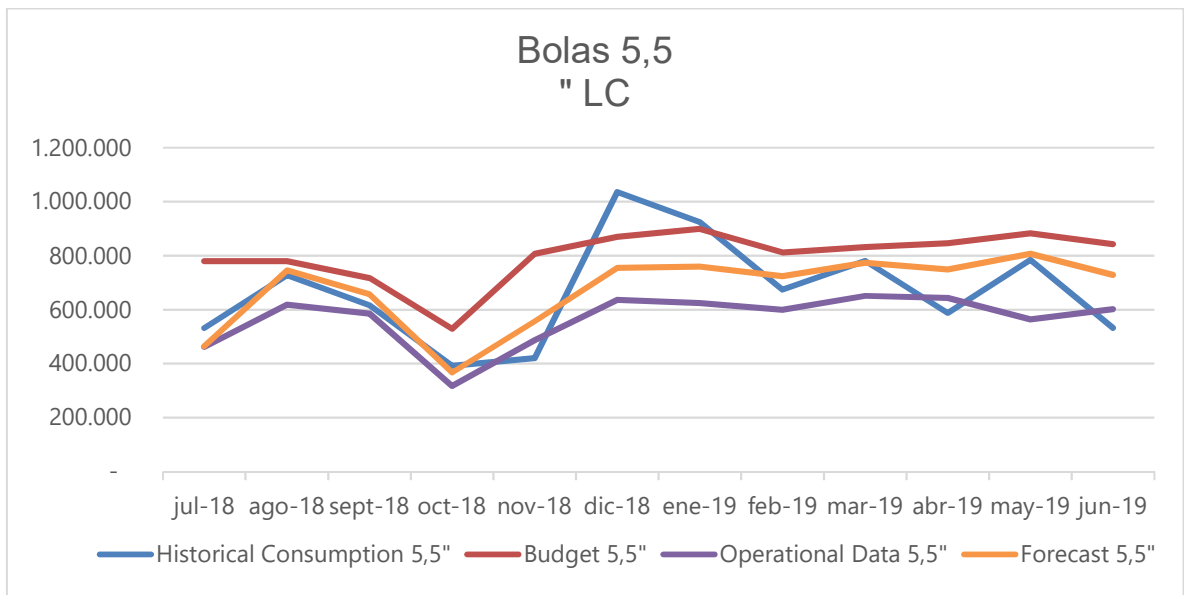


**Figura 14: Forecast para año fiscal 2019 en Concentradora Laguna Seca 2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 5,5" en Laguna Seca 2 es:

Molinos SAG LS2	1-MAPE
Budget	83,59%
Historical Consumption	84,31%
Forecast	91,20%

**Tabla 18: Accuracy bolas 5,5" en Laguna Seca 2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa**



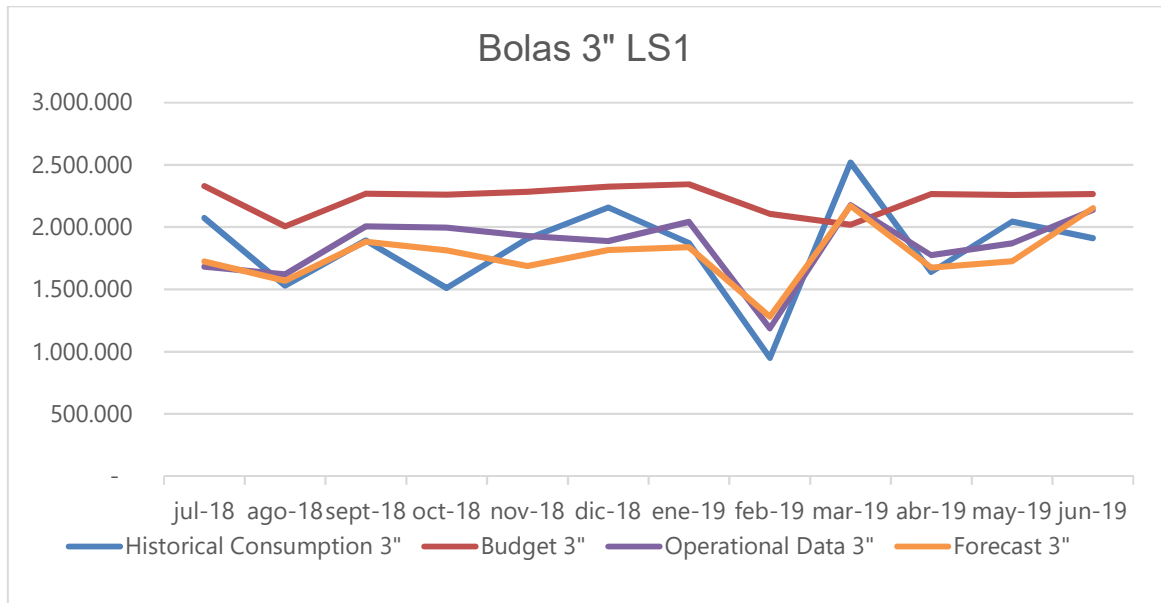
**Figura 15: Forecast para año fiscal 2019 en Concentradora Los Colorados.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 5,5" en Los Colorados es:

<b>Molinos SAG LC</b>	<b>1-MAPE</b>
Budget	56,46%
Historical Consumption	76,8%
Forecast	81,29%

**Tabla 19: Accuracy bolas 5,5" en Los Colorados.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Molinos de Bolas:

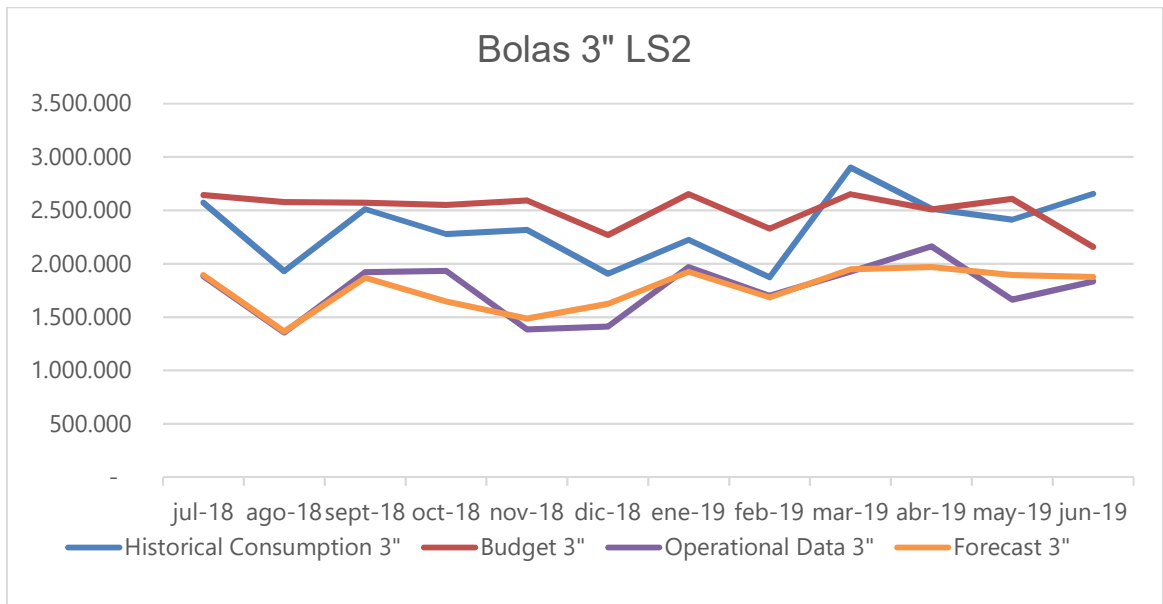


**Figura 16: Forecast para año fiscal 2019 en Concetradora Laguna Seca 1.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 3" en Laguna Seca 1 es:

<b>Molinos de Bolas LS1</b>	<b>1-MAPE</b>
Budget	76,32%
Historical Consumption	87,84%
Forecast	94,23%

**Tabla 20: Accuracy bolas 3" en Laguna Seca 1.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

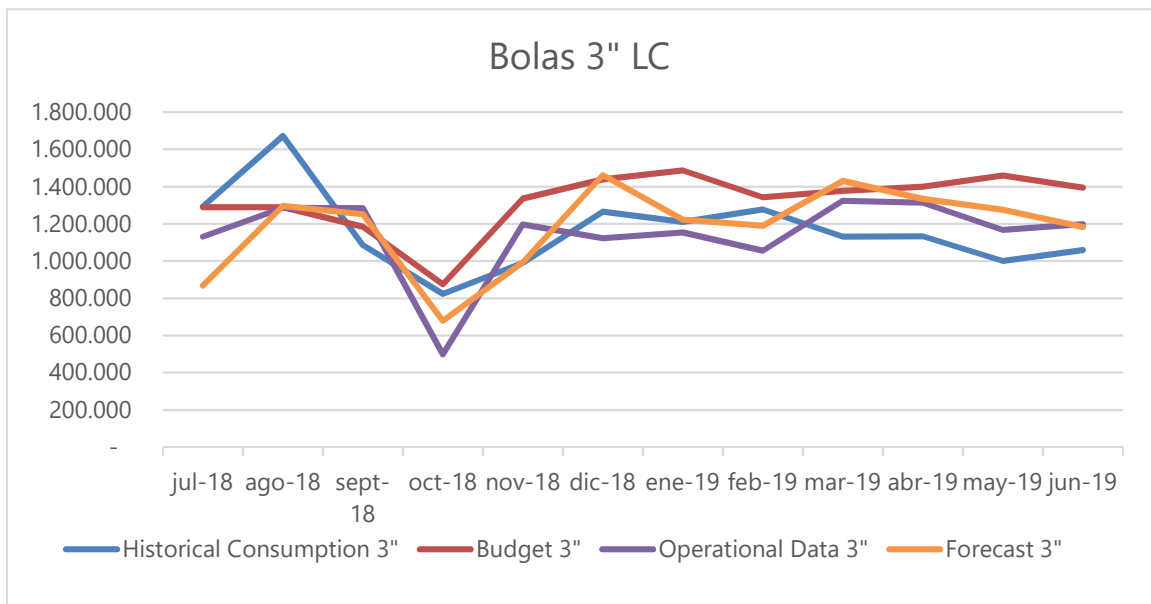


**Figura 17: Forecast para año fiscal 2019 en Concentradora Laguna Seca 2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 3" en Laguna Seca 2 es:

