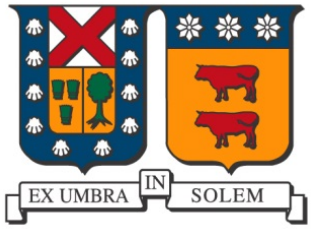
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE ELECTRÓNICA
VALPARAÍSO - CHILE



"CONSTRUCCIÓN DE ÍNDICES DE DEGRADACIÓN PARA UN MOLINO SAG"

SANDRA VALENTINA SILVA GONZÁLEZ

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL EN ELECTRÓNICA

Profesor Guía: Juan Yuz Eissmann
Profesor Correferente: Marcos Orchard Concha

Marzo - 2022

RESUMEN

En la industria minera se requiere una gran exigencia de las máquinas que hacen posible el proceso minero. Los complejos procesos ejecutados a diario exigen realizar mantenimiento a las máquinas, lo que a menudo implica la interrupción del proceso completo. Por tanto, es útil conocer el estado de la maquinaria para realizar mantenimiento en el momento adecuado.

Esta memoria se basa en la metodología para obtener un modelo de degradación para un molino SAG propuesta por Eduardo Grendi, en su memoria de título. El modelo se diseñó para ser usado durante la operación del activo, estimando su tasa de desgaste de acuerdo al modo de operación del molino SAG para obtener una predicción del momento óptimo para realizarle mantenimiento. El objetivo central de esta memoria de título es perfeccionar la metodología previamente desarrollada por Eduardo Grendi. Para ello se revisa analíticamente el modelo de degradación y los fundamentos detrás de la construcción del índice de degradación propuestos. Luego, utilizando el modelo en un activo diferente (otro molino SAG), se cambian algunas decisiones tomadas con el fin de mejorar la metodología. Además, se utiliza información adicional de los operadores del SAG y del origen del mineral en su interior.

Se dispone de datos reales proporcionados por CODELCO división Chuquicamata. Estos se manipulan mediante PCA para observar el comportamiento de las variables y escoger las que más información aporten en el análisis. Luego, se utiliza UMAP para visibilizar los datos y encontrar modos de operación. Además, el replicar los análisis en dos molinos diferentes permite comprobar que la metodología puede ser comparable entre activos. Esto permitirá crear un indice de degradación en futuras investigaciones.

ABSTRACT

Abstract— In the mining industry there are big requirements for the machines that make the mining process possible. That complex processes are executed daily implies that maintenance is needed for these machines, this often implies the full interruption of the process. This is why knowing the state of the machinery is convenient for making the maintenance in the precise moments.

This degree project is based on the methodology for obtaining a degradation model for a SAG mill proposed by Eduardo Grendi, in his degree thesis. The model was designed to be used during the operation of the asset, estimating its wear rate according to the SAG mill´s operation mode to obtain a prediction of the optimal moment to perform maintenance.

The main goal of this degree project is to improve the methodology developed by Eduardo Grendi. For this purpose, the degradation model and the foundations behind the construction of the degradation rate proposed by him are analytically revised. Then, using the model on a different machine (another SAG mill), some decisions are changed in order to improve the methodology. Moreover, additional information from the SAG operators and the origin of the ore inside the SAG is used.

Real data provided by CODELCO Chuquicamata division is available. These are manipulated using PCA to observe the behavior of the variables and choose the ones that provide the most information in the analysis. UMAP is then used to make the data visible and find modes of operation. In addition, replicating the analyzes in two different mills makes it possible to verify that the methodology can be comparable between assets. This will allow to create a degradation index in future investigations.

ÍNDICE DE CONTENIDOS

RESUMEN I

ABSTRACT II

CAPÍTULO 1: INTRODUCCIÓN **1**

1.1 ANTECEDENTES 1

1.2 MOLINOS SAG 5

1.3 TRABAJO A DESARROLLAR 7

1.4 OBJETIVOS DE LA MEMORIA 8

1.5 ESTRUCTURA DEL DOCUMENTO 9

CAPÍTULO 2: CIENCIA DE DATOS **10**

2.1 CICLO DEL PROCESO 10

2.2 TIPOS DE ANÁLISIS 12

2.3 PYTHON 13

2.4 ALGORITMOS UTILIZADOS 15

2.4.1 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA) 16

2.4.2 Uniform Manifold Approximation and Projection (UMAP) . . . 22

2.4.3 K-MEANS 23

CAPÍTULO 3: REVISIÓN A LA METODOLOGÍA DE ANÁLISIS . **25**

3.1 OBTENCIÓN VARIABLES INICIALES 25

3.2 FILTRO MODO MOLINO APAGADO 36

3.3 SEÑALES EN CERO 41

CAPÍTULO 4: ANÁLISIS CON ALGORITMOS Y OBTENCIÓN MODOS DE OPERACIÓN **44**

4.1 ANÁLISIS PCA 44

4.1.1 INDICADOR PESO 48

4.1.2 LOADING PLOT 51

4.2 ANÁLISIS UMAP 56

CAPÍTULO 5: CONCLUSIONES 65

5.1 APORTES A LA METODOLOGÍA PREVIA 65

5.2 LIMITACIONES 66

5.3 RESULTADOS OBTENIDOS 66

5.4 TRABAJO A FUTURO 67

REFERENCIAS BIBLIOGRÁFICAS 69

CAPÍTULO 1

INTRODUCCIÓN

1.1. ANTECEDENTES

Es primordial que las empresas mineras, tengan una mirada hacia el futuro. Según [1], la innovación en esta área en particular, ayuda a enfrentar los desafíos de incrementar sistemáticamente los niveles de productividad y la eficiencia operativa en sus procesos. Además, la minería es un proceso desafiante que en su evolución va planteando nuevas dificultades, las cuales se manejan principalmente invirtiendo en maquinarias y con el buen uso de la mejor tecnología disponible. Esto permite que las empresas estén en constante perfeccionamiento y puedan tener procesos cada vez más eficientes y en sintonía con la evolución de los mercados.

Este proyecto de título se centra en la maquinaria de molinos SAG. Actualmente, en CODELCO se realiza un mantenimiento preventivo en sus activos. Se realizan rutinas de mantención con una determinada periodicidad, independiente de lo que suceda con los molinos en el periodo de operación y, sin importar el nivel de agresividad con la que se pueda haber utilizado. En el sentido de hacer procesos más eficientes, en el trabajo de título [2] se propuso un índice cuantitativo que permite estimar la degradación del activo. Esto da la posibilidad de realizar una mantención predictiva a un molino SAG, en particular, para el reemplazo de sus revestimientos.

Debido a la exigencia de uso y producción, los molinos SAG se exponen a fallos que detienen toda la producción. Es por esto, que en esta memoria se propuso buscar y mejorar la metodología desarrollada para la estimación de tasas de degradación, y hacer posible la predicción del tiempo de vida útil del revestimiento de los molinos SAG.

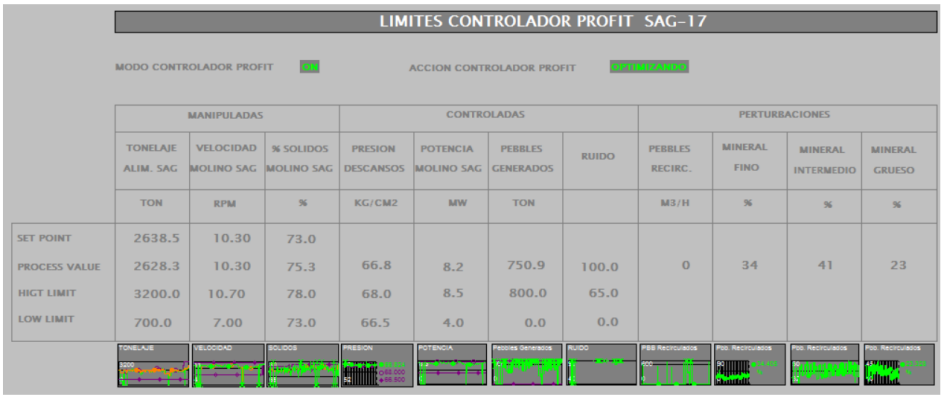
La presente memoria se realiza con datos reales de CODELCO proporcionados por la

Gerencia Digital y Analítica Avanzada. CODELCO y Honeywell han formado Kairos Mining [3], una alianza estratégica que ha permitido maximizar y mantener en el tiempo los beneficios de los sistemas de automatización en las plantas de la minera, reduciendo las fluctuaciones de las variables críticas de los procesos de concentración con el fin de tener procesos más estables y mejorar el desempeño del negocio. Se destaca específicamente la aplicación en la planta A2 de CODELCO Chuquicamata, donde las soluciones de control avanzado implementadas han sido efectivas en altos índices de utilización. Una de las ventajas que esto arroja, es la obtención de datos e información muestreadas al minuto de la planta de interés para este trabajo de memoria.

En las figuras 1.1a y 1.1b, se muestran interfaces que indican los valores deseados para variables esenciales en el sistema. Esta es la zona de controladores que sugiere los SetPoints a los operadores de los molinos.



(a) Despliegue Controlador SAG16



(b) Despliegue Controlador SAG17

Figura 1.1

La figura 1.2, representa la planta A2, donde se ubica la sección de molienda de interés. Se observan en verde las áreas activas y en rojo las áreas inactivas en un determinado momento. Cabe destacar, que a través de este software se tiene conocimiento sobre el silo que almacena el mineral que luego ingresa a los molinos SAG. Esta es la sección de la planta de CODELCO que se considera para análisis a lo largo del presente trabajo de memoria.

Las figuras 1.3a y 1.3b, corresponden al monitoreo en tiempo real de cada sensor de los molinos SAG. Esta es la toma de muestras que se van actualizando cada un minuto, dando el dataframe con las muestras de cada una de las variables que se utilizan a lo largo de todo el presente trabajo.