Molinos de Bolas LS2	1-MAPE
Budget	54,58%
Historical Consumption	65,8%
Forecast	94,07%

**Tabla 21: Accuracy bolas 3" en Laguna Seca 2.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**



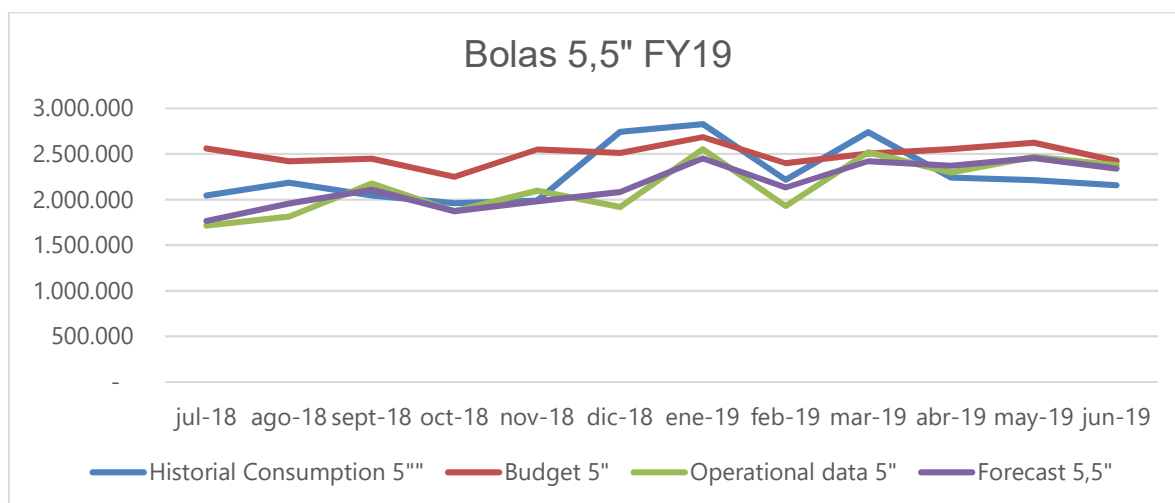
**Figura 18: Forecast para año fiscal 2019 en Concentradora Los Colorados.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 3" en Los Colorados es:

Molinos de Bolas Los Colorados	1-MAPE
Budget	79,52%
Historical Consumption	80,39%
Forecast	87,57%

**Tabla 22: Accuracy bolas 3" en Los Colorados.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Finalmente, si se realiza el análisis a nivel agregado para bolas 3" y 5,5" con el fin de obtener una visión macro del insumo, los resultados son los siguientes:



**Figura 19: Forecast para año fiscal 2019 de bolas 5,5" en Minera Escondida.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 5,5" en el año fiscal 2019 es:

Molinos SAG Escondida	Minera	Consumo Bolas 5,5" [tons]	1-MAPE
		Budget	81,91%
		Historical Consumption	87,05%
		Operational Data	-
		Forecast	95,69%

**Tabla 23: Accuracy bolas 5,5" en Minera Escondida.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Luego, podemos calcular el ahorro en términos de budget mal asignado para insumos (Budget – Forecast) y ahorro real (Historical Consumption – Forecast):

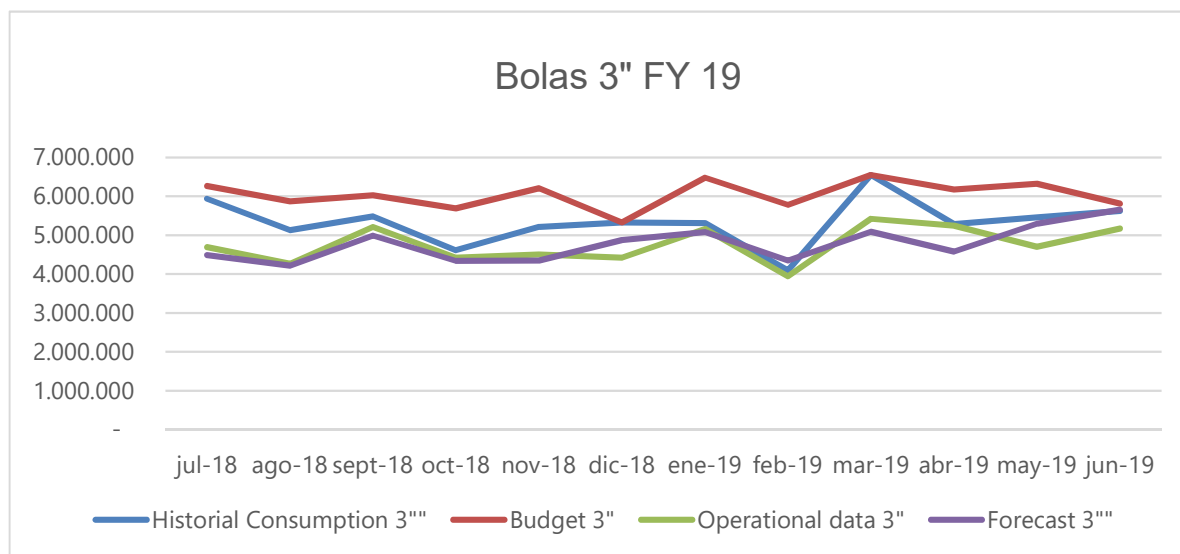
Ahorro	Tons	Money (USD) <sup>4</sup>
Budget – Forecast	4.221	\$ 4.931.495
Bought - Forecast	1.415	\$ 1.653.620

**Tabla 24: Ahorro de bolas 5,5" en Minera Escondida Limitada.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

<sup>4</sup> Precios basados en FY19 Budget.



Ahora, el análisis agregando para molinos de bolas:



**Figura 20: Forecast para año fiscal 2019 de bolas 3" en Minera Escondida Limitada.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Donde el accuracy para bolas 3" en el año fiscal 2019 es:

Molinos de Bolas Minera Escondida	Consumo Bolas 3" [tons]	1-MAPE
Budget	72.519	72,39%
Historical Consumption	64.037	87,85%
Operational Data	57.173	-
Forecast	57.304	95,37%

**Tabla 25: Accuracy bolas 3" en Minera Escondida Limitada.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Luego, podemos calcular el ahorro en términos de presupuesto mal asignado para insumos (Budget – Forecast) y ahorro real (Historical Consumption – Forecast):

Ahorro	Tons	Money (USD) <sup>5</sup>
Budget – Forecast	15.215	\$ 16.001.511
Bought - Forecast	6.733	\$ 7.081.468

**Tabla 26: Ahorro de bolas 5,5" en Minera Escondida Limitada.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

Entonces, para el año fiscal 2019, de haber utilizado los modelos con técnicas de data mining ya presentados, se podría haber reducido el presupuesto de bolas de molienda 3"

<sup>5</sup> Precios basados en FY19 Budget.

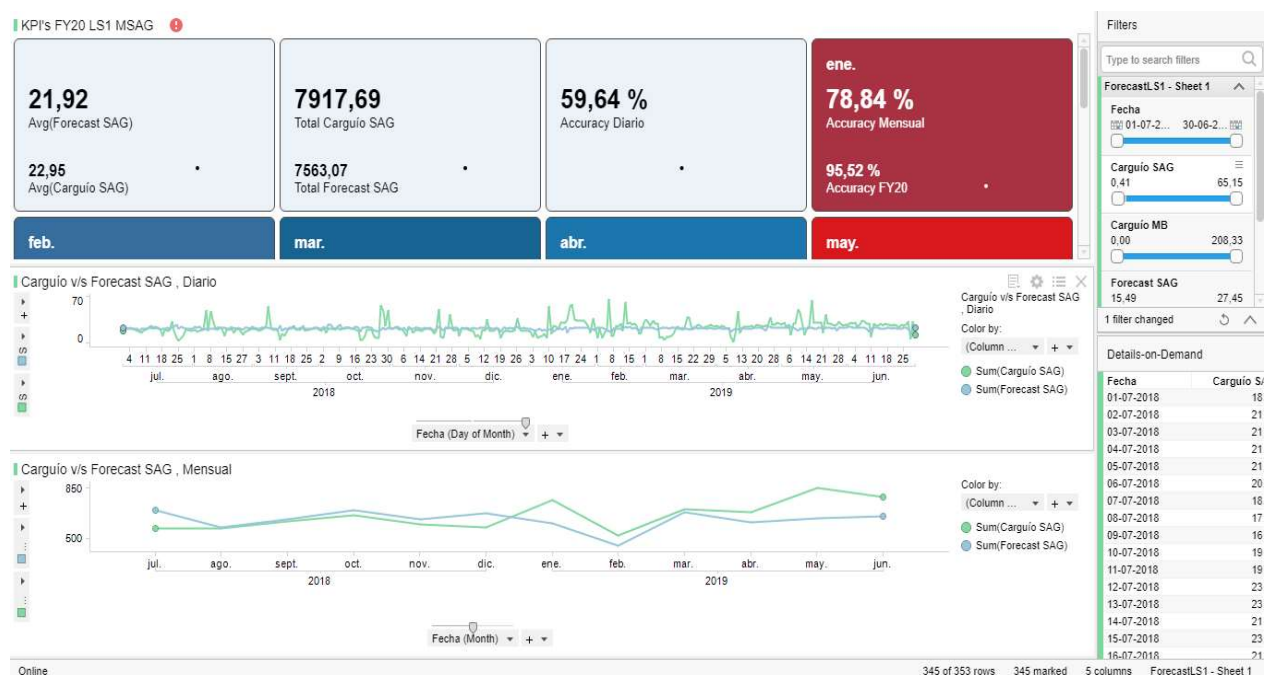
y 5,5” en \$23.082.978 USD, los cuales podrían haber sido destinados a otro insumo en algún activo de la compañía. Esto, a su vez, es un ahorro efectivo de \$ 8.735.088 USD, producto del aumento considerable en la precisión del forecast acerca de la cantidad de bolas a utilizar en Minera Escondida.

## V.V Dashboard para herramienta predictiva

Una etapa necesaria para pasar del modelo teórico a una herramienta útil para el usuario final, es diseñar una interfaz que permita revisar los resultados del modelo predictivo en las respectivas rutinas de planificación.

De esta forma, se construye un dashboard mediante el software TIBCO Spotfire dado que es el software oficial en BHP para visualización y reportabilidad de data, lo que implica que será útil y amigable para el cliente final, en este caso el equipo de planificación y control de procesos.

El dashboard cuenta de 3 áreas: KPI's para el año fiscal 20, gráfico de visualización del carguío real v/s forecast a nivel diario y gráfico de visualización carguío real v/s forecast a nivel mensual. Junto a esto, en la parte derecha se cuenta con una sección de filtros que permite filtrar las variables utilizadas en el dashboard (fecha, carguío real SAG, carguío real molino de bolas, forecast SAG, forecast molino de bolas) y una sección de Details-on-Demand que permite visualizar el valor de las variables por fecha.

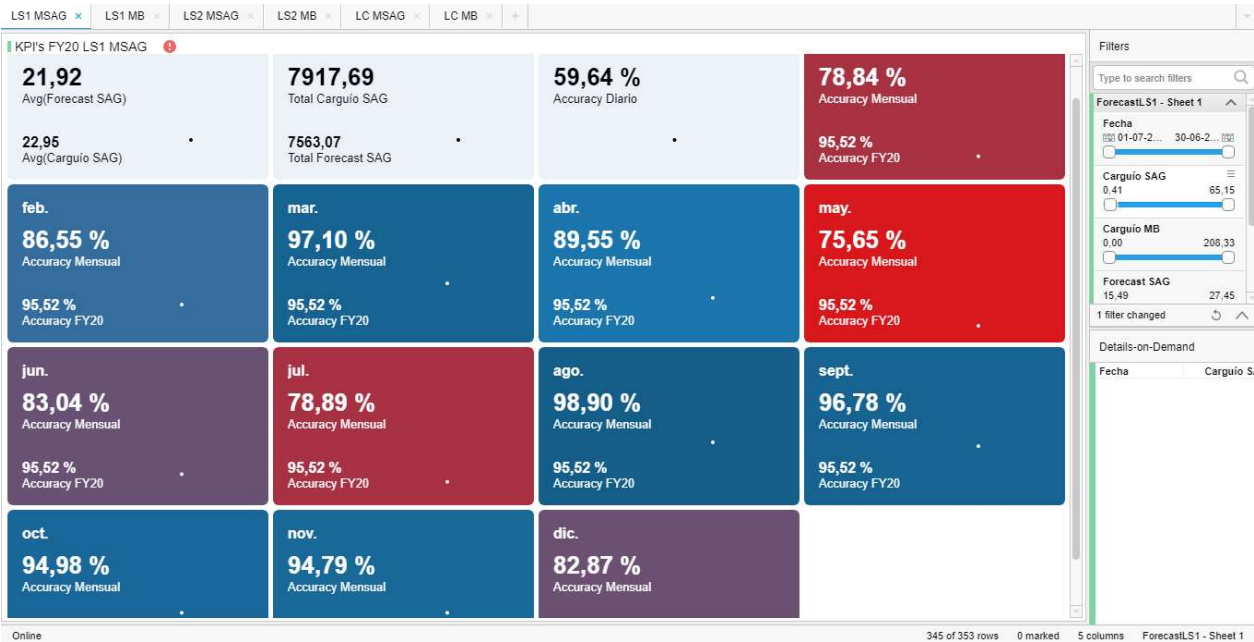


**Figura 21: Dashboard para la visualización de la herramienta predictiva.**

**Fuente: Elaborado por el autor en base a datos de la empresa.**

El recuadro de KPI's para el año fiscal 20 cuenta con el cálculo de promedios y totales de carguío y forecast, en conjunto a los accuracies a nivel diario, mensual y anual con el

objetivo de tener una visualización agregada y desagregada de bolas de molienda. Este se puede observar en detalle en la siguiente figura:



**Figura 22: Sección KPI's FY20 – Dashboard.**  
**Fuente: Elaborado por el autor en base a datos de la empresa.**

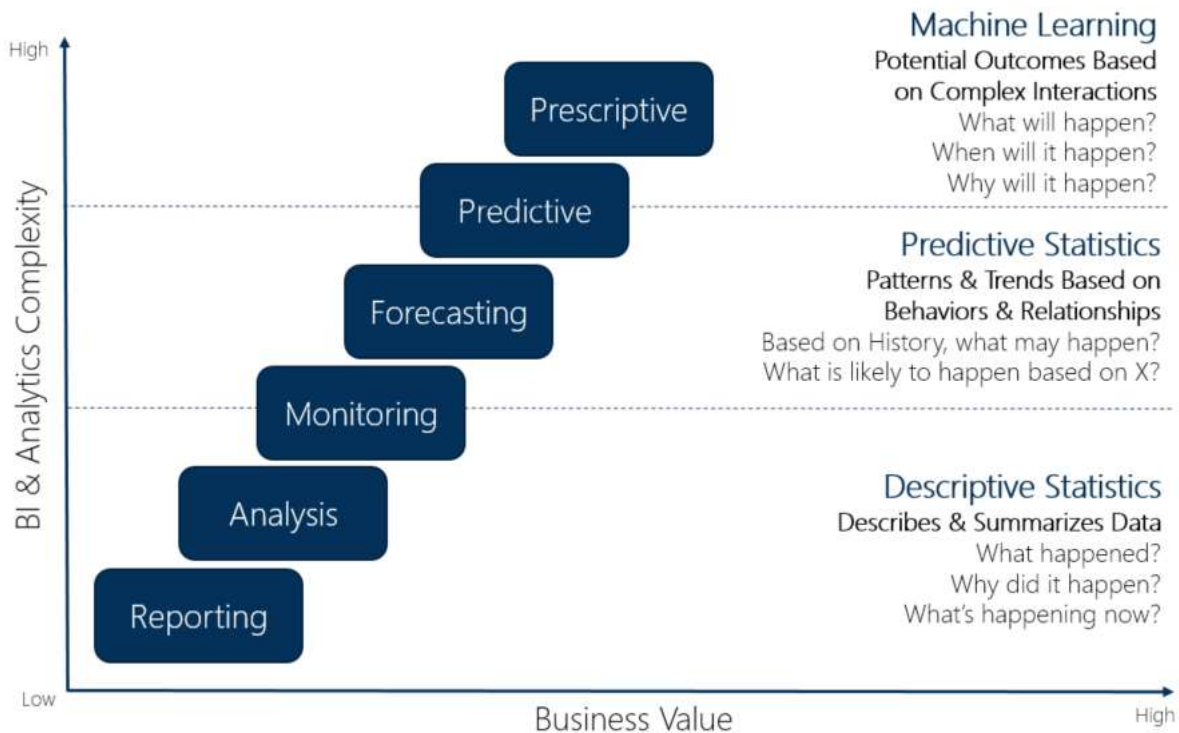
Finalmente, cabe destacar que el dashboard tiene 6 pestañas (LS1 MSAG, LS1 MB, LS2 SAG, LS2 MB, LC SAG, LC MB) cada una con la visualización del respectivo forecast tanto de bolas 5,5” y 3” para cada Concentradora de Minera Escondida.

### V.VI Visión Estratégica

Dentro de la realización del proyecto de título, se observó como las herramientas de analytics pueden agregar valor al negocio mediante la automatización de procesos de análisis y el aumento en la eficiencia dada la utilización de los datos disponibles.

En otras áreas de la empresa, tales como Enterprise & Improvement, se utilizan herramientas analíticas de tipo descriptivas, de diagnóstico y de monitoreo para detectar, entender en base a datos y finalmente abordar los problemas que surgen en el área operacional.

Dado esto, existe la oportunidad dentro del área comercial de seguir adoptando nuevas tecnologías de analytics y data science con el fin de agregar valor al negocio, ya sea mediante la optimización de procesos o reducción de costos, realizando planificaciones de demanda de excelencia, convirtiendo a los analistas de data science en partners del negocio y, por ende, dedicando más tiempo a la gestión que al análisis del mismo.



**Figura 23: Evolución de Analytics. Valor del negocio v/s complejidad.**  
**Fuente: Obtenido de [18].**

La figura anterior muestra la evolución del desarrollo de data science, relacionando el valor del negocio con la complejidad de la inteligencia de negocios y analytics. Entonces, se espera que la visión estratégica de la implementación de modelos de machine learning para la predicción de demanda futura de bolas de molienda, sirva para mostrar en el área comercial el potencial de invertir en el desarrollo de data science, el cual puede ser replicable a otros insumos estratégicos como bolas 1,25", Cal, Floculante, NaHS, Colectores, entre otros ; aumentando la precisión de los forecast mediante la automatización de este proceso , optimizando el uso del tiempo para invertirlo en gestión y toma de decisiones más que en generación de reportes y trabajo manual .

## **VI. CONCLUSIÓN**

### **VI.I Cumplimiento de objetivos propuestos**

En relación a los objetivos específicos propuestos al comienzo del trabajo de título, ya terminado el desarrollo del proyecto el estado de cada uno es el siguiente:

Se analizó el proceso de molienda que se lleva a cabo en las concentradoras de Minera Escondida, específicamente como se realiza el carguío diario de bolas, entendiendo cuales son los inputs que se utilizan y que factores pueden introducir variabilidad en este proceso.