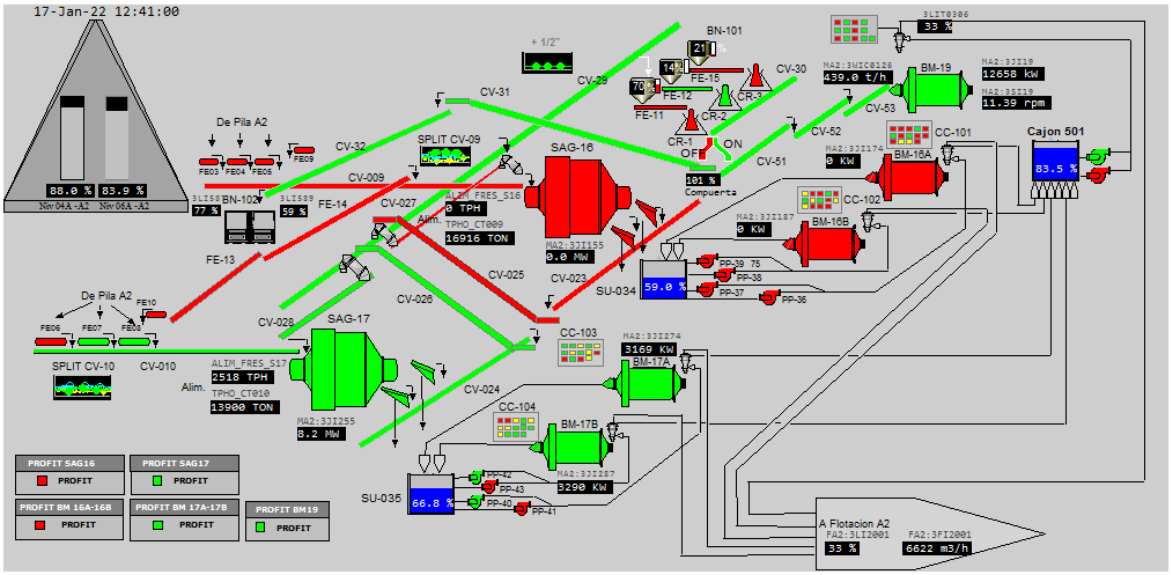
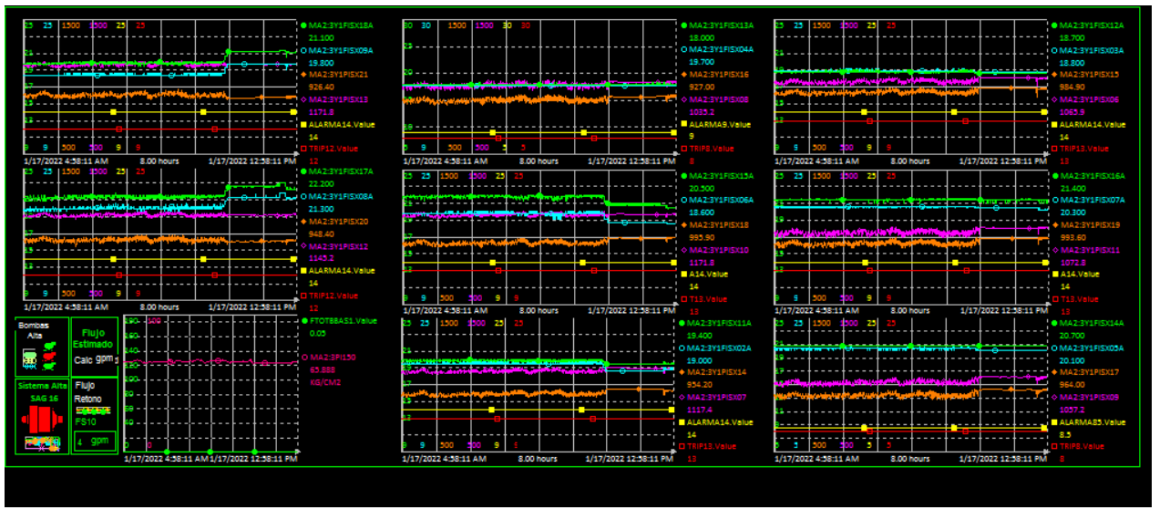
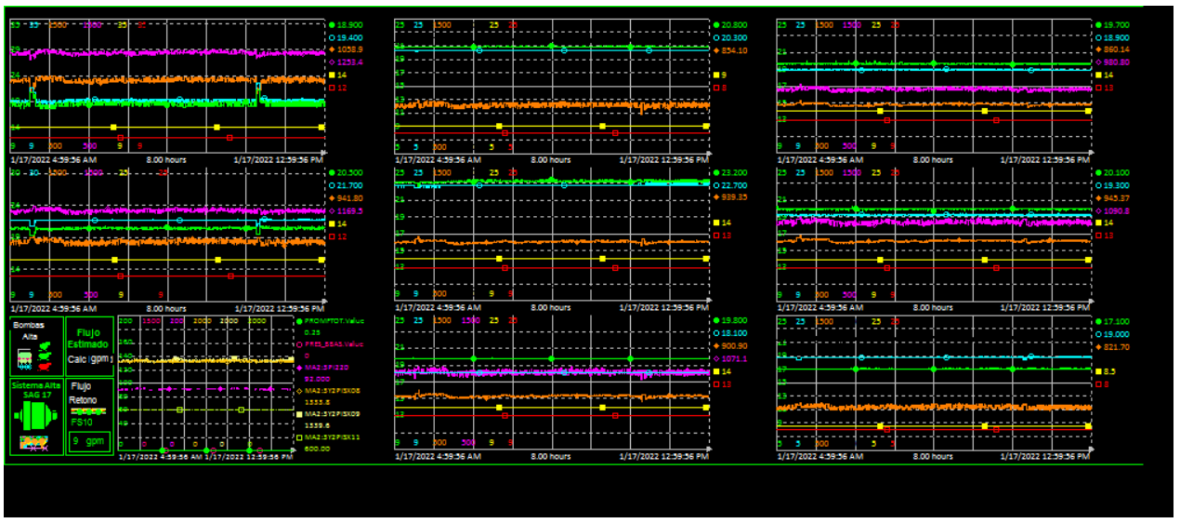


Figura 1.2: Sección planta A2 de CODELCO Chuquicamata.



(a) Variables SAG16



(b) Variables SAG17

Figura 1.3

1.2. MOLINOS SAG

Las primeras etapas del proceso productivo del cobre son la exploración de los suelos y la extracción de mineral mediante el uso de explosivos. Posteriormente, se tiene el chancado, con tres fases y, la molienda, donde se utiliza maquinaria para disminuir el tamaño de las rocas extraídas. Específicamente en la etapa de molienda se utiliza el molino SAG (semiautógeno). Este tipo de molino es una mezcla de los otros dos molinos utilizados en minería: los que utilizan bolas de algún material pesado para moler el mineral y, los autógenos que, a través de su revestimiento poroso, elevan la pulpa y la dejan caer a una determinada altura para molerla. La selección de los molinos SAG en la minería ha permitido aumentar los niveles de productividad, pues son capaces de tratar el mineral que se obtiene como producto del chancado primario descartando las etapas del chancado secundario y terciario. Esto implica un ahorro de costos producto de la menor cantidad de equipos considerados durante el chancado. Además, debido a su dureza y a las grandes toneladas que maneja el SAG es capaz de ajustarse a crecientes demandas productivas y a minerales de distintas durezas, concentrando su uso preferentemente en los minerales de cobre. Sin embargo, los revestimientos interiores del molino se degradan debido a su funcionamiento.

En esta memoria de título se consideran los molinos SAG16 y SAG17 de CODELCO, ubicados en la división de Chuquicamata. Estos dos molinos en términos de diseño son idénticos, aunque a medida que se van usando de forma diferente estos van teniendo un desgaste distinto. Es importante señalar que ambos molinos se alimentan desde una misma pila de mineral: el silo que se observa en la figura 1.2. Sin embargo, el SAG17 se alimenta principalmente de mineral grueso y el SAG16 de mineral fino. Esto sucede de manera natural, ya que las rocas en la pila de mineral se sitúan de tal forma, que el mineral grueso queda en la parte externa de la pila donde circula la correa transportadora del SAG17. Esto se traduce en una consecuencia operacional importante: el molino SAG16 procesa entre un 20-25 % más de mineral que el molino SAG17, ya que es más sencillo triturar mineral fino. A lo largo del presente trabajo,

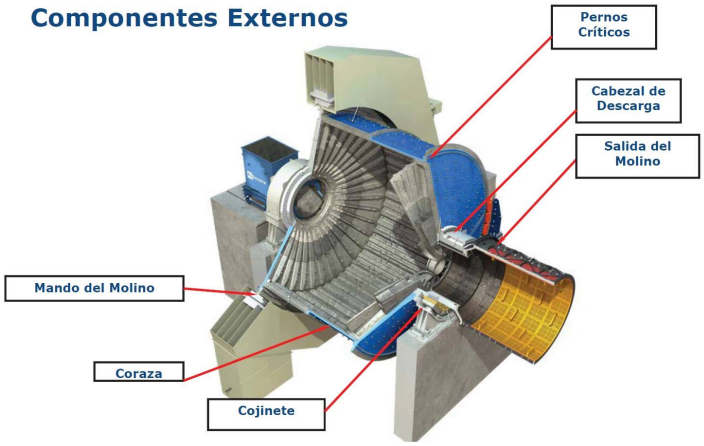
se harán análisis comparativos entre los molinos SAG16 y SAG17. En la figura 1.4, se muestra una fotografía externa de un molino SAG, y en la figura 1.5a las piezas principales que corresponen al exterior del molino.



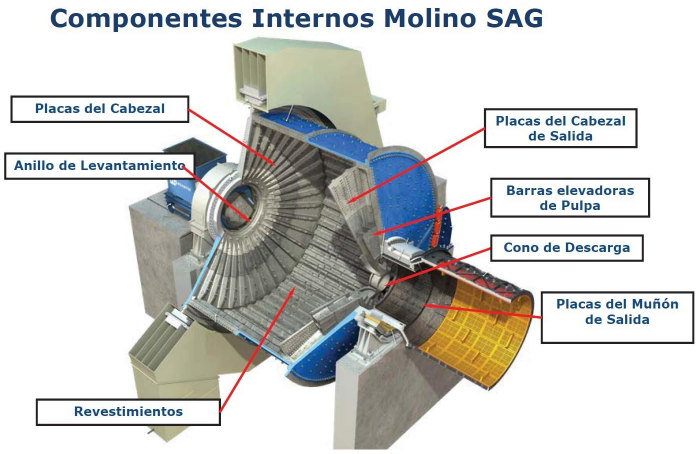
Figura 1.4: Exterior de un molino SAG.

El interior de un molino SAG se muestra en la figura 1.5b junto con el nombre de las piezas que están en su interior. El problema central a resolver en este proyecto, es mejorar un modelo de degradación o desgaste de los revestimientos de un molino SAG que permite estimar el estado actual del activo. Este modelo de degradación podría ser usado posteriormente para predecir una eventual mantención. La mantención del molino SAG implica un cambio de revestimiento, pieza de gran tamaño y valor, por lo que, tal mantención significa una suspensión del proceso minero que trae consigo grandes costos para la empresa. Más aún, detenciones no programadas para reemplazar y revisar revestimientos alterando la planificación de la operación minera.

En el presente proyecto de título, interesa caracterizar modos de operación del molino SAG que permitan obtener un indicador de su tasa de desgaste actual en base a condiciones de uso. Para esto, es importante determinar modos de operación mediante el análisis de datos y complementando con modelos físicos que permitan caracterizar matemáticamente el desgaste de revestimientos del molino SAG.



(a) Principales componentes externas del molino SAG.



(b) Principales componentes internas del molino SAG.

Figura 1.5

1.3. TRABAJO A DESARROLLAR

En primera instancia, el foco de esta memoria está en comprender el proceso minero y, en particular la operación del molino SAG. Para esto se revisó literatura sobre molinos SAG, tanto modelado fenomenológico como basado en datos. Además, se estudia la metodología propuesta previamente en [2]. Superando esta etapa, se comienzan a analizar las variables y características propias de los molinos a considerar en la memoria. Cabe destacar que el procedimiento y los análisis, se realizan tanto en el molino SAG16 como en el SAG17. Esto para ir constantemente comparando resultados y obtener conclusio-

nes que permitan tomar decisiones en la creación del modelo de degradación.

La gran cantidad de datos reales disponibles para el análisis (más de un año de operación, con datos al minuto) hace necesario utilizar herramientas que permitan reducir la alta dimensionalidad. Debido a esto se estudian algoritmos para una reducción de dimensionalidad como *Principal Component Analysis* PCA, que según [4] es un método estadístico que permite reducir la complejidad de espacios muestrales con muchas dimensiones a la vez. Asimismo, se considera *Uniform Manifold Approximation and Projection* UMAP, una técnica desarrollada para la reducción de dimensionalidad capaz de mantener la estructura global de los datos [5]. Además, se estudia el algoritmo de clustering *K-means*, que particiona un conjunto de n datos en k grupos, en el que cada dato pertenece al grupo cuyo promedio esté más cerca [6]. En particular, esto permite obtener los posibles modos de operación del molino SAG.

En esta memoria se proponen mejoras a la metodología propuesta en [2], por tanto se toman algunas decisiones diferentes contemplando nuevos análisis que perfeccionan lo ya realizado: Se añaden datos nuevos, como información sobre los operadores de los molinos SAG y del mineral que circula por los activos. Luego, se sigue el procedimiento de [2] tomando, en algunos casos, distintos criterios de selección de variables. Además, se eligen ventanas de datos para entrenamiento acordes a la influencia de las decisiones tomadas a lo largo del trabajo. Más aún, se compara la aplicación de la metodología en dos molinos diferente, SAG16 y SAG17.

1.4. OBJETIVOS DE LA MEMORIA

El proyecto "Construcción de índices de Degradación para un Molino SAG", tiene como objetivo general, proponer nuevos análisis en un molino SAG en base a señales disponibles, que permitan obtener nuevos índices de degradación en un posterior trabajo.

Los objetivos específicos a tratar en el presente trabajo de titulación son:

1. Proponer mejoras a la metodología desarrollada previamente, en especial, estudiar en profundidad la fenomenología del proceso.
2. Aplicar la metodología a un activo diferente al previamente considerado, en particular, el SAG16 en vez del SAG17.
3. Realizar análisis, para proponer la creación de un índice de degradación, en base a los resultados parciales.

1.5. ESTRUCTURA DEL DOCUMENTO

La presente memoria consta de cinco capítulos, los cuales en orden pretenden contextualizar sobre el proceso y las herramientas técnicas a utilizar e, ilustrar el problema concreto y su solución. Todo esto mientras se van cumpliendo los objetivos antes descritos.

Para un mejor entendimiento, el Capítulo 2 explica el procedimiento y tipos de análisis que se utilizan a lo largo del trabajo de memoria, junto con una descripción de las herramientas utilizadas.

El Capítulo 3 ilustra el proceso de revisión de la metodología en el molino SAG17 y explica los cambios en algunas decisiones tomadas en [2]. Luego, en el Capítulo 4 se muestran los nuevos análisis con los algoritmos y se obtienen los modos de operación de los molinos SAG.

Finalmente, el Capítulo 5 presenta comentarios y conclusiones sobre los resultados obtenidos. También se discuten posibles futuras aplicaciones, revisiones y mejoras.