Luego, en conjunto al conocimiento experto de distintos profesionales del área operacional y comercial de BHP se logró incorporar nuevas variables para predecir el consumo de bolas de molienda, tales como el índice de SAG y Bond; la mineralogía de alimentación y elementos menores presentes en esta.

Se gestionó distintas bases de datos, dando como resultado un Data Warehouse compuesto de data relativa al carguío de bolas de molienda, el cual fue utilizado como input para realizar un challenge de modelos de machine learning, siendo el que entregó mejores resultados independiente de la concentradora y tipo de molino; Gradient Boosting Trees, seleccionado de acuerdo al mínimo RMSE y mejor interpretabilidad de los resultados.

Los modelos recién mencionados se utilizaron para predecir la demanda del año fiscal 20, donde se alcanzó un accuracy de 95,69% para bolas 5,5" y 95,37% para bolas 3", lo que se traduce en una reducción de presupuesto de \$23.082.978 USD, los cuales podrían haber sido destinados a otro insumo estratégico en Minera Escondida u otro activo de la compañía. Este aumento en el accuracy en la predicción de demanda de bolas de molienda es un ahorro efectivo de \$ 8.735.088 USD, entre lo que se compró y lo que realmente se usó en los molinos.

Finalmente, se utilizó los resultados de la herramienta predictiva para desarrollar un dashboard que permite al área operacional visualizar la predicción del carguío a nivel diario y al área comercial a nivel mensual y por año fiscal, permitiendo hacer un seguimiento del aumento en la eficiencia de la predicción de este insumo estratégico.

## **VI.II Resultados Observados**

Considerando que las bolas de molienda corresponden a un insumo estratégico vital dentro del proceso de molienda y por ende de producción de concentrado de cobre, es importante predecir la demanda futura de estas de la manera más exacta posible, ya que tener un forecast de demanda robusto y confiable permite facilitar las conversaciones de valor entre la operación y área comercial, alineados con la cobertura, riesgos y oportunidades comerciales, es decir, gestionar proactivamente la compra y entrega de este insumo y también negociar con mayor certeza los volúmenes (y por ende mejores precios) con los proveedores. Los modelos desarrollados en este trabajo de título permiten lograr este objetivo, dado que el accuracy más bajo en el testeado realizado para el año fiscal 2019 fue de 81,29% en bolas 5,5" de Los Colorados y el más alto de 94,23% en bolas 3" de Laguna Seca 1, obteniendo en conjunto un 95,69% y 95,37% para bolas 5,5" y 3", respectivamente.

Es importante considerar también quien es el cliente final del proyecto de minería de datos y diseñar la solución en conjunto a este, considerando de qué forma será utilizada la herramienta, ya que los requerimientos y necesidades de la parte operacional y comercial son distintos.

En fin, la gestión del cambio es un proceso fundamental en la incorporación de nuevas tecnologías. En BHP se está trabajando para transformar las relaciones basadas en confianza y colaboración, para tomar decisiones de valor en base a data, lo cual considera incorporar nuevas capacidades en la organización para poder soportar el cambio tecnológico.

## **VI.III Recomendaciones a Futuro**

Quedan propuestos como trabajo a futuro los siguientes puntos, los cuales están fuera del alcance de esta memoria, pero permitirían agregar más valor a la herramienta y al negocio:

- Incorporar en el modelo variables relativas al rajo del que proviene el mineral, para anticipar características mineralógicas.
- Desarrollar modelos predictivos para el consumo de bolas por molino, incorporando variables operacionales de estos, para analizar si hay otras características que se puedan agregar a nivel macro e incorporar a los modelos expuestos en este trabajo de título.
- Realizar la debida gestión del cambio, generando las nuevas capacidades necesarias para utilizar y continuar desarrollando modelos de data science más

complejos y que permitan automatizar el análisis en otros insumos estratégicos para cualquier activo de la compañía.

- Utilizar los modelos y data disponible para realizar el forecast del año fiscal 2021, testeando la eficacia y eficiencia de los modelos.

## VI. BIBLIOGRAFÍA

- [1] DATOS Y RECURSOS SOBRE Producción minera nacional por empresa y pasta, desde 1998 al 2017 [en línea] <<http://datos.gob.cl/dataset/produccion-minera-nacional-por-empresa>> [consulta: 05 enero 2020].
- [2] SERVICIO NACIONAL DE GEOLOGÍA Y MINERÍA 2018. Anuario de la Minería de Chile [en línea] < [https://www.sernageomin.cl/wp-content/uploads/2019/06/Libro\\_Anuario\\_2018\\_.pdf](https://www.sernageomin.cl/wp-content/uploads/2019/06/Libro_Anuario_2018_.pdf)> [consulta: 05 enero 2020] .
- [3] BHP 2017. Informe de Sustentabilidad [en línea] < [https://www.bhp.com/-/media/documents/media/reports-and-presentations/2018/180720\\_informedesustentabilidadbhpchile.pdf?>](https://www.bhp.com/-/media/documents/media/reports-and-presentations/2018/180720_informedesustentabilidadbhpchile.pdf?>) [consulta: 05 enero 2020].
- [4] COCHILCO 2018. Análisis del mercado de insumos críticos en la minería del cobre [en línea] < <https://www.cochilco.cl/Mercado%20de%20Metales/Analisis%20Mercado%20de%20los%20Insumos%20Cr%C3%ADticos%202019vf2.pdf>> [consulta: 05 enero 2020].
- [5] HAN, J., KAMBER, M., & PEI, J. (2011). Data mining: concepts and techniques: concepts and techniques.
- [6] AZEVEDO, A. I. R. L. (2008). KDD, SEMMA and CRISP-DM: a parallel overview.
- [7] KIMBALL “et al”. John Wiley and Sons Building the data warehouse (2da edición). New York, 1996.
- [8] KIMBALL “et al”. John Wiley and Sons The data warehouse lifecycle toolkit: expert methods for designing, developing and deploying data warehouses. , New York, USA. 1998.
- [9] VASSILIADIS, P., SIMITSIS, A. y SKIADOPOULOS, S. On the logical modeling of ETL processes. In Proc. 14th Int. Conf. on Advanced Information Systems Engineering, London, UK. pp. 782-786.2002.
- [10] SMART VISION EUROPE. The CRISP-DM process model. 1999. [en línea] <http://crisp-dm.eu/> [consulta : 05 enero 2020].
- [11] BREIMAN, L. Bagging predictors. Machine Learning forthcoming, 1996.
- [12] BREIMAN, L. 2001. Random Forests. Mach. Learn. 45, 1 (October 2001).
- [13] FRIEDMAN, J. H. Greedy Function Approximation: A Gradient Boosting Machine. Stanford, 2010.
- [14] CHAPMAN “et al”. 2000. CRISP-DM 1.0: Step-by-step Data Mining Guide.
- [15] MUÑOZ, A. 2019. Producción de Cobre. En: REUNIÓN MENSUAL: planificación de control de procesos: 3 de marzo de 2019. Antofagasta. BHP.
- [16] BECHTEL CHILE LTDA [2012] . Criterio de diseño para filosofía de operación – OGP1. 71p.
- [17] RIDGEWAY, G. Generalized Boosted Models: A guide to the gbm package, 2005.
- [18] GITHUB. Model training and tuning. [en línea] <https://topepo.github.io/caret/model-training-and-tuning.html> [consulta : 10 mayo 2020].
- [19] ARBELA. Advance your Analytics Journey [en línea] <https://www.arbelatech.com/services/bi-analytics.html> [consulta : 10 enero 2020].



## VII. ANEXOS

### VII.I Código herramienta predictiva en R

#### *#Load Librarys*

```
library(readxl)
library(tidyverse)
library(ggplot2)
library(stargazer)
library(plm)
library(lmtest)
library(plyr)
library(car)
library(CORElearn)
library(rpart)
library(forecast)
library(rpart.plot)
library(DMwR)
library(Metrics)
library(performanceEstimation)
library(caret)
library(lubridate)
library(outliers)
library(MLmetrics)
library(openxlsx)
```

#### *#Directory*

```
setwd("C:/Users/NAVAJI9/Desktop/BHP/R- Budget")
```

#### *#DATA LAGUNA SECA 1*

```
CARGUIO_BOLAS_L1 <- read_excel("CARGUIO BOLAS L1-L2 27-01-2020.xlsx", sheet =
"L1", col_types = c("date", "numeric", "numeric", "numeric", "numeric", "numeric", "
numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "text"), col_nam
es = TRUE, cell_limits(c(2,2),c(NA,22)))
```

```
names(CARGUIO_BOLAS_L1)[2] <- "TPD"
names(CARGUIO_BOLAS_L1)[3] <- "BWI"
names(CARGUIO_BOLAS_L1)[4] <- "SPI"
names(CARGUIO_BOLAS_L1)[11] <- "MSAG"
names(CARGUIO_BOLAS_L1)[12] <- "MB01"
names(CARGUIO_BOLAS_L1)[13] <- "MB02"
names(CARGUIO_BOLAS_L1)[14] <- "MB03"
names(CARGUIO_BOLAS_L1)[15] <- "MB04"
```

```
CARGUIO_BOLAS_L1 <- CARGUIO_BOLAS_L1[, c(1,2,3,4,11,12,13,14,15)]
```

```
CARGUIO_BOLAS_L1[["MB01"]][is.na(CARGUIO_BOLAS_L1[["MB01"]])] <- 0
```

```

CARGUIO_BOLAS_L1[["MB02"]][is.na(CARGUIO_BOLAS_L1[["MB02"]])] <- 0
CARGUIO_BOLAS_L1[["MB03"]][is.na(CARGUIO_BOLAS_L1[["MB03"]])] <- 0
CARGUIO_BOLAS_L1[["MB04"]][is.na(CARGUIO_BOLAS_L1[["MB04"]])] <- 0
CARGUIO_BOLAS_L1[["MSAG"]][is.na(CARGUIO_BOLAS_L1[["MSAG"]])] <- 0

attach(CARGUIO_BOLAS_L1)
CARGUIO_BOLAS_L1 <- cbind(CARGUIO_BOLAS_L1,MB01+MB02+MB03+MB04)
CARGUIO_BOLAS_L1$Concentradora="LS1"
names(CARGUIO_BOLAS_L1)[10] <- "MB"
detach(CARGUIO_BOLAS_L1)

#DATA LAGUNA SECA 2

CARGUIO_BOLAS_L2 <- read_excel("CARGUIO BOLAS L1-L2 27-01-2020.xlsx", sheet =
"L2",col_types = c("date","numeric","numeric","numeric","numeric","numeric","
numeric","numeric","numeric","numeric","numeric","numeric","numeric","nume
ric","numeric","numeric","numeric","numeric","numeric","text"),col_nam
es = TRUE, cell_limits(c(2,2),c(NA,22)))
names(CARGUIO_BOLAS_L2)[2] <- "TPD"
names(CARGUIO_BOLAS_L2)[3] <- "BWI"
names(CARGUIO_BOLAS_L2)[4] <- "SPI"
names(CARGUIO_BOLAS_L2)[10] <- "MSAG"
names(CARGUIO_BOLAS_L2)[11] <- "MB05"
names(CARGUIO_BOLAS_L2)[12] <- "MB06"
names(CARGUIO_BOLAS_L2)[13] <- "MB07"
names(CARGUIO_BOLAS_L2)[14] <- "MB08"
CARGUIO_BOLAS_L2<-CARGUIO_BOLAS_L2[,c(1,2,3,4,10,11,12,13,14)]
CARGUIO_BOLAS_L2[["MB05"]][is.na(CARGUIO_BOLAS_L2[["MB05"]])] <- 0
CARGUIO_BOLAS_L2[["MB06"]][is.na(CARGUIO_BOLAS_L2[["MB06"]])] <- 0
CARGUIO_BOLAS_L2[["MB07"]][is.na(CARGUIO_BOLAS_L2[["MB07"]])] <- 0
CARGUIO_BOLAS_L2[["MB08"]][is.na(CARGUIO_BOLAS_L2[["MB08"]])] <- 0
CARGUIO_BOLAS_L2[["MSAG"]][is.na(CARGUIO_BOLAS_L2[["MSAG"]])] <- 0

attach(CARGUIO_BOLAS_L2)
CARGUIO_BOLAS_L2 <- cbind(CARGUIO_BOLAS_L2,MB05+MB06+MB07+MB08)
CARGUIO_BOLAS_L2$Concentradora="LS2"
names(CARGUIO_BOLAS_L2)[10] <- "MB"
detach(CARGUIO_BOLAS_L2)

#DATA LOS COLORADOS

CARGUIO_BOLAS_LC1 <- read_excel("CARGUIO BOLAS LC 2017-2019.xlsx", sheet = "
Læ"%ã¹ænea 1 CLC",col_types = c("date","numeric","numeric","numeric","numeric
","numeric","numeric","numeric","numeric","numeric","numeric","numeric","n
umeric","numeric","numeric","numeric","numeric","numeric","numeric","nume
ric","numeric","text"),col_names = TRUE, cell_limits(c(2,2),c(NA,24)))
CARGUIO_BOLAS_LC1 <- CARGUIO_BOLAS_LC1[,c(1,2,3,4,11,12,13,20,21,22)]
names(CARGUIO_BOLAS_LC1)[5] <- "MSAG"
names(CARGUIO_BOLAS_LC1)[6] <- "MB-1"
names(CARGUIO_BOLAS_LC1)[7] <- "MB-2"
names(CARGUIO_BOLAS_LC1)[2] <- "TPDSAG"
CARGUIO_BOLAS_LC1[["MB-1"]][is.na(CARGUIO_BOLAS_LC1[["MB-1"]])] <- 0

```

```

CARGUIO_BOLAS_LC1[["MB-2"]][is.na(CARGUIO_BOLAS_LC1[["MB-2"]])] <- 0
CARGUIO_BOLAS_LC1[["MSAG"]][is.na(CARGUIO_BOLAS_LC1[["MSAG"]])] <- 0
CARGUIO_BOLAS_LC1[["TPDSAG"]][is.na(CARGUIO_BOLAS_LC1[["TPDSAG"]])] <- 0

CARGUIO_BOLAS_LC1 <- cbind(CARGUIO_BOLAS_LC1,CARGUIO_BOLAS_LC1$`MB-1`+ CARGUI
O_BOLAS_LC1$`MB-2`)
names(CARGUIO_BOLAS_LC1)[11] <- "MB"
names(CARGUIO_BOLAS_LC1)[3] <- "TPDMB1"
names(CARGUIO_BOLAS_LC1)[4] <- "TPDMB2"
CARGUIO_BOLAS_LC1[["TPDMB1"]][is.na(CARGUIO_BOLAS_LC1[["TPDMB1"]])] <- 0
CARGUIO_BOLAS_LC1[["TPDMB2"]][is.na(CARGUIO_BOLAS_LC1[["TPDMB2"]])] <- 0
CARGUIO_BOLAS_LC1 <- cbind(CARGUIO_BOLAS_LC1,CARGUIO_BOLAS_LC1$TPDMB1+CARGUIO
_BOLAS_LC1$TPDMB2)
names(CARGUIO_BOLAS_LC1)[12] <- "TPDMB"
CARGUIO_BOLAS_LC1$Concentradora="LC1"
CARGUIO_BOLAS_LC2 <- read_excel("CARGUIO BOLAS LC 2017-2019.xlsx", sheet = "
Læ"%ã¹ænea 2 CLC",col_types = c("date","numeric","numeric","numeric","numeric
","numeric","numeric","numeric","numeric","numeric","numeric","numeric","nume
ric","numeric","numeric","numeric","numeric","numeric","numeric","text"),col_
names = TRUE, cell_limits(c(3,2),c(NA,21)))
CARGUIO_BOLAS_LC2 <- CARGUIO_BOLAS_LC2[,c(1,2,3,4,11,12,13,17,18,19)]
names(CARGUIO_BOLAS_LC2)[5] <- "MSAG"
names(CARGUIO_BOLAS_LC2)[6] <- "MB-3"
names(CARGUIO_BOLAS_LC2)[7] <- "MB-4"
names(CARGUIO_BOLAS_LC2)[2] <- "TPDSAG"
CARGUIO_BOLAS_LC2[["MSAG"]][is.na(CARGUIO_BOLAS_LC2[["MSAG"]])] <- 0
CARGUIO_BOLAS_LC2[["MB-3"]][is.na(CARGUIO_BOLAS_LC2[["MB-3"]])] <- 0
CARGUIO_BOLAS_LC2[["MB-4"]][is.na(CARGUIO_BOLAS_LC2[["MB-4"]])] <- 0
CARGUIO_BOLAS_LC2[["TPDSAG"]][is.na(CARGUIO_BOLAS_LC2[["TPDSAG"]])] <- 0
CARGUIO_BOLAS_LC2 <- cbind(CARGUIO_BOLAS_LC2,CARGUIO_BOLAS_LC2$`MB-3`+ CARGUI
O_BOLAS_LC2$`MB-4`)
names(CARGUIO_BOLAS_LC2)[11] <- "MB"
CARGUIO_BOLAS_LC2[["TPD MB 03"]][is.na(CARGUIO_BOLAS_LC2[["TPD MB 03"]])]
<- 0
CARGUIO_BOLAS_LC2[["TPD MB 04"]][is.na(CARGUIO_BOLAS_LC2[["TPD MB 04"]])]
<- 0
CARGUIO_BOLAS_LC2 <- cbind(CARGUIO_BOLAS_LC2,CARGUIO_BOLAS_LC2$`TPD MB 03`+
CARGUIO_BOLAS_LC2$`TPD MB 04`)
names(CARGUIO_BOLAS_LC2)[12] <- "TPDMB"
CARGUIO_BOLAS_LC2$Concentradora="LC2"
CARGUIO_BOLAS_LC3 <- read_excel("CARGUIO BOLAS LC 2017-2019.xlsx", sheet = "
Læ"%ã¹ænea 3 CLC",col_types = c("date","numeric","numeric","numeric","numeric
","numeric","numeric","numeric","numeric","numeric","numeric","numeric","n
umeric","numeric","numeric","numeric","numeric","numeric","numeric","n
umeric","numeric","numeric","numeric","numeric","numeric","text"),col_names = TRUE, cel
l_limits(c(3,2),c(NA,27)))
CARGUIO_BOLAS_LC3 <- CARGUIO_BOLAS_LC3[,c(1,2,3,4,5,14,15,16,17,22,23,24,25)]
names(CARGUIO_BOLAS_LC3)[6] <- "MSAG"
names(CARGUIO_BOLAS_LC3)[7] <- "MB-5"
names(CARGUIO_BOLAS_LC3)[8] <- "MB-6"
names(CARGUIO_BOLAS_LC3)[9] <- "MB-7"
names(CARGUIO_BOLAS_LC3)[2] <- "TPDSAG"