CAPÍTULO 2

CIENCIA DE DATOS

La ciencia de datos o data science es el proceso de obtener información o conocimiento a partir de datos. Según [7], esta disciplina nació ante la necesidad de utilizar, interpretar y comprender de alguna forma la gran cantidad de datos que las empresas generan cada día. Esta ciencia es un campo interdisciplinario que implica el uso de métodos y sistemas científicos para extraer conocimiento y comprender mejor los datos. Utiliza y unifica técnicas y teorías de varios campos como las matemáticas, la estadística, la ciencias de la información, el análisis de datos, las ciencias de la computación y todos sus métodos relacionados para comprender y analizar los fenómenos del mundo real.

El conocimiento en estadística es muy necesario en el campo de la ciencia de datos, ya que, esta se encarga de investigar y desarrollar métodos para recoger, organizar y analizar datos. Mediante la implementación de modelos, las estadísticas ayudan a obtener conclusiones relevantes para tomar decisiones con evidencia empírica. Mientras más se conozca de estadísticas, el estudio será más sólido y podrá llegar más lejos. Otra área importante de conocer para desarrollar la ciencia de los datos, es el conocimiento del área en particular donde se está haciendo el análisis. Tener un buen dominio en la materia que se está trabajado facilitará la interpretación de los datos.

2.1. CICLO DEL PROCESO

Al iniciar algún proceso de análisis de datos, siempre es necesario hacer una preparación de los datos con que se trabajará. Es común trabajar con datos imperfectos, muchas veces los datos pueden no haberse tomado correctamente o no estar ordenados ni bien completados. Esto provoca datos en blanco, erróneos, contradictorios, redundantes o con una naturaleza diferente. Sin embargo, con las herramientas adecuadas, estos datos

pueden ser trabajados para darles más coherencia, es decir, se les debe hacer un preprocesamiento: una limpieza, imputación, edición, integración, reducción o transformación a los datos, según sea la necesidad. El proceso de preprocesamiento es muy importante, pues el no realizarlo podría provocar que se consideren datos erróneos y se obtengan análisis poco descriptivos o inadecuados.

Dependiendo de la naturaleza de los datos iniciales, podría ser necesario normalizarlos, o sea, llevar todos los valores a una escala de 0 a 1 y; estandarizarlos, llevar la varianza de los datos también en una escala de 0 a 1. Esto es útil para que todas las dimensiones del registro tengan la misma relevancia y, poder encontrar patrones en los análisis sin que una de las dimensiones con valores muy grandes acapare todo el algoritmo. Luego de preprocesar y estandarizar los datos si fuese necesario, siempre es útil visualizarlos. Lo ideal es elegir herramientas que permitan visualizar los datos de manera creativa y que puedan lidiar con la cantidad de dimensiones con las que se esté trabajando. De esta manera, y sumándole los conocimientos sobre estadísticas y el área en la que se trabaja, será posible conocer de manera más profunda los datos y su comportamiento.

Los modelos normalmente siguen alguna de estas dos técnicas [8]: aprendizaje supervisado, donde cada dato tiene asociado un rótulo o, no supervisado, que tiene datos sin rotular. En el aprendizaje supervisado, se asume que se comienza con un conjunto de datos previamente etiquetado, esto quiere decir, que para cada dato de entrenamiento se conoce el valor de su atributo objetivo. Esto le permitirá al algoritmo "aprender" una función capaz de predecir el atributo objetivo para un conjunto de datos nuevos. Algunos ejemplos de aprendizaje supervisado son los modelos de clasificación y regresión lineal. Por otra parte, los métodos no supervisados son algoritmos que basan su proceso de entrenamiento en un conjunto de datos sin etiquetas, es decir, no se conoce ningún valor objetivo. El aprendizaje no supervisado es más complejo y, está dedicado a las tareas de agrupamiento, también llamadas clustering o segmentación, donde su objetivo es encontrar grupos similares en el conjunto de datos. Ejemplos de este tipo de métodos serían el agrupamiento jerárquico y algoritmos de clustering. Finalmente,

siempre es bueno comunicar los resultados, verificar en que medida se alcanzaron los objetivos y difundir los resultados.

Especificamente en este trabajo de memoria, se preprocesa la data rellenando los datos faltantes con ceros, se normalizan y estandarizan los datos y, se visualizan mediante algoritmos. El problema se trata de un aprendizaje no supervisado, ya que se quiere predecir el comportamiento del molino SAG, sin conocer cual es el modo de operación del molino a lo largo del tiempo.

2.2. TIPOS DE ANÁLISIS

Existen diferentes tipos de análisis, según lo que el problema requiera [9]. Se observa una ilustración representativa de estos análisis en la Figura 2.6 y se describe cada uno de ellos [10].

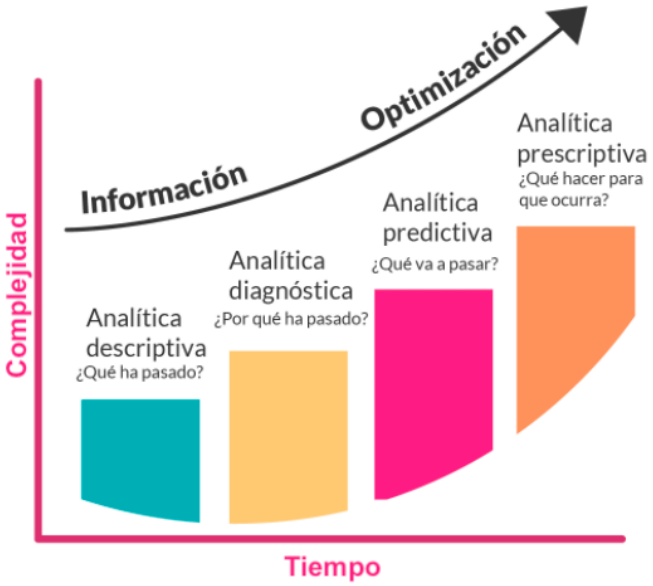


Figura 2.6: Relación entre el tiempo y la dificultad de los procesos de análisis.

El análisis descriptivo se utiliza en eventos pasados. Esta técnica permite analizar una gran variedad de datos históricos (estructurados o no estructurados). Con diferentes métodos estadísticos se recopila toda la información relevante a la situación, se analiza

y se obtiene información sobre lo que sucedió. Está en el primer nivel de complejidad, pues la situación a analizar, es un evento que ya ha sucedido.

El análisis de diagnóstico responde a "¿Por qué sucedió el fenómeno?". En este nivel de complejidad se encuentran las técnicas de inferencia estadística, que mediante la construcción de modelos permite estudiar la relación y comparación de variables para obtener conclusiones útiles.

El análisis predictivo trata de eventos que aún no ocurren y, ya que el futuro siempre presenta algún tipo de incertidumbre, es un análisis desafiante. Por lo tanto, este análisis tiene mayor complejidad que los anteriores debido a la situación dinámica en la que se encuentra. Mediante algoritmos permite identificar patrones, detectar tendencias y realizar pronósticos para orientar el comportamiento de fenómenos futuros.

El nivel de mayor complejidad es el análisis prescriptivo. Es un sistema que mediante la simulación de escenarios y estimación de probabilidades de ocurrencia, entrega recomendaciones de acciones que se deben tomar para lograr un objetivo deseado. En otras palabras, ayuda a entender qué se debe hacer para obtener los resultados que se quieren en el futuro. Para este análisis, se utiliza aprendizaje de máquinas (o machine learning), que mediante la iteración de estas operaciones logra ajustar los modelos para automatizar los procesos de toma de decisiones.

2.3. PYTHON

Python es un lenguaje de programación muy popular a nivel mundial, principalmente por su estructura y sintaxis que son relativamente fáciles de comprender. Es un lenguaje de propósito general muy versátil. Se utiliza para desarrollar códigos en diversas áreas como, conexión con bases de datos, desarrollo web, desarrollo de aplicaciones, juegos, inteligencia artificial, data análisis, finanzas, etc. También, tiene la particularidad de ser de código abierto y ser mantenible en el tiempo por una comunidad. Esto da mayo-

res certezas de que el software no quedará obsoleto y seguirá funcionando en el futuro, debido a las actualizaciones por parte de sus usuarios. Además, entrega herramientas gratuitas y libres que no necesitan permiso ni licencias. Cabe destacar también, que Python permite una fácil detección de errores, escalabilidad y visualización en comparación con otros programas que se usan para manejo de datos. Python tiene variadas librerías de programación que permiten añadir implementos funcionales de gran ayuda para conseguir códigos más rápidos, cortos y óptimos. Constan de diversos módulos que permiten el acceso de funcionalidades específicas del sistema como entrada y salida de archivos, soluciones estandarizadas a problemas de programación, etc. Estas librerías son un conjunto de funciones desarrolladas por miembros de la comunidad y que han sido disponibilizados para que cualquiera pueda utilizarlo de forma gratuita [11]. Son de fácil acceso y simples de utilizar. Para acceder a estas librerías, es necesario incluirlas en el código con la palabra reservada *import*, como se muestra en la Figura 2.7. Algunas librerías de interés:

- TensorFlow: Para el aprendizaje automático. Puede construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones.
- Matplotlib: Permite la generación personalizada de gráficas en general.
- Numpy: Extensión de funciones matemáticas avanzadas que son más rápidas y utilizan menos memoria.
- Seaborn: Para visualización de datos (basada en matplotlib).
- Scipy: Estadística y algoritmos matemáticos.
- Sklearn: Para aprendizaje automático.
- Pandas: Para manipulación y análisis de datos.

Todo lo anteriormente relatado hace que Python sea un lenguaje de programación recomendable para el área del data science. Pues permite manipular, precisar y visualizar

datos de una manera relativamente sencilla y excepcional. Además, mediante sus herramientas y librerías, permite hacer diversos análisis y obtener variados resultados interpretables.

```
from datetime import datetime
from datetime import timedelta
import math
import numpy as np
import pandas as pd
from pandas import read_excel
import operator
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
```

Figura 2.7: Ejemplo para importar librerías a Python.

2.4. ALGORITMOS UTILIZADOS

Antes de utilizar cualquier tipo de algoritmo, en el presente trabajo se realiza una elección de variables de los molinos SAG considerando su fenomenología. Luego, se hace necesario analizar los datos disponibles con herramientas más sofisticadas, para obtener conclusiones que permitan cumplir con los objetivos planteados en la Sección 1.4.

Primero, se utiliza el Análisis de Componentes Principales o Principal Component Analysis (PCA) en inglés, para reducir la dimensionalidad de los datos. Luego, se utiliza el algoritmo Uniform Manifold Approximation and Projection (UMAP) para visualizar los datos disponibles después de la reducción con PCA. Finalmente, se utiliza K-means para clasificar los datos visualizados por UMAP y, obtener los modos de operación del molino SAG. A continuación se presentan estas tres herramientas de análisis.

2.4.1. ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Para este algoritmo, es fundamental conocer y entender el concepto de reducción de dimensionalidad. Este proceso, realiza una reducción en el número de variables con las que se esté trabajando.

La reducción de dimensionalidad, según [12], se utiliza principalmente porque ciertos algoritmos podrían no ser capaces de procesar una gran cantidad de datos, lo que relentiza o no hace posible los análisis deseados. También considera la redundancia o correlación entre las señales disponibles. Es importante tener una cantidad adecuada de variables para trabajar con data science, ya que, al trabajar con una cantidad pequeña de variables se facilitarán los análisis alivianando la carga computacional del procesamiento. Algunas ventajas para realizar este proceso son, la reducción del espacio de tiempo y almacenamiento requerido, el mejoramiento del rendimiento del modelo de aprendizaje tras la eliminación de multicolienalidad (relación de dependencia lineal fuerte entre más de dos variables) y, se hace más fácil visualizar los datos cuando se reduce a dimensiones muy bajas tales como 2D o 3D.

En la presenta memoria, se trabaja con un dataframe con muchas variables y, se hace necesario reducirlas. En este caso entonces, se considera indispensable utilizar alguna técnica que logre una transformación de datos de un espacio de alta a uno de baja dimensión, como se observa en la Figura 2.8. Además, es importante que la representación de baja dimensión conserve la mayor parte de propiedades significativas de los datos originales, pues es indispensable que las nuevas columnas no pierdan mucha información.

Según [13], la reducción de dimensionalidad con análisis PCA, observa el comportamiento de las variables del punto de vista de la variabilidad y correlación, para así poder notar señales que no son relevantes al análisis y puedan ser descartadas. PCA es un método estadístico que permite simplificar la complejidad de espacios muestrales con muchas dimensiones a la vez, conservando en su mayor parte la información.