```

```

CARGUIO_BOLAS_LC3[["MSAG"]][is.na(CARGUIO_BOLAS_LC3[["MSAG"]])] <- 0
CARGUIO_BOLAS_LC3[["MB-5"]][is.na(CARGUIO_BOLAS_LC3[["MB-5"]])] <- 0
CARGUIO_BOLAS_LC3[["MB-6"]][is.na(CARGUIO_BOLAS_LC3[["MB-6"]])] <- 0
CARGUIO_BOLAS_LC3[["MB-7"]][is.na(CARGUIO_BOLAS_LC3[["MB-7"]])] <- 0
CARGUIO_BOLAS_LC3[["TPDSAG"]][is.na(CARGUIO_BOLAS_LC3[["TPDSAG"]])] <- 0
CARGUIO_BOLAS_LC3 <- cbind(CARGUIO_BOLAS_LC3,CARGUIO_BOLAS_LC3$`MB-5`+CARGUIO
_BOLAS_LC3$`MB-6`+CARGUIO_BOLAS_LC3$`MB-7`)
names(CARGUIO_BOLAS_LC3)[14] <- "MB"
CARGUIO_BOLAS_LC3[["TPD MB 05"]][is.na(CARGUIO_BOLAS_LC3[["TPD MB 05"]])]
<- 0
CARGUIO_BOLAS_LC3[["TPD MB 06"]][is.na(CARGUIO_BOLAS_LC3[["TPD MB 06"]])]
<- 0
CARGUIO_BOLAS_LC3[["TPD MB 07"]][is.na(CARGUIO_BOLAS_LC3[["TPD MB 07"]])]
<- 0
CARGUIO_BOLAS_LC3 <- cbind(CARGUIO_BOLAS_LC3,CARGUIO_BOLAS_LC3$`TPD MB
05`+CARGUIO_BOLAS_LC3$`TPD MB 06`+CARGUIO_BOLAS_LC3$`TPD MB 07`)
names(CARGUIO_BOLAS_LC3)[15] <- "TPDMB"
CARGUIO_BOLAS_LC3$Concentradora="LC3"
CARGUIO_BOLAS_LC1 <- CARGUIO_BOLAS_LC1[,c(1,2,12,13,5,11)]
CARGUIO_BOLAS_LC2 <- CARGUIO_BOLAS_LC2[,c(1,2,12,13,5,11)]
CARGUIO_BOLAS_LC3 <- CARGUIO_BOLAS_LC3[,c(1,2,15,16,6,14)]

```

```

A <- merge(CARGUIO_BOLAS_LC1,CARGUIO_BOLAS_LC2, by=c("Fecha"), all.x = TRUE,
all.y= TRUE)
BDLC <- merge(A,CARGUIO_BOLAS_LC3,by=c("Fecha"), all.x = TRUE, all.y = TRUE)
rm(A)

```

```

BDLC[["TPDSAG.x"]][is.na(BDLC[["TPDSAG.x"]])] <- 0
BDLC[["TPDSAG.y"]][is.na(BDLC[["TPDSAG.y"]])] <- 0
BDLC[["TPDSAG"]][is.na(BDLC[["TPDSAG"]])] <- 0
BDLC[["TPDMB.x"]][is.na(BDLC[["TPDMB.x"]])] <- 0
BDLC[["TPDMB.y"]][is.na(BDLC[["TPDMB.y"]])] <- 0
BDLC[["TPDMB"]][is.na(BDLC[["TPDMB"]])] <- 0
BDLC[["MSAG.x"]][is.na(BDLC[["MSAG.x"]])] <- 0
BDLC[["MSAG.y"]][is.na(BDLC[["MSAG.y"]])] <- 0
BDLC[["MSAG"]][is.na(BDLC[["MSAG"]])] <- 0
BDLC[["MB.x"]][is.na(BDLC[["MB.x"]])] <- 0
BDLC[["MB.y"]][is.na(BDLC[["MB.y"]])] <- 0
BDLC[["MB"]][is.na(BDLC[["MB"]])] <- 0
BDLC <- cbind(BDLC, BDLC$TPDSAG.x+BDLC$TPDSAG.y+BDLC$TPDSAG)
BDLC <- cbind(BDLC, BDLC$TPDMB.x+BDLC$TPDMB.y+BDLC$TPDMB)
BDLC <- cbind(BDLC, BDLC$MSAG.x+BDLC$MSAG.y+BDLC$MSAG)
BDLC <- cbind(BDLC, BDLC$MB.x+BDLC$MB.y+BDLC$MB)
BDLC <- BDLC[,c(1,17,18,19,20)]
names(BDLC)[2] <- "TPDSAG"
names(BDLC)[3] <- "TPDMB"
names(BDLC)[4] <- "MSAG"
names(BDLC)[5] <- "MB"

```

*#DATA MD's*

```
Route.Difot<-"C:/Users/NAVAJI9/Desktop/BHP/R/Molinos de Bolas/MD"
```

```

filenames<- list.files(Route.Difot)

MD_LS1<- data.frame()

for (f in filenames){
  tryCatch(
    {
      file.route<-paste(Route.Difot,f,sep="/")
      print(paste("reading",f,sep=" "))
      difot<-read_excel(file.route, sheet = "LS", col_types = c("date", "num
eric", "numeric", "numeric", "numeric", "numeric", "numeric", "nume
ric", "numeric", "numeric", "numeric", "numeric", "numeric", "nume
ric", "numeric", "numeric", "numeric", "numeric"),col_names = TRUE, cell_limi
ts(c(1,1),c(32,20)))
      difot$filename<-f
      MD_LS1<-rbind.fill(MD_LS1,difot)
      print("Read Succesfully")

    },
    error = function(cond){return(paste("Error en", f, sep=" "))}
  )
}

names(MD_LS1)[1] <- "Fecha"

LS1 <- merge(CARGUIO_BOLAS_L1, MD_LS1, by.x = "Fecha")

LS1 <- LS1[,c(1,2,21,22,5,10,14,15,16,17,18,19,20,26,27,28,29,30)]
LS1$`Py (%)` <- LS1$`Py (%)`*100
LS1$`Cpy (%)` <- LS1$`Cpy (%)`*100
LS1$`Cc (%)` <- LS1$`Cc (%)`*100
LS1$`Cv (%)` <- LS1$`Cv (%)`*100

names(LS1) = c("Fecha", "TPD", "BWI", "SPI", "MSAG", "MB", "CuT", "CuS", "Fe", "As", "M
o", "Au", "Ag", "Pirita", "Calcopirita", "Calcosita", "Covelina", "Cuproferrico")
LS1 <- LS1[,c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)]

MD_LS2<- data.frame()

for (f in filenames){
  tryCatch(
    {
      file.route<-paste(Route.Difot,f,sep="/")
      print(paste("reading",f,sep=" "))
      difot<-read_excel(file.route, sheet = "OGP1", col_types = c("date", "n
umeric", "numeric", "numeric", "numeric", "numeric", "numeric", "nu
meric", "numeric", "numeric", "numeric", "numeric", "numeric", "nu
meric", "numeric", "numeric", "numeric", "numeric"),col_names = TRUE, cell_li
mits(c(1,1),c(32,20)))

```

```

difot$filename<-f
MD_LS2<-rbind.fill(MD_LS2,difot)
print("Read Successfully")

},
error = function(cond){return(paste("Error en", f, sep=" "))}
)
}

names(MD_LS2)[1] <- "Fecha"
LS2 <- merge(CARGUIO_BOLAS_L2, MD_LS2, by.x = "Fecha")
LS2 <- LS2[,c(1,2,21,22,5,10,14,15,16,17,18,19,20,26,27,28,29,30)]
LS2$`Py (%)` <- LS2$`Py (%)`*100
LS2$`Cpy (%)` <- LS2$`Cpy (%)`*100
LS2$`Cc (%)` <- LS2$`Cc (%)`*100
LS2$`Cv (%)` <- LS2$`Cv (%)`*100

names(LS2) = c("Fecha", "TPD", "BWI", "SPI", "MSAG", "MB", "CuT", "CuS", "Fe", "As", "Mo", "Au", "Ag", "Pirita", "Calcopirita", "Calcosita", "Covelina", "Cuproferrico")
LS2 <- LS2[,c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17)]

MD_LC<- data.frame()

for (f in filenames){
  tryCatch(
    {
      file.route<-paste(Route.Difot,f,sep="/")
      print(paste("reading",f,sep=" "))
      difot<-read_excel(file.route, sheet = "LC", col_types = c("date", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric", "numeric"),col_names = TRUE, cell_limits(c(1,1),c(32,20)))
      difot$filename<-f
      MD_LC<-rbind.fill(MD_LC,difot)
      print("Read Successfully")

    },
    error = function(cond){return(paste("Error en", f, sep=" "))}
  )
}

names(MD_LC)[1] <- "Fecha"
LC <- merge(BDLC, MD_LC, by.x = "Fecha")
LC <- LC[,c(1,2,3,4,5,8,9,10,11,12,13,14,15,16,20,21,22,23,24)]
LC$`Py (%)` <- LC$`Py (%)`*100
LC$`Cpy (%)` <- LC$`Cpy (%)`*100
LC$`Cc (%)` <- LC$`Cc (%)`*100
LC$`Cv (%)` <- LC$`Cv (%)`*100

```

```

names(LC) = c("Fecha", "TPDSAG", "TPDMB", "MSAG", "MB", "CuT", "CuS", "Fe", "As", "Mo",
,"Au", "Ag", "BWI", "SPI", "Pirita", "Calcopirita", "Calcosita", "Covelina", "Cuprofe
rrico")

rm(CARGUIO_BOLAS_LC1)
rm(CARGUIO_BOLAS_LC2)
rm(CARGUIO_BOLAS_LC3)
rm(difot)
rm(BDLC)
rm(MD_LC)

LC <- LC[,c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18)]

#Data Potencias

Route.Difot<-"C:/Users/NAVAJI9/Desktop/BHP/R/Molinos de Bolas/Potencias"
filenames<- list.files(Route.Difot)

Potencias_LS1<- data.frame()

for (f in filenames){
  tryCatch(
    {
      file.route<-paste(Route.Difot,f,sep="/")
      print(paste("reading",f,sep=" "))
      difot<-read_excel(file.route, sheet = "Laguna Seca L1", col_types = c(
"date", "numeric", "numeric"),col_names = TRUE, cell_limits(c(1,1),c(500,3)))
      difot$filename<-f
      Potencias_LS1<-rbind.fill(Potencias_LS1,difot)
      print("Read Succesfully")

    },
    error = function(cond){return(paste("Error en", f, sep=" "))}
  )
}

Potencias_LS1 <- Potencias_LS1[,c(1,3)]
names(Potencias_LS1)[2] <- "PotenciaMB"
Potencias_LS1 <- na.omit(Potencias_LS1, cols="PotenciaMB")

LS1 <- merge(LS1, Potencias_LS1, by.x = "Fecha")
rm(CARGUIO_BOLAS_L1)
rm(difot)
rm(MD_LS1)
rm(Potencias_LS1)

Potencias_LS2<- data.frame()

for (f in filenames){
  tryCatch(

```

```

{
  file.route<-paste(Route.Difot,f,sep="/")
  print(paste("reading",f,sep=" "))
  difot<-read_excel(file.route, sheet = "Laguna Seca L2", col_types = c(
"date", "numeric", "numeric"),col_names = TRUE, cell_limits(c(1,1),c(500,3)))
  difot$filename<-f
  Potencias_LS2<-rbind.fill(Potencias_LS2,difot)
  print("Read Successfully")

},
error = function(cond){return(paste("Error en", f, sep=" "))}
)
}

Potencias_LS2 <- Potencias_LS2[,c(1,3)]
names(Potencias_LS2)[2] <- "PotenciaMB"
Potencias_LS2 <- na.omit(Potencias_LS2, cols="PotenciaMB")

LS2 <- merge(LS2, Potencias_LS2, by.x = "Fecha")
rm(CARGUIO_BOLAS_L2)
rm(difot)
rm(MD_LS2)
rm(Potencias_LS2)

Potencias_LC<- data.frame()

for (f in filenames){
  tryCatch(
  {
    file.route<-paste(Route.Difot,f,sep="/")
    print(paste("reading",f,sep=" "))
    difot<-read_excel(file.route, sheet = "Los Colorados", col_types = c("
date", "numeric", "numeric"),col_names = TRUE, cell_limits(c(1,1),c(500,3)))
    difot$filename<-f
    Potencias_LC<-rbind.fill(Potencias_LC,difot)
    print("Read Successfully")

  },
  error = function(cond){return(paste("Error en", f, sep=" "))}
  )
}

Potencias_LC <- Potencias_LC[,c(1,3)]
names(Potencias_LC)[2] <- "PotenciaMB"
Potencias_LC <- na.omit(Potencias_LC, cols="PotenciaMB")

LC <- merge(LC, Potencias_LC, by.x = "Fecha")
rm(difot)
rm(Potencias_LC)

```



```

#FORECAST
LS1 <- na.omit(LS1)
LS1 <- LS1[LS1$`TPD` !=0,]

LS2 <- na.omit(LS2)
LS2 <- LS2[LS2$`TPD` !=0,]

LC <- na.omit(LC)

#MEMORIA

LS1$Fecha <- as.Date(LS1$Fecha, format="%Y-%m-%d ")
LS1 <- subset(LS1, Fecha < as.Date("2019-07-01"))

LS2$Fecha <- as.Date(LS2$Fecha, format="%Y-%m-%d ")
LS2 <- subset(LS2, Fecha < as.Date("2019-07-01"))

LC$Fecha <- as.Date(LC$Fecha, format="%Y-%m-%d ")
LC <- subset(LC, Fecha < as.Date("2019-07-01"))

# OUTLIERS TREATMENT

grubbs.outliers <- function(x, p.thresh=0.05) {require(outliers, quietly=TRUE)
}
  x <- x[!is.na(x)]
  n <- length(x)
  zs <- abs(x - mean(x)) / sd(x)
  outs <- 1 - sapply(zs, function(z) pgrubbs(z, n, type=10))
  posOuts <- which(outs <= p.thresh)
  return(list(zs=zs,pvals=outs,outliers=x[posOuts],positions=posOuts))
}
outliers1 <-grubbs.outliers(LS1$MSAG)$outliers
LS1 <- LS1[-which(LS1$MSAG %in% outliers1),]
outliers2 <-grubbs.outliers(LS1$MB)$outliers
LS1 <- LS1[-which(LS1$MB %in% outliers2),]
outliers3 <-grubbs.outliers(LS2$MSAG)$outliers
LS2 <- LS2[-which(LS2$MSAG %in% outliers3),]
outliers4 <-grubbs.outliers(LS2$MB)$outliers
LS2 <- LS2[-which(LS2$MB %in% outliers4),]
outliers5 <-grubbs.outliers(LC$MSAG)$outliers
LC <- LC[-which(LC$MSAG %in% outliers5),]
outliers6 <-grubbs.outliers(LC$MB)$outliers
LC <- LC[-which(LC$MB %in% outliers6),]

#FY 20 TEST ANALYSIS

tr1 <- subset(LS1, Fecha < as.Date("2018-07-01"))
ts1 <- subset(LS1, Fecha > as.Date("2018-06-30"))

tr2 <- subset(LS2, Fecha < as.Date("2018-07-01"))
ts2 <- subset(LS2, Fecha > as.Date("2018-06-30"))

```

```

trlc <- subset(LC, Fecha < as.Date("2018-07-01"))
tslc <- subset(LC, Fecha > as.Date("2018-06-30"))
tslc <- tslc[tslc$`MSAG` + tslc$`MB` != 0,]

#ESTADÍSTICA DESCRIPTIVA
stargazer(subset(LS1[,2:18]),title = "Estadísticas descriptivas Laguna Seca 1",
  digits = 1, type= "text", out= "LS1.htm")
stargazer(subset(LS2[,2:18]),title = "Estadísticas descriptivas Laguna Seca 2",
  digits = 1, type= "text", out= "LS2.htm")
stargazer(subset(LC[,2:19]),title = "Estadísticas descriptivas Los Colorados",
  digits = 1, type= "text", out= "LC.htm")

#CORRELACIONES
correlacionesLS1=cor(LS1[,2:18])
stargazer(correlacionesLS1,type = "text", title = "Matriz de Correlación LS1",
  digits = 3, out = "correlacionesLS1")
summary(LS1)

correlacionesLS2=cor(LS2[,2:18])
stargazer(correlacionesLS2,type = "text", title = "Matriz de Correlación LS2",
  digits = 3, out = "correlacionesLS2")

correlacionesLC=cor(LC[,2:19])
stargazer(correlacionesLS2,type = "text", title = "Matriz de Correlación LS2",
  digits = 3, out = "correlacionesLC")

#BAGGING WITH CARET

#LS1
set.seed(1234)
ctrl <- trainControl(method= "cv", number = 5)

bagged <- train(
  MSAG ~ .,
  data = LS1[,-c(1,6,19)],
  method = "treebag",
  trControl = ctrl,
  importance = TRUE
)

bagged
#MOLINO DE BOLAS

set.seed(123)
ctrl <- trainControl(method = "cv", number = 5)
bagged <- train(
  MB ~ .,
  data = LS1[,-c(1,5,18)],
  method = "treebag",
  trControl = ctrl,
  importance = TRUE
)

```

```

)

bagged

#LS2
set.seed(123)
ctrl <- trainControl(method= "cv", number = 5)

bagged <- train(
  MSAG ~ .,
  data = LS2[,-c(1,6,19)],
  method = "treebag",
  trControl = ctrl,
  importance = TRUE
)

bagged
#MOLINO DE BOLAS

bagged <- train(
  MB ~ .,
  data = LS2[,-c(1,5,18)],
  method = "treebag",
  trControl = ctrl,
  importance = TRUE
)

bagged
plot(varImp(bagged))
varImp(bagged)

#LC
set.seed(123)
ctrl <- trainControl(method = "cv", number = 5)

bagged <- train(
  MSAG ~ .,
  data = LC[,-c(1,3,5,20)],
  method = "treebag",
  trControl = ctrl,
  importance = TRUE
)

bagged

#Molino de Bolas

bagged <- train(
  MB ~ .,
  data = LC[,-c(1,2,4,19)],
  method = "treebag",
  trControl = ctrl,
  importance = TRUE
)

```

```

)
bagged

#RANDOM FOREST

#LS1
rf <- train(
  MSAG ~ .,
  data = LS1[,-c(1,6,19)],
  method = "rf",
  trControl = ctrl,
  importance = TRUE,
  n.trees = 100000
)

rf
plot(varImp(rf))
varImp(rf)

rf2 <- train(
  MB ~ .,
  data = LS1[,-c(1,5,18)],
  method = "rf",
  trControl = ctrl,
  importance = TRUE,
  n.trees = 100000
)

rf2
plot(varImp(rf))
varImp(rf)

#LS2
ctrl <- trainControl(method = "cv", number = 5)

rf <- train(
  MSAG ~ .,
  data = LS2[,-c(1,6,19)],
  method = "rf",
  trControl = ctrl,
  importance = TRUE,
  n.trees = 100000
)

rf
plot(varImp(rf))
varImp(rf)

rf2 <- train(
  MB ~ .,
  data = LS2[,-c(1,5,18)],
  method = "rf",

```

```

    trControl = ctrl,
    importance = TRUE,
    n.trees = 100000
  )

rf2
plot(varImp(rf))
varImp(rf)

#LC
ctrl <- trainControl(method = "cv", number = 5)
rf <- train(
  MSAG ~ .,
  data = LC[, -c(1,3,5,20)],
  method = "rf",
  trControl = ctrl,
  importance = TRUE,
  n.trees = 100000
)

rf
plot(varImp(rf))
varImp(rf)

rf2 <- train(
  MB ~ .,
  data = LC[, -c(1,2,4,19)],
  method = "rf",
  trControl = ctrl,
  importance = TRUE,
  n.trees = 100000
)

rf2
plot(varImp(rf))
varImp(rf)

#GBM LS1

#TUNE GBM

gbmGrid <- expand.grid(interaction.depth = c(1,3,6,9,12),
  n.trees = c(100,250, 500,1000,5000),
  shrinkage = c(0.005, 0.001),
  n.minobsinnode = 10)

gbmGrid1 <- expand.grid(interaction.depth = 9 ,
  n.trees = 250,
  shrinkage = 0.005,
  n.minobsinnode = 10)

gbmGrid2 <- expand.grid(interaction.depth = 9 ,