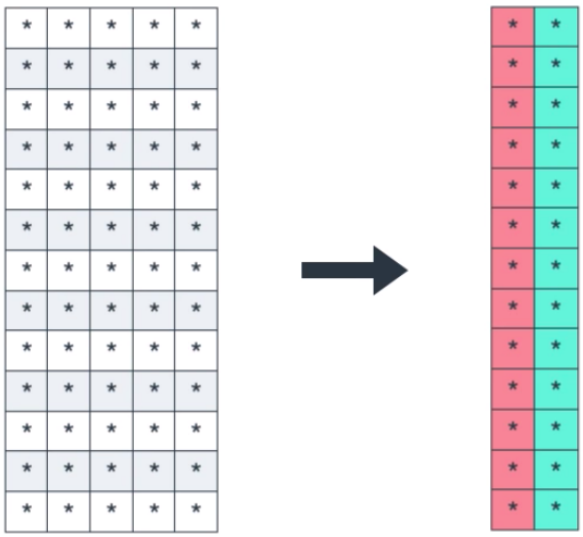


Figura 2.8: Ejemplo reducción de dimensionalidad.

Existen tres conceptos de estadística fundamentales que permiten explicar el funcionamiento del algoritmo PCA para llegar a un espacio de baja dimensión. Estos son; el promedio, la varianza que representa la variabilidad de una serie de datos respecto a su promedio y, la covarianza que determina si existe una dependencia entre las variables.

$$Promedio = \frac{1}{n} \cdot \sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + ... + a_n}{n} \tag{2.1}$$

$$Varianza = \frac{1}{n} \cdot \sum_{i=1}^{n} {a_i}^2 = \frac{{a_1}^2 + {a_2}^2 + ... + {a_n}^2}{n} \tag{2.2}$$

$$Covarianza = \frac{1}{n} \cdot \sum_{i=1}^{n} a_i = \frac{(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n} \tag{2.3}$$

Lo que realiza el algoritmo internamente utilizando los conceptos anteriormente explicados se explica a continuación. Se crea la matriz de covarianza, una matriz simétrica que en su diagonal contiene la varianza de cada una de las dimensiones y, la covarianza entre cada par de dimensiones en el resto de entradas de la matriz. PCA, realiza una

transformación lineal de los datos, y la forma en la que están proyectados los datos cambia. Este cambio es indicado por los valores y vectores propios. A la matriz de covarianza se le calculan los vectores propios, que indican dirección y, los valores propios, que indican magnitud. Estos vectores propios se denominan componentes principales. Luego, se ordenan los valores propios de mayor a menor. Geométricamente la data estará más cerca del plano creado por los vectores propios asociados a los autovalores más grandes. Finalmente, se proyectan los datos con la nueva dimensión reducida y se da origen a la tabla con las componentes creadas. Esto pierde un poco de información, porque si bien la data está más cercana a ciertas dimensiones no lo estará de forma perfecta.

En resumen, PCA es una técnica que describe datos multidimensionales en términos de nuevas variables, llamadas componentes principales. El nuevo sistema ordenado permitirá descartar variables dependiendo de la interpretación que se haga. Esto permite reducir la dimensionalidad asociada a un conjunto de datos. Sin embargo, este enfoque trae consigo una pérdida de información que puede o no ser significativo. En la literatura existen diversos métodos y técnicas como criterio para hacer la reducción de dimensionalidad. A continuación, se describen algunos métodos utilizados en esta memoria [14][15]:

- **Heatmap:** Este diagrama de calor, permite observar datos desde el punto de vista de la variabilidad y la correlación de las variables. La cantidad de componentes principales existentes en este caso, serán la misma cantidad de variables que se estén analizando. El hacer PCA con tal cantidad de componentes, arroja una matriz cuadrada que creará un nuevo sistema coordenado en función de la variabilidad y correlación de los datos. La matriz se crea desde el comando *pca.components_*, con el que se tiene el conjunto de todos los vectores propios para su espacio de proyección. Un ejemplo de este diagrama se observa en la Figura 2.9.

  Se observa graficamente por la barra lateral de la Figura, que los valores del sistema son tanto valores positivos como valores negativos. El signo de los componentes

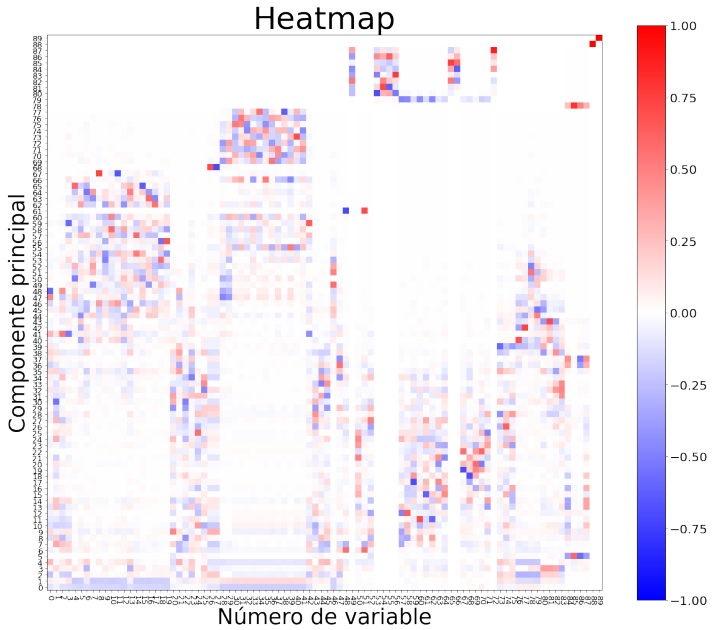


Figura 2.9: Ejemplo Diagrama de Calor (Heatmap).

en cada uno de los ejes calculados informará sobre la correlación de la componente y la variable del sistema. Si el signo es negativo, existirá correlación negativa entre la variable y la componente, esto quiere decir, que al crecer una la otra decrece, y viceversa. Mientras que si el signo es positivo, existirá una correlación positiva, en donde si la variable crece la componente también lo hará.

Por otro lado, el valor numérico representado en colores muestra el valor del peso de cada señal del sistema. Por lo tanto, que el color sea blanco, significará que el peso de la respectiva variable no es relevante en la respectiva componente y, si el color está bien definido que, la variable tiene un peso significativo en la componente.

A modo de ejemplo se puede observar en la Figura 2.9, que la primera componente tiene correlación negativa con casi todas las variables del sistema asociadas, pues están pintadas de azul. Además, esta misma componente no depende de todas las variables del sistema, pues tiene muchos valores cero (color blanco).

- **Scree Plot:** Esta gráfica muestra cuánta variación de los datos captura cada componente principal. Las componentes principales se crean en orden de la cantidad

de variación que cubren. El eje y contiene los valores propios, que representan la cantidad de variación. Con este tipo de gráfico, se pueden seleccionar los componentes principales que se desean conservar. En la Figura 2.10 se observa un ejemplo del Scree Plot con una barra por cada componente principal que indica su variación.

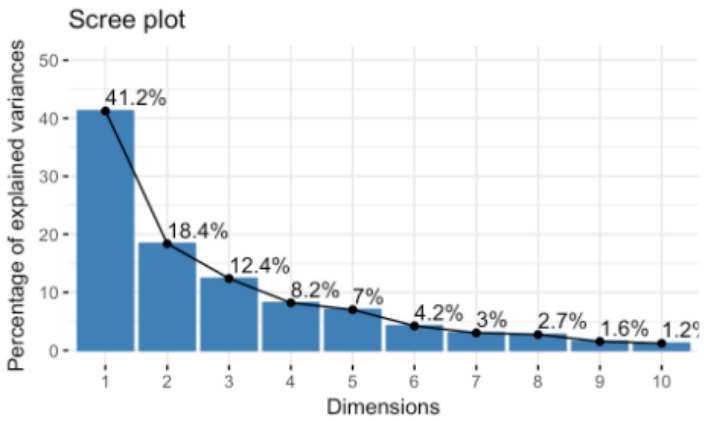


Figura 2.10: Ejemplo Gráfica de Sedimentación (Scree Plot) [16].

- **Score Plot:** El Score Plot es un gráfico de dispersión. Se encarga de indicar grupos de muestras en función de su similitud. Además, muestra la proyección de los datos multidimensionales a un espacio dimensional reducido de 2 o 3 dimensiones. Los componentes principales elegidos para crear el gráfico son los de mayor varianza, o sea, los primeros. A modo de ejemplo, se tiene la Figura 2.11. Se observa el comportamiento de los datos en las primeras dos componentes principales donde hay más varianza, por lo tanto, resulta representativo generar decisiones a partir de esta visualización. Este gráfico no indica las variables a descartar, es una forma de visualizar los datos al reducir la abrumadora cantidad de dimensiones mediante la construcción de componentes principales.

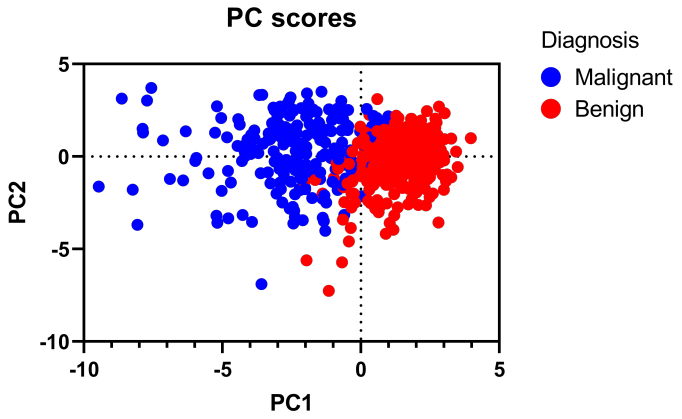


Figura 2.11: Ejemplo Gráfica de Dispersión (Score Plot)[17]

- **Loadings Plot:** Este gráfico se analiza de manera visual. Indica la fuerza con la que cada característica (variable) influye en un componente principal. Un ejemplo se observa en la Figura 2.12.

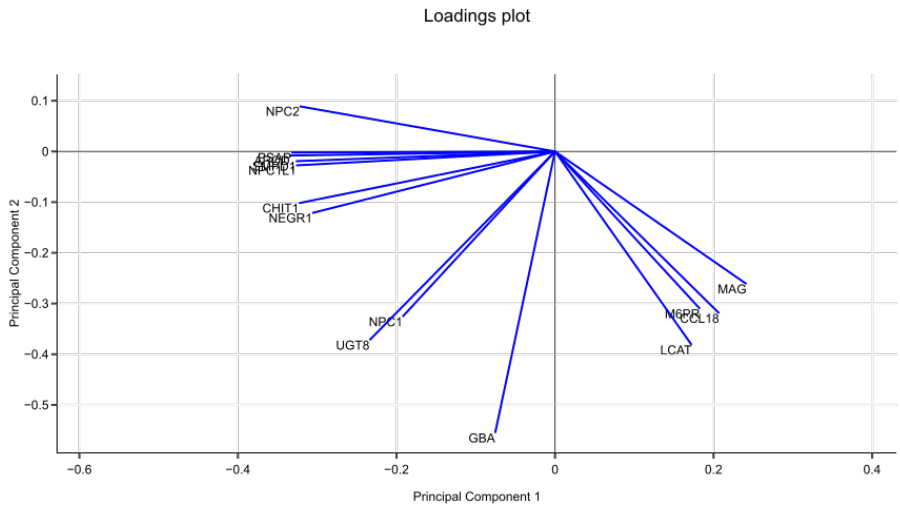


Figura 2.12: Ejemplo Gráfica de Cargas (Loadings Plot) [18]

Las cargas tienen diferentes largos, que pueden variar de -1 a 1. Las más grandes cercanas a los límites, indican que la variable influye fuertemente en el respectivo componente. Las cercanas a cero, indican que la variable tiene una influencia débil sobre el componente. Además del largo, importa el ángulo que haya entre las cargas y los ejes del gráfico. Las variables más cercanas al eje de las ordena-

das influirán más en la segunda componente principal, mientras que las variables cercanas al eje de las abscisas, influirán en la primera componente principal. También, este gráfico indica como las variables se correlacionan entre si dependiendo del ángulo que haya entre las cargas. Si hay ángulos pequeños entre ellas estarán correlacionadas positivamente, si el ángulo es cercano a 90^{o} no estarán correlacionadas y, si el ángulo es cercano a 180^{o} están correlacionadas negativamente.