```

```

        n.trees = 500,
        shrinkage = 0.005,
        n.minobsinnode = 10)

LS1SAG <- LS1[LS1$`MSAG` > 1 ,]
tr1SAG <- tr1[tr1$`MSAG` > 1 ,]

set.seed(123)
ctrl <- trainControl(method="cv", number = 5)

fit <- train(MSAG ~., data=LS1SAG[,-c(1,3,6,18)], method= "gbm", verbose= FALSE,
trControl= ctrl, tuneGrid=gbmGrid )
fit <- train(MSAG ~., data=tr1SAG[,-c(1,3,6,18)], method= "gbm", verbose= FALSE,
tuneGrid=gbmGrid1 )
predicted1 = predict(fit,ts1)
regressionMetrics(ts1$MSAG, predicted1, train.y = tr1$MSAG)
summary(fit, cBars=10, las=2, xlim=c(0,25))
summary(predicted1)

plot(fit$finalModel, "TPD")
plot(fit$finalModel, "Fe")

SAG1=data.frame(ts1$Fecha, predicted1)
SAG1$ts1.Fecha <- as.Date(SAG1$ts1.Fecha, format="%Y-%m-%d ")

SAG1%>%
  mutate(month = month(ts1.Fecha), year= year(ts1.Fecha)) %>%
  group_by(month,year) %>%
  dplyr::summarise(totalps=sum(predicted1))

#molino de bolas
set.seed(1234)

LS1MB <- LS1[LS1$`MB` > 1 ,]
LS1MB <- LS1MB[LS1MB$`PotenciaMB` > 1 ,]
tr1MB <- tr1[tr1$`MB` > 1 ,]
tr1MB <- tr1MB[tr1MB$`PotenciaMB` > 1 ,]

fitcontrol <- trainControl( method = "cv", number = 5)
fit2 <- train(MB ~., data=LS1MB[,-c(1,2,4,5)], method= "gbm", verbose=FALSE,
tuneGrid = gbmGrid, trControl= ctrl)
fit2 <- train(MB ~., data=tr1MB[,-c(1,2,4,5)], method= "gbm", verbose=FALSE,
tuneGrid = gbmGrid2)

predicted2 = predict(fit2,ts1)
regressionMetrics(ts1$MB, predicted2, train.y = tr1$MB)
summary(fit2, cBars=10, las=2, xlim=c(0,20))
fit2
summary(predicted)

MB2=data.frame(ts1$Fecha, predicted2)
MB2$ts1.Fecha <- as.Date(MB2$ts1.Fecha, format="%Y-%m-%d ")

```

```

MB2%>%
  mutate(month = month(ts1.Fecha), year= year(ts1.Fecha)) %>%
  group_by(month,year) %>%
  dplyr::summarise(totalps=sum(predicted2))

plot(fit$finalModel, "PotenciaMB")
plot(fit$finalModel, "PotenciaMo")

#LS2
gbmGrid3 <- expand.grid(interaction.depth = 12 ,
                        n.trees = 250 ,
                        shrinkage = 0.005 ,
                        n.minobsinnode = 10)

gbmGrid4 <- expand.grid(interaction.depth = 12 ,
                        n.trees = 500 ,
                        shrinkage = 0.005 ,
                        n.minobsinnode = 10)

LS2SAG <- LS2[LS2$`MSAG` > 1 ,]
tr2SAG <- tr2[tr2$`MSAG` > 1 ,]

set.seed(123)
ctrl <- trainControl(method="cv", number = 5)
fit3 <- train(MSAG ~., data=LS2SAG[,-c(1,3,6,18)], method= "gbm", verbose= FA
LSE , trControl = ctrl, tuneGrid= gbmGrid3 )
fit3 <- train(MSAG ~., data=tr2SAG[,-c(1,3,6,18)], method= "gbm", verbose= FA
LSE , tuneGrid= gbmGrid3 )

predicted3 = predict(fit3,ts2)
regressionMetrics(ts2$MSAG, predicted3, train.y = tr2SAG$MSAG)
summary(fit3, cBars=10, las=2, xlim=c(0,15))
summary(predicted3)
fit3

plot(fit3$finalModel, "TPD")
plot(fit3$finalModel, "Fe")

MSAG2=data.frame(ts2$Fecha,predicted3)
MSAG2$ts2.Fecha <- as.Date(MSAG2$ts2.Fecha, format="%Y-%m-%d ")
MSAG2%>%
  mutate(month = month(ts2.Fecha), year= year(ts2.Fecha)) %>%
  group_by(month,year) %>%
  dplyr::summarise(totalps=sum(predicted3))

#molino de bolas
LS2MB <- LS2[LS2$`MB` > 1 ,]
LS2MB <- LS2MB[LS2MB$`PotenciaMB` > 1 ,]
tr2MB <- tr2[tr2$`MB` > 1 ,]
tr2MB <- tr2MB[tr2MB$`PotenciaMB` > 1 ,]

set.seed(1210)

```

```

fitcontrol <- trainControl( method = "cv", number = 5)
fit4 <- train(MB ~., data=LS2MB[,-c(1,2,4,5)], method= "gbm", trControl= fitc
ontrol, verbose=FALSE, tuneGrid = gbmGrid4)
fit4 <- train(MB ~., data=tr2MB[,-c(1,2,4,5)], method= "gbm", verbose=FALSE,
tuneGrid = gbmGrid4)
predicted4 = predict(fit4,ts2)
regressionMetrics(ts2$MB, predicted4, train.y = tr2MB$MB)
summary(fit4, cBars=10, las=2, xlim=c(0,17))
fit4
summary(predicted4)

MB2=data.frame(ts2$Fecha,predicted4)
MB2$ts2.Fecha <- as.Date(MB2$ts2.Fecha, format="%Y-%m-%d ")
MB2%>%
  mutate(month = month(ts2.Fecha), year= year(ts2.Fecha)) %>%
  group_by(month,year) %>%
  dplyr::summarise(totalps=sum(predicted4))

plot(fit$finalModel, "PotenciaMB")
plot(fit$finalModel, "Covelina")

#LC
gbmGrid5 <- expand.grid(interaction.depth = 3,
                        n.trees = 500,
                        shrinkage = 0.005,
                        n.minobsinnode = 10)

gbmGrid6 <- expand.grid(interaction.depth = 9 ,
                        n.trees = 1000 ,
                        shrinkage = 0.001,
                        n.minobsinnode = 10)

LCSAG <- LC[LC$`TPDSAG` !=0,]
LCSAG <- LCSAG[LCSAG$`MSAG` > 1 ,]
trlcSAG <- trlc[trlc$`MSAG` > 1 ,]

set.seed(123)
ctrl <- trainControl(method="cv", number = 5)
fit5 <- train(MSAG ~., data=LCSAG[,-c(1,3,5,13,19)], method= "gbm", verbose=
FALSE , trControl = ctrl , tuneGrid = gbmGrid5)
fit5 <- train(MSAG ~., data=trlcSAG[,-c(1,3,5,13,19)], method= "gbm", verbose
= FALSE, tuneGrid = gbmGrid5)
predicted5 = predict(fit5,tslc)
summary(fit5, cBars=10, las=2, xlim=c(0,40))
summary(predicted5)
fit5

plot(fit5$finalModel, "TPDSAG")
plot(fit5$finalModel, "As")

MSAGLC=data.frame(tslc$Fecha,predicted5)
MSAGLC$tslc.Fecha <- as.Date(MSAGLC$tslc.Fecha, format="%Y-%m-%d ")

```



```

MSAGLC%>%
  mutate(month = month(tslc.Fecha), year= year(tslc.Fecha)) %>%
  group_by(month,year) %>%
  dplyr::summarise(totalps=sum(predicted5))

#molino de bolas
LCMB <- LC[LC$`TPDMB` !=0,]
LCMB <- LCMB[LCMB$`MB` > 1 ,]
LCMB <- LCMB[LCMB$`PotenciaMB` > 1 ,]

trlcMB <- trlc[trlc$`MB` > 1 ,]
trlcMB <- trlcMB[trlcMB$`PotenciaMB` > 1 ,]

set.seed(1234)
fitcontrol <- trainControl( method = "cv", number = 5)
fit6 <- train(MB ~., data=LCMB[,-c(1,2,3,4,14)], method= "gbm", trControl= fi
tcontrol, verbose=FALSE, tuneGrid= gbmGrid6)
fit6 <- train(MB ~., data=trlcMB[,-c(1,2,3,4,14)], method= "gbm", verbose=FAL
SE, tuneGrid= gbmGrid6)
predicted6 = predict(fit6,tslc)
regressionMetrics(ts$MB, predicted2, train.y = tr$MB)
summary(fit6, cBars=10, las=2, xlim=c(0,25))
fit6
summary(predicted6)

plot(fit6$finalModel, "PotenciaMB")
plot(fit6$finalModel, "BWI")

MBLC=data.frame(tslc$Fecha,predicted6)
MBLC$tslc.Fecha <- as.Date(MBLC$tslc.Fecha, format="%Y-%m-%d ")
MBLC%>%
  mutate(month = month(tslc.Fecha), year= year(tslc.Fecha)) %>%
  group_by(month,year) %>%
  dplyr::summarise(totalps=sum(predicted6))

#SPOTFIRE

#LS1
ForecastLS1= data.frame(ts1$Fecha,ts1$MSAG,ts1$MB, predicted1, predicted2)
names(ForecastLS1)=c("Fecha", "Carguæ"%ã¹¸o SAG", "Carguæ"%ã¹¸o MB", "Forecast
SAG", "Forecast MB")
write.xlsx(ForecastLS1, "ForecastLS1.xlsx")

#LS2
ForecastLS2= data.frame(ts2$Fecha,ts2$MSAG,ts2$MB, predicted3, predicted4)
names(ForecastLS2)=c("Fecha", "Carguæ"%ã¹¸o SAG", "Carguæ"%ã¹¸o MB", "Forecast
SAG", "Forecast MB")
write.xlsx(ForecastLS2, "ForecastLS2.xlsx")

#LC
ForecastLC= data.frame(tslc$Fecha,tslc$MSAG,tslc$MB, predicted5, predicted6)
names(ForecastLC)=c("Fecha", "Carguæ"%ã¹¸o SAG", "Carguæ"%ã¹¸o MB", "Forecast S

```

```
AG", "Forecast MB")  
write.xlsx(ForecastLC, "ForecastLC.xlsx")
```