2.4.2. Uniform Manifold Approximation and Projection (UMAP)

UMAP es una técnica creada para la reducción de dimensionalidad de un conjunto n dimensional de datos, de tal forma que se puedan representar en 2 o 3 dimensiones. Esta técnica mantiene la estructura global de los datos y se logra realizando transformaciones no lineales, para tener una visualización certera del espacio original del conjunto de datos. Para el caso de señales en el tiempo, cada punto resultante de UMAP es un instante de tiempo y cada uno de sus puntos contiene la información del resto de puntos, esto lo hace un método reversible, ya que no hay pérdida de información. Es un algoritmo más complejo que necesita más recursos computacionales en comparación a otros algoritmos como PCA. Por esto, en esta memoria, para ahorrar tiempo en la ejecución del algoritmo UMAP es que antes se realiza una reducción de dimensionalidad con PCA.

En general, UMAP consta de dos procedimientos principales. Por una parte, la construcción de la superficie o Manifold, en donde se crea un nuevo espacio proveniente de la transformación no lineal utilizada para la reducción de dimensionalidad. Por otra parte, la proyección de puntos, que a través del mismo o diferente conjunto de datos aplicando a la transformación no lineal, resulta el gráfico de dos dimensiones [19].

El aplicar UMAP permite representar en una nube de puntos de dos dimensiones todo el conjunto de datos de alta dimensionalidad, como se observa en la Figura 2.13.

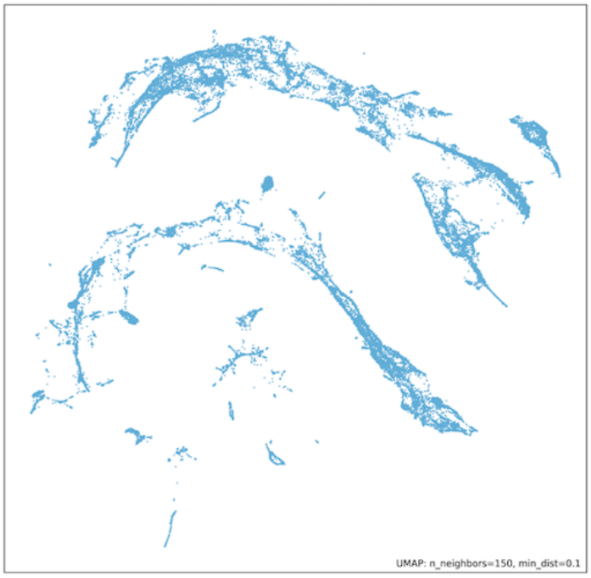


Figura 2.13: Ejemplo nube de puntos. (UMAP)

2.4.3. K-MEANS

El algoritmo de k-medias o k-means permite organizar los datos en diferentes grupos (clusters) dependiendo de sus características. En la Figura 2.14 se observa cada cluster con un color y número diferente. K-means permite agrupar variables y datos en base a características similares o patrones que estos posean, permitiendo una mejor disposición de los datos.

Existe una variedad de métodos de clustering, los cuales varían según el tipo de agrupamiento que se necesita, el parámetro de aplicación del mismo e incluso la dispersión y número de datos. K-means es uno de los más utilizados para agrupación y, se caracteriza por tener una gran escalabilidad de datos. Este algoritmo se centra en particionar un número de datos en k grupos, donde los datos se almacenan dentro de estos asignando el promedio de cada grupo k, lo cual lo hace útil para analizar los gráficos generados por UMAP. La manera en la que opera el algoritmo k-means es la siguiente:

1. Inicialización: Primero, se asignan los grupos k, centrando el valor del centroide a emplear.

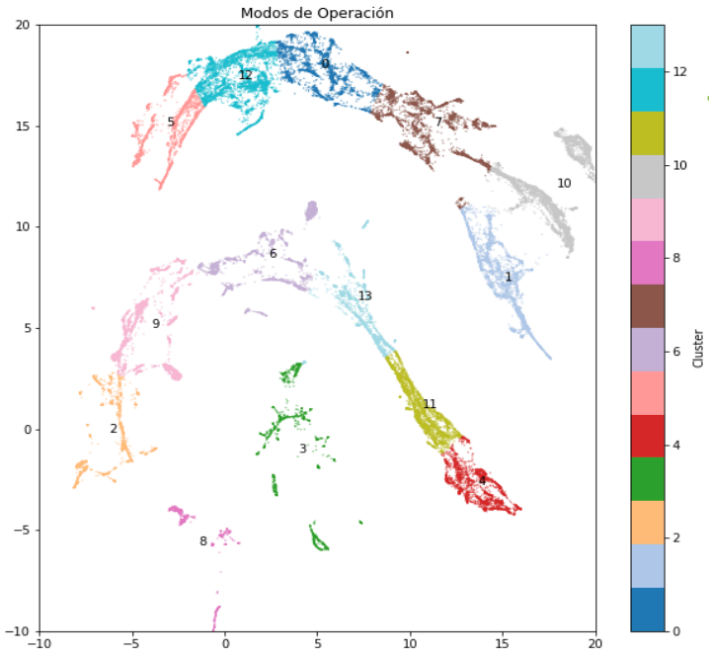


Figura 2.14: Ejemplo clusters. (K-means)

2. Asignación: En base a los centroides empleados, se clasifican los datos en torno a los valores de estos. Esta asignación puede ser realizada minimizando la varianza entre los datos, o utilizando el método de los mínimos cuadrados para asignar estos valores, como está detallado en [20].

3. Actualización: A partir de los valores asignados, se calculan los nuevos centroides, y se repite el algoritmo hasta que se tengan los grupos en base a el método usado en la asignación.

CAPÍTULO 3

REVISIÓN A LA METODOLOGÍA DE ANÁLISIS

Uno de los objetivos específicos de este proyecto es revisar la metodología propuesta en la memoria [2]. Esto para considerar nuevas decisiones que permitan mejoras en el análisis de datos y por lo tanto, en el desarrollo para construir un índice de degradación. Es por esto que en el presente capítulo se explican los detalles encontrados a lo largo de la revisión que afectan en el análisis. Además, se describen algunas de las decisiones tomadas, antes de comenzar con los análisis utilizando los algoritmos mencionados en la sección 2.4.

3.1. OBTENCIÓN VARIABLES INICIALES

En la primera etapa de esta memoria, se dispone de una gran cantidad de datos reales proporcionados por CODELCO, obtenidos desde los molinos SAG en la división de Chuquicamata. Dichas variables corresponden a distintos sensores instalados en el molino, sensores externos del molino relacionados al proceso minero y, pesómetros en las correas transportadoras de mineral, entre otros.

Se debe buscar una manera de discriminar dichos datos, puesto que CODELCO entrega alrededor de 800 variables. Se debe tener en cuenta, que no todas las variables reflejan información relevante para los molinos SAG. Además, al trabajar con demasiadas variables, los algoritmos podrían requerir demasiado tiempo para procesarlas. Es por esto que se hace necesario hacer una selección de las variables iniciales. Cabe destacar, que la lista de 800 variables son las mismas consideradas en [2].

Revisando la metodología propuesta en [2] se observa que la discriminación inicial de las variables fue en base a sus etiquetas. Esto quiere decir que se filtraron las señales

por el nombre de estas. El criterio de selección para las variables a utilizar en el modelo de degradación del molino SAG17 se hizo sobre las variables que tienen "SAG17" en su descripción. Podría ser suficientemente válido pensar que las variables con dicha etiqueta serían las que afectarían principal y directamente al modelo del activo. Para respaldar esta decisión, se tuvo apoyo de los expertos de CODELCO para que, el analizar la degradación a través de los datos escogidos sea coherente con la fenomenología del proceso. Las 61 variables seleccionadas son las enumeradas en la Tabla 3.1.

Tags SAG17		
1. Potencia Mol. Sag 17	22. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	43. Temp Enrollado Motor 1v B Sag 17
2. Pres. Lub. Descsos. Sag2	23. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	44. Temp Enrollado Motor 1w A Sag 17
3. Flujo Mineral A Sag17	24. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	45. Temp Enrollado Motor 1w B Sag 17
4. Flujo Mineral Rec. Sag17	25. Presion Aceite Descanso Alimentacion Lado Norte (Terreno) Sag-17	46. Temp Enrollado Motor 2u A Sag 17
5. Pebbles 16 Transf De S17	26. Presion Aceite Descanso Alimentacion Lado Norte (Terreno) Sag-17	47. Temp Enrollado Motor 2u B Sag 17
6. Pebbles 17 Recir A S17	27. Presion Aceite Descanso Alimentacion Lado Norte (Terreno) Sag-17	48. Temp Enrollado Motor 2v A Sag 17
7. Pebbles Generados Molino Sag-17	28. Presion Aceite Descanso Alimentacion Lado Sur (Terreno) Sag-17	49. Temp Enrollado Motor 2v B Sag 17
8. Pebbles Retornados A Molino Sag-17	29. Presion Aceite Descanso Alimentacion Lado Sur (Terreno) Sag-17	50. Temp Enrollado Motor 2w A Sag 17
9. Ruido Molino Sag 17.	30. Presion Aceite Descanso Alimentacion Lado Sur (Terreno) Sag-17	51. Temp Enrollado Motor 2w B Sag 17
10. Pesómetro Fe-103	31. Presion Aceite Descanso Descarga Lado Norte (Terreno) Sag-17	52. Temp. Aceite Retorno Desc. Alimentacion Sag 17
11. Mineral Grueso Sag17	32. Presion Aceite Descanso Descarga Lado Sur (Terreno) Sag-17	53. Temp. Aceite Retorno Desc. Descarga Sag 17
12. Sp Flujo Mineral A Sag17	33. Set Point Presion Descansos Sag-17	54. Temp. Aceite Salida Sist. Alta Presion Sag 17
13. Sp Solido A Sag-17	34. Temp Aceite Descanso Sag17	55. Temp. Agua Entrada Inter. Calor (Agua-aceite) Sag 17
14. Porcentaje De Solidos Molino Sag-17	35. Temp Aceite Descanso Sag17	56. Temp. Agua Intercamb. Aire-agua Sag 17
15. Sp Velocidad A Sag-17	36. Temp Agua Enfriamiento Motor Entrada 1 Sag 17	57. Temp. Agua Intercamb. Aire-agua Sag 17
16. Veloc. Molino Sag17	37. Temp Agua Enfriamiento Motor Entrada 2 Sag 17	58. Potencia Mol. Sag 17
17. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	38. Temp Aire Circulacion Motor Entrada 1 Sag 17	59. Velocidad Feeder 6 Sag-17
18. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	39. Temp Aire Circulacion Motor Entrada 2 Sag 17	60. Velocidad Feeder 7 Sag-17
19. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	40. Temp Enrollado Motor 1u A Sag 17	61. Velocidad Feeder 8 Sag-17
20. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	41. Temp Enrollado Motor 1u B Sag 17	
21. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-17	42. Temp Enrollado Motor 1v A Sag 17	

Tabla 3.1: Listado de principales variables consideradas para el molino SAG17

La elección junto con los operadores y expertos, indican las variables que consideran importantes en la práctica para estimar, en base a su experiencia, el estado del molino. Por todo lo anterior, se consideran para este análisis, las variables relacionadas puramente con las entradas, salidas y parámetros del molino. Algunas de ellas son de potencia eléctrica, velocidad de giro del motor, presión del aceite en los descansos, alimentación de carga, flujo de mineral de entrada, flujo de pebbles generados, temperaturas dentro de los motores, temperaturas en los aceites de lubricación, flujo y temperatura de aire de ventilación, etc. Se considera válido entonces, hacer la selección mencionada de variables. Por lo que en la réplica de la metodología en el molino SAG16, se utilizan las mismas variables que en la Tabla 3.1, pero para este otro activo. Al momento de buscar las variables equivalentes del SAG17 en el SAG16 en la lista de aproximadamente 800 variables, se observa que faltan algunas. Con la ayuda del Director de Analítica Avanzada de CODELCO, se obtienen algunas de las variables faltantes que tienen un nombre diferente. Sin embargo, hay variables consideradas en el molino SAG17 que no tienen equivalencias en el molino SAG16. Es decir, no existen todas las variables del SAG17 en el SAG16. Esto hace que la cantidad inicial de variables a utilizar en el análisis del activo SAG16 sean 53, las que se detallan en la Tabla 3.2.

Cabe destacar que la información disponible de estas variables está muestreada cada un minuto desde Enero del 2020 hasta Noviembre del 2021, o sea, 22 meses. Esto para ambos activos. En la Figura 3.15, se presentan gráficas que describen la variabilidad y correlación entre las 61 variables del molino SAG17 y, las 53 variables del molino SAG16. Se observa en el heatmap un grupo de variables con peso en las componentes principales de la 3 hasta la 30, aproximadamente, y otro grupo con más peso en el resto de las componentes principales. Este resto de variables son mayoritariamente temperaturas del motor de los molinos y presiones de los aceites.

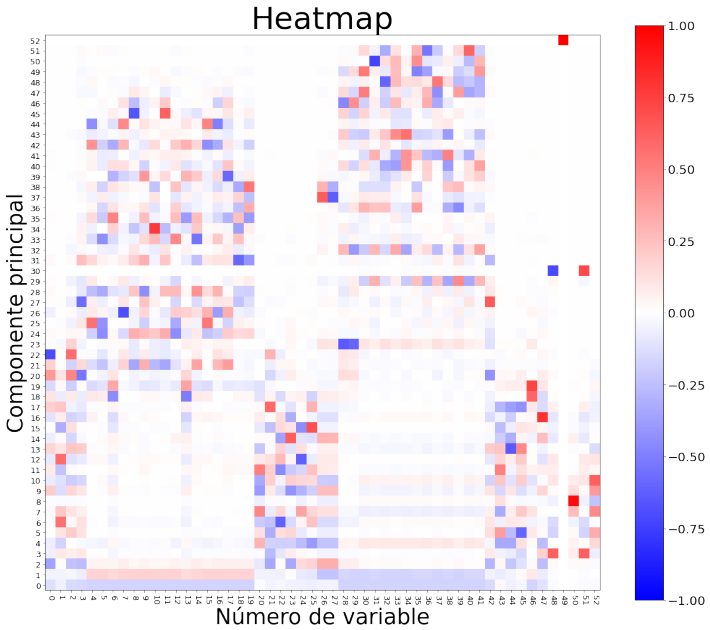
Si bien las variables consideradas hasta ahora son válidas para desarrollar la metodología, se busca añadir más información teniendo en cuenta la fenomenología del proceso. Ya que con una mayor cantidad de información que describa el desgaste del revesti-

Tags SAG16		
1. Flujo Mineral A Sag16	19. Presion Aceite Descanso Descarga Lado Norte (Terreno) Sag-16	37. Temp Enrollado Motor 2u A Sag 16
2. Mineral Grueso Sag16	20. Presion Aceite Descanso Descarga Lado Sur (Terreno) Sag-16	38. Temp Enrollado Motor 2u B Sag 16
3. Porcentaje de solidos Molino SAG-16	21. Ruido Sag-16	39. Temp Enrollado Motor 2v A Sag 16
4. Potencia Mol. Sag 16	22. Sp Flujo Mineral a Sag16	40. Temp Enrollado Motor 2v B Sag 16
5. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	23. Sp Solido a Sag-16	41. Temp Enrollado Motor 2w A Sag 16
6. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	24. Sp Velocidad a Sag-16	42. Temp Enrollado Motor 2w B Sag 16
7. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	25. Temp Aceite Descanso Sag16	43. Veloc. Molino Sag16
8. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	26. Temp Aceite Descanso Sag16	44. Velocidad Feeder 3 Sag-16
9. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	27. Temp Agua Enfriamiento Motor Entrada 1 Sag16	45. Velocidad Feeder 4 Sag-16
10. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	28. Temp Agua Enfriamiento Motor Entrada 2 Sag 16	46. Velocidad Feeder 5 Sag-16
11. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	29. Temp Aire Circulacion Motor Entrada 1 Sag16	47. Pres. Lub. Descsos Sag1
12. Presion Aceite Descanso Alimentacion (Sala De Lubricacion) Sag-16	30. Temp Aire Circulacion Motor Entrada 2 Sag16	48. Flujo Mineral Pebbles Generados SAG 16
13. Presion Aceite Descanso Alimentacion Lado Norte (Terreno) Sag-16	31. Temp Enrollado Motor 1u A Sag 16	49. Pebbles 16 Recir A S16
14. Presion Aceite Descanso Alimentacion Lado Norte (Terreno) Sag-16	32. Temp Enrollado Motor 1u B Sag 16	50. Pebbles 17 Transf De S16
15. Presion Aceite Descanso Alimentacion Lado Norte (Terreno) Sag-16	33. Temp Enrollado Motor 1v A Sag 16	51. Pebbles 16 Transf De S17
16. Presion Aceite Descanso Alimentacion Lado Sur (Terreno) Sag-16	34. Temp Enrollado Motor 1v B Sag 16	52. Pebbles Recirculados SAG 16
17. Presion Aceite Descanso Alimentacion Lado Sur (Terreno) Sag-16	35. Temp Enrollado Motor 1w A Sag 16	53. Pesómetro Fe-104
18. Presion Aceite Descanso Alimentacion Lado Sur (Terreno) Sag-16	36. Temp Enrollado Motor 1w B Sag 16	

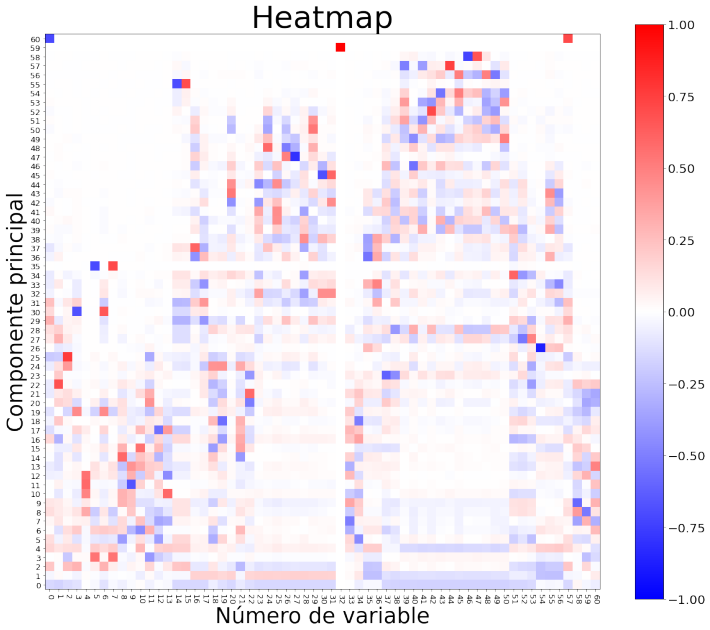
Tabla 3.2: Listado de principales variables consideradas para el molino SAG16

miento de los molinos, se podrían caracterizar de mejor manera los modos de operación de los activos.

En base a conversaciones con profesores del DIMM (Departamento de Ingeniería Metalúrgica y de Materiales) de la Universidad Técnica Federico Santa María, se cree útil y necesario añadir información sobre las personas que operan los molinos SAG. De estas conversaciones se conocen experiencias y evidencias de que, la manera en la que se opera la maquinaria podría influir en la tasa de desgaste del molino. Esto porque puede haber personas que operen los molinos con diferente agresividad, o sea, no todos los operadores manipulan el molino de igual forma. Es un factor humano que cambiará



(a) Heatmap SAG16 con variables iniciales.



(b) Heatmap SAG17 con variables iniciales.

Figura 3.15

dependiendo del operador.

Parte del dataframe entregado por CODELCO con la información de los operadores, se observa en la Figura 3.16. Se debe tener en cuenta algunas consideraciones al utilizar

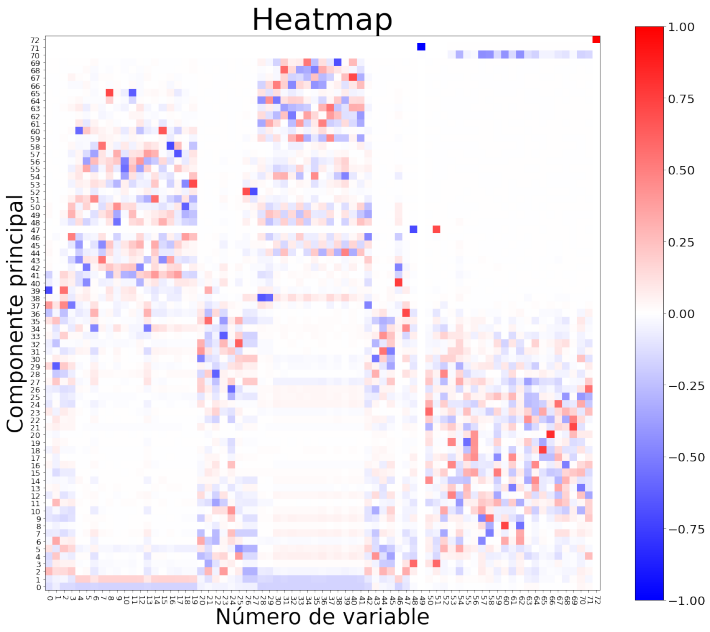
esta información. La información obtenida es sobre ambos activos, es decir, el operador de turno está al mando de ambos molinos al mismo tiempo. Por lo que servirá la misma información tanto para el molino SAG16 como para el molino SAG17. Además, se hace el supuesto de que el respectivo turno terminará al iniciar el siguiente. De esta manera, no habrá problemas en que hayan dos operadores en el mismo instante ni haya momentos sin un operador cuando se añada la información al dataframe con las variables iniciales. También, se realiza un preprocesamiento al dataframe de la Figura 3.16. Se hace un filtro con las filas que contengan "MA2" en la columna tag_id, esto indica que la información es sobre la planta A2, donde se ubica la sección de molienda de interés. En segundo lugar, se completan los datos "NaN" con el operador de la fila siguiente.

	timestamp	tag_id	actual	recomendado	username	updated_at
0	2021-11-12 08:00:00.000	ma2_3si254	10,1103485891	10,76	Manriquez Rodriguez Sebastian (Codelco-Chuquic...	2021-11-12 09:29:59.459
1	2021-11-12 08:00:00.000	fa0_4lic0152	258,1489445729	198	Ramirez Arancibia Walter (Codelco-Chuquicamata)	2021-11-12 08:10:47.114
2	2021-11-12 08:00:00.000	calc_sum_s4_tph	205,1793066014	253	Lagos Suarez Sergio Fernando (Codelco-Chuquica...	2021-11-12 08:03:57.599
3	2021-11-12 08:00:00.000	ma0_2pic27d_op	82,9900665283	82	Lagos Suarez Sergio Fernando (Codelco-Chuquica...	2021-11-12 08:13:37.514
4	2021-11-12 08:00:00.000	fa2_3ai1301	10,6760560547	11,838	Ramirez Arancibia Walter (Codelco-Chuquicamata)	2021-11-12 08:17:33.596
...	...	...	...	...	...	...
56824	2020-11-17 08:00:00.000	fa2_3lc1338	194,9941041598	195,8341318162	NaN	NaN
56825	2020-11-17 08:00:00.000	ma1_xpsc1425	80,9838377446	80,3752605066	Jefes de Turno Concentradora (Codelco-Chuquica...	2020-11-17 12:14:08.530
56826	2020-11-17 08:00:00.000	calc_sum_s5_tph	209,1756746463	225,3842315674	Jefes de Turno Concentradora (Codelco-Chuquica...	2020-11-17 12:05:44.068
56827	2020-11-17 08:00:00.000	calc_re_8fic13_dosif	23,6085949543	23,3769350225	Carrillo Rojas Daniela (Codelco-Chuquicamata)	2020-11-17 12:21:42.845
56828	2020-11-17 08:00:00.000	fa1_7lc1219	89,9598987693	80,2214827572	Carrillo Rojas Daniela (Codelco-Chuquicamata)	2020-11-17 12:01:06.074

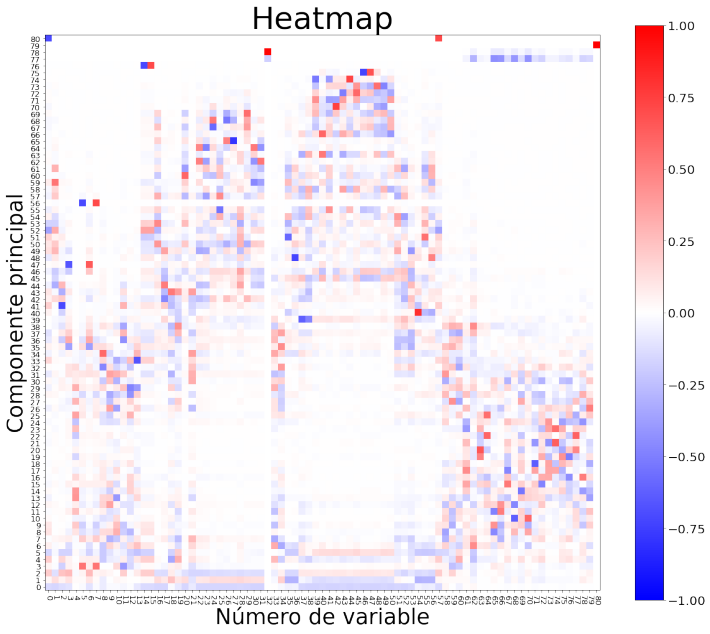
Figura 3.16: Dataframe con información de operadores.

La cantidad de columnas, y por lo tanto de variables, que se añaden al dataframe inicial, son 20 pues corresponde una columna por operador. Cada una de las 20 nuevas variables, indica el momento en el que un operador utiliza los molinos. Las columnas contienen un 1 al momento en el que la persona está operando y el resto de instantes de tiempo en los que no opera se rellenan con ceros. Cabe destacar, que el añadir esta información hace que se acorte la ventana de tiempo de datos disponibles. Es decir, de los 22 meses de datos disponibles, ahora se dispone solo de 12 meses, de noviembre del 2020 a noviembre del 2021, para ambos molinos.

En la Figura 3.17, se observa la gráfica de variabilidad y correlación tanto del molino SAG16 como del SAG17. Ambas realizadas considerando la información de los operadores.



(a) Heatmap SAG16 con operadores.



(b) Heatmap SAG17 con operadores.

Figura 3.17

Las variables con la información de los operadores tienen peso mayoritariamente en

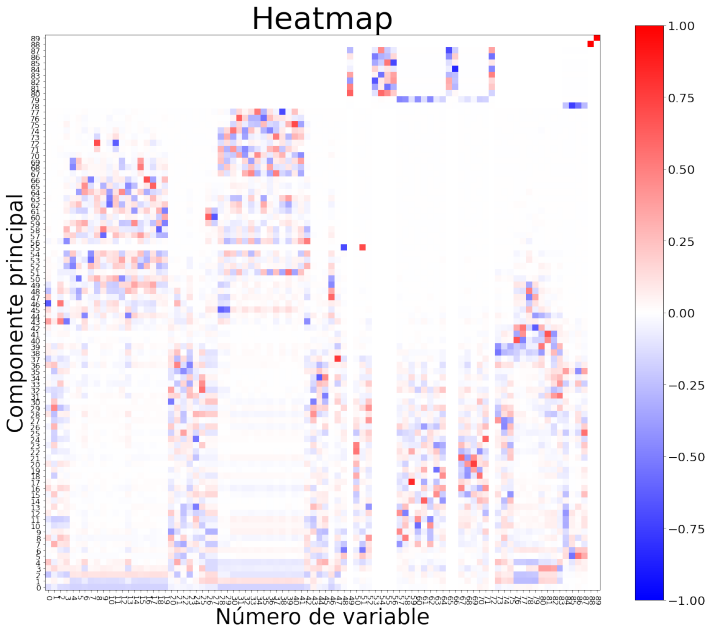
las componentes principales donde no tienen peso las temperaturas del motor de los molinos y las presiones de los aceites. Esto tiene sentido pues, la manipulación de los operadores no influye y, por lo tanto, no está correlacionada con las temperaturas ni presiones, ya que estas variables dependerán de si el molino está encendido o apagado.

Luego, se añade más información que aporta en el desgaste del revestimiento de los molinos SAG. Se trata especificamente de 17 variables, donde hay información sobre el origen, el tipo (cobre o molibdeno), la dureza y la granulometría del mineral. Con respecto al origen, los datos disponibles aportan información sobre el mineral que sale de diferentes fuentes. A la división Chuquicamata llega mineral proveniente de la división Radomiro Tomic también ubicada en la ciudad de Calama, de la mina subterránea y del rajo. En cada uno de estos escenarios el mineral es distinto, por lo que la utilización de los molinos puede cambiar y por lo tanto, también lo hará la tasa de desgaste del revestimiento. CODELCO dispone modelos de trazabilidad que cuentan con estimaciones de cuando llega el mineral a la mina. Además, se añaden indicadores muy potentes en metalurgia, como la granulometría, y la dureza del mineral. La granulometría indicará el tamaño del mineral, puede ser fino, intermedio, grueso o estar entre estas clasificaciones. El indicador spi señala la dureza del mineral y, aunque no es tan fácil encontrar en la industria sensores que midan este parámetro, CODELCO tiene una estimación de esta variable que, cabe destacar, puede no ser tan exacta. Sin embargo, la dureza del mineral aporta información útil para conocer el estado de los molinos, pues independientemente del tamaño de las rocas a moler, si es muy dura será dificil triturarla para los molinos.

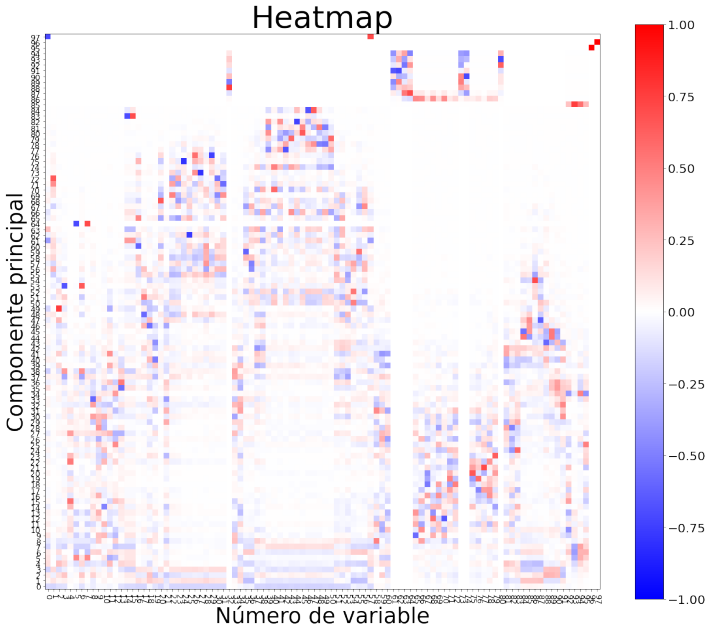
Algunas consideraciones importantes para utilizar la información del mineral detallada en el párrafo anterior, son que al igual que la información de los operadores, las variables nuevas aportan información de ambos molinos, por lo que se utilizan los mismos datos en los dos activos. Además, nuevamente hay un recorte en la ventana de tiempo del análisis. Añadiendo esta información, se trabaja con datos tomados desde Enero del 2021 a Noviembre del 2021, es decir, se cuenta con aproximadamente 10 meses de datos.

Otra consideración importante a tener en cuenta es que las 17 señales incorporadas son estimaciones creadas a partir de modelos desarrollados por CODELCO. Esto puede ocasionar una desincronización entre los datos de los sensores iniciales y esta información nueva que se añade. Por una parte, los sensores de los molinos crean muestras cada un minuto, mientras que las estimaciones con la información del mineral se tienen cada una hora. Esto hizo que se tuvieran que rellenar instantes de tiempo con datos que podrían no corresponder exactamente al mineral que entra al molino.

En la Figura 3.18 se observa el diagrama de calor con la información del mineral incorporada, habiendo un total de 90 variables para el molino SAG16 y 98 variables para el SAG17.



(a) Heatmap SAG16 con información del mineral.



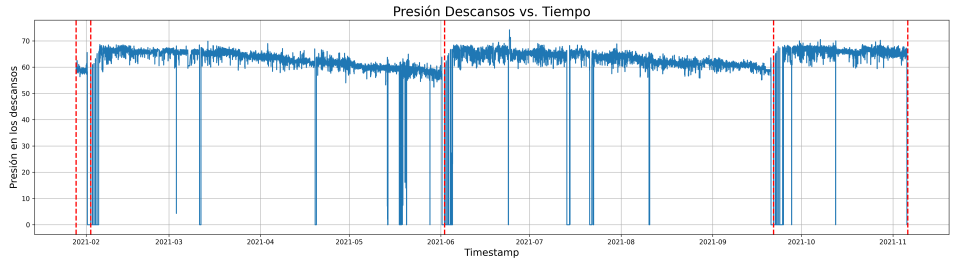
(b) Heatmap SAG17 con información del mineral.

Figura 3.18

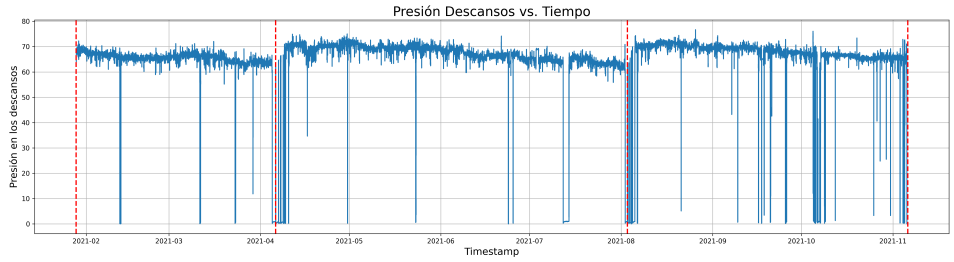
El comportamiento en el heatmap de las variables añadidas es bastante particular, pues no encaja en ninguno de los dos grupos anteriormente mencionados. Por lo tanto, se requiere profundizar en el análisis de otras formas que se verán más adelante. Además,

se observa que hay varias variables de operadores y dos sobre la información del mineral que solo tienen peso en las últimas componentes principales. Esto podría evidenciar un descarte en las variables que tengan peso en estas últimas componentes.

Una consideración importante para el posterior análisis que, al haberse acortado el tiempo de muestras y por lo tanto el largo de la data, se cuenta con menos campañas completas dentro del análisis. Una campaña es el periodo en que el revestimiento del molino SAG funciona antes de ser necesario su reemplazo. Para conocer los inicios y finales de campaña, se grafica en el tiempo la presión de los descansos de cada molino. Esta señal, se muestra en la Figura 3.19, está relacionada directamente con el peso total del molino, pues el revestimiento tiene gran masa. Por lo tanto, los momentos donde los revestimientos de los molinos SAG se cambian, se pueden apreciar como un cambio abrupto en la pendiente que decae, marcado con líneas verticales rojas en la Figura 3.19.



(a) Molino SAG16.



(b) Molino SAG17.

Figura 3.19: Señal presión de los descansos,

A partir de los analisis de la presión en los descansos, se obtienen los inicios y finales de campaña para cada molino SAG que se señalan en las Tablas 3.3 y 3.4.

Campaña	Inicio	Final	Estado
1	28-01-2021	02-02-2021	Incompleta
2	02-02-2021	02-06-2021	Completa
3	02-06-2021	21-09-2021	Completa
4	21-09-2021	05-11-2021	Incompleta

Tabla 3.3: Fechas de inicio y término de las campañas del molino SAG16.

Campaña	Inicio	Final	Estado
1	28-01-2021	06-04-2021	Incompleta
2	06-04-2021	03-08-2021	Completa
3	03-08-2021	05-11-2021	Incompleta

Tabla 3.4: Fechas de inicio y término de las campañas del molino SAG17.

3.2. FILTRO MODO MOLINO APAGADO

Al momento de caracterizar los modos de operación en los molinos, podría haber algún modo que describa al molino cuando se encuentra apagado. Debido a que no hay interés en analizar este estado del molino, se busca un modo para detectar tales datos. Todo esto, debido a que cuando los activos están apagados no hay degradación en el revestimiento. En la revisión de la memoria [2], se utiliza un filtro para dejar los modos de operación del molino apagado fuera del análisis, después de haber utilizado PCA, para no visualizar el modo de operación apagado en UMAP. Sin embargo, se cree que al momento de usar PCA ya se deben filtrar los datos que no interesan, pues se desea analizar la variabilidad de las señales sólo cuando el molino esté funcionando. Por lo tanto, se desarrolla un filtro para detecar cuando el molino está apagado antes de realizar el análisis con algoritmos.

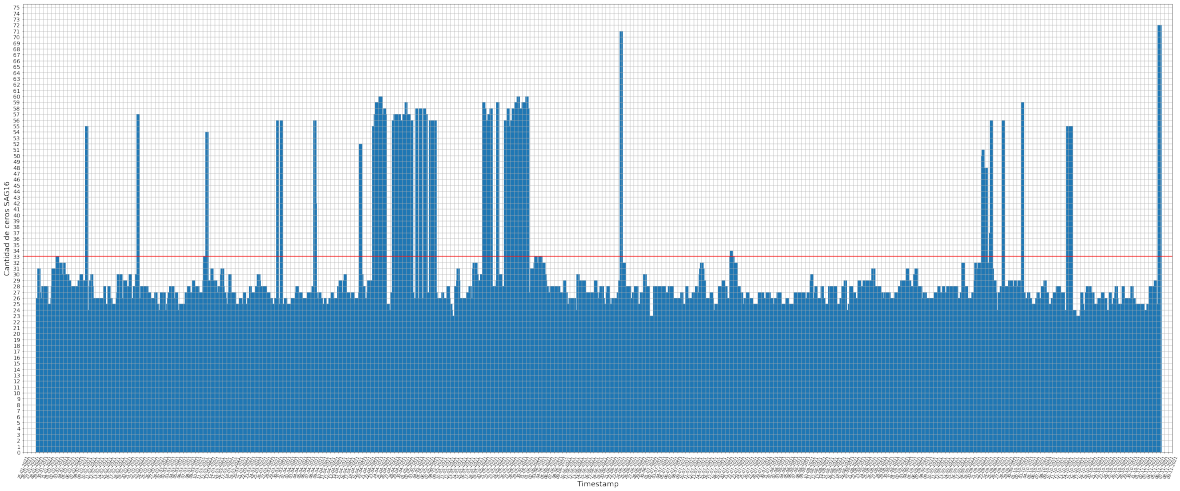
Para la creación del filtro se hace un análisis con la data disponible, se crea un contador de ceros por instante de tiempo, como muestra la Figura 3.20. Los ceros de este contador, equivalen a la cantidad de ceros en cada una de las filas del dataframe utilizado. Cabe destacar, que en algunos instantes hay elevadas cantidades de ceros, esto sucede por la

información de los operadores. Existen 20 operadores, por lo que hay 20 variables en donde solo estará operando una persona a la vez. Por lo tanto, hay 19 ceros en cada una de las filas del dataframe que se refieren a los operadores inactivos.

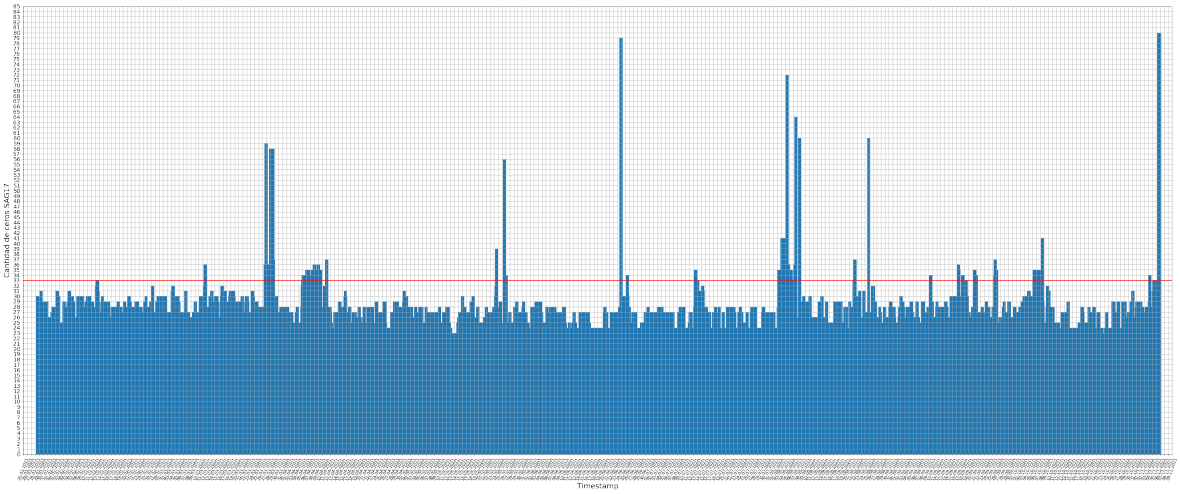
	Timestamp	Cantidad ceros
0	2021-01-28 14:00:40	29
1	2021-01-28 14:01:40	29
2	2021-01-28 14:02:40	29
3	2021-01-28 14:03:40	29
4	2021-01-28 14:04:40	29
...	...	...
405207	2021-11-05 20:55:12	80
405208	2021-11-05 20:56:12	80
405209	2021-11-05 20:57:12	80
405210	2021-11-05 20:58:12	80
405211	2021-11-05 20:59:12	80

Figura 3.20: Dataframe contador ceros.

La Figura 3.21 muestra el valor del contador de ceros en cada instante de tiempo, es decir, en cada minuto.



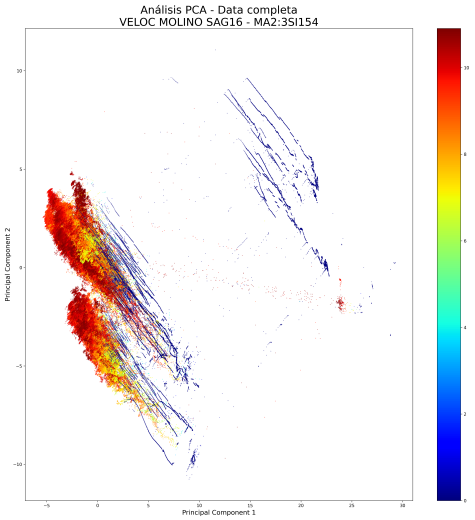
(a) Contador de ceros del molino SAG16.



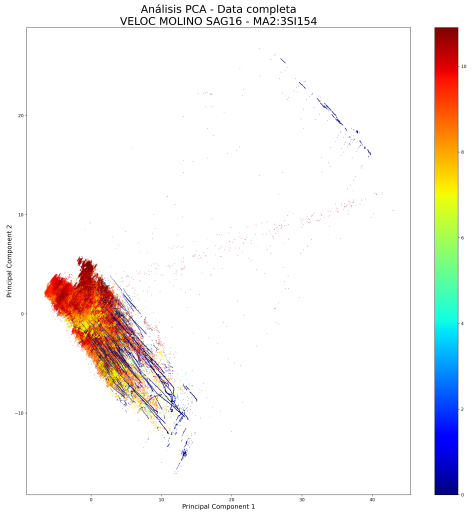
(b) Contador de ceros del molino SAG17.

Figura 3.21

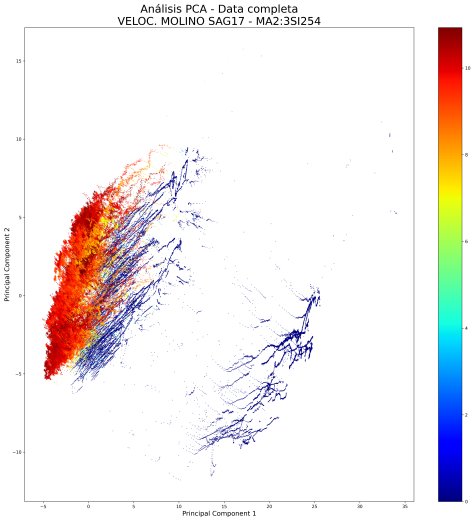
De manera gráfica, se observa que aproximadamente el umbral para detectar que el molino está apagado podría hacerse en 33 ceros. Se señala con la línea horizontal roja en cada Figura. Para verificar que este filtro encontrado tiene coherencia, se realiza un Score Plot antes y después de haber usado el filtro y, se colorea con respecto a la velocidad de cada molino. Estos gráficos se observan en la Figura 3.22.



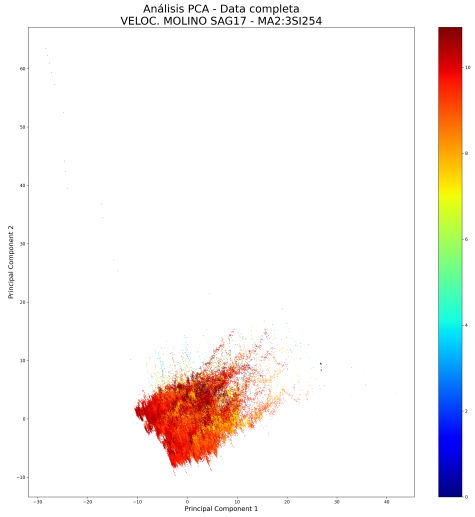
(a) Score Plot coloreado del SAG16 sin filtro.



(b) Score Plot coloreado del SAG16 con filtro.



(c) Score Plot coloreado del SAG17 sin filtro.



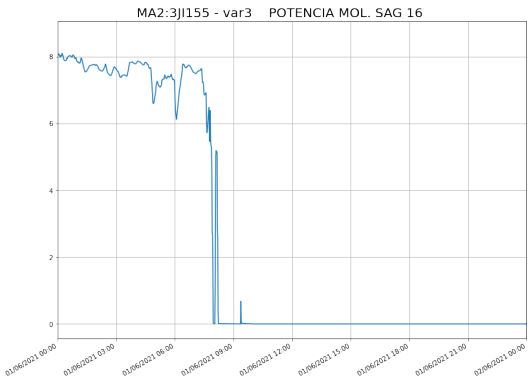
(d) Score Plot coloreado del SAG17 con filtro.

Figura 3.22

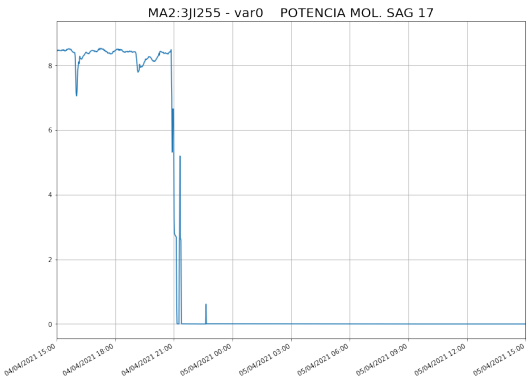
En las gráficas de la Figura 3.22, se observa como dismiuyen los puntos azules de manera crítica al utillizar el filtro. Los puntos azules representan los instantes de tiempo en donde el molino se encuentra detenido.

Luego, para verificar que los momentos que el filtro deja fuera corresponden realmente a que el molino dejó de funcionar, se realiza un análisis exhaustivo de algunas variables

graficadas en el tiempo, observadas específicamente en los instantes de tiempo en los que hay 33 o más variables en cero. Se observa este análisis en los gráficos de las Figuras 3.23, 3.24 y 3.25.

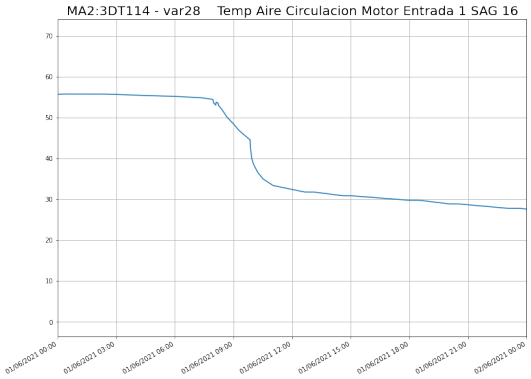


(a) Potencia en el tiempo SAG 16.

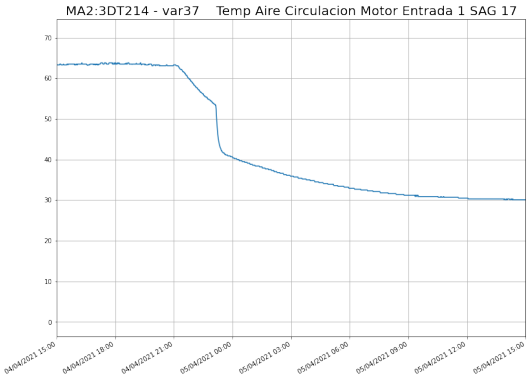


(b) Potencia en el tiempo SAG 17.

Figura 3.23: Gráficas en un instante con 33 o más variables en cero.



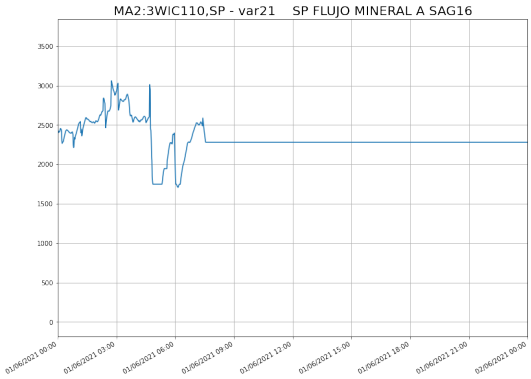
(a) Temperatura en el tiempo SAG 16.



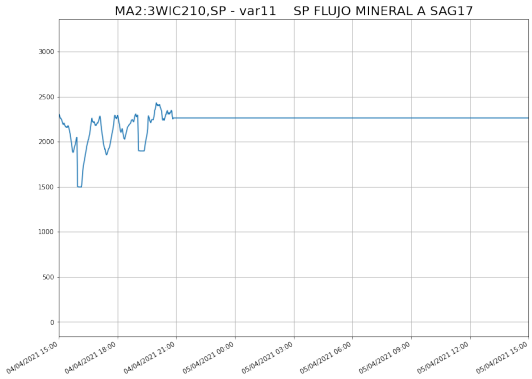
(b) Temperatura en el tiempo SAG 17.

Figura 3.24: Gráficas en un instante con 33 o más variables en cero.

En estas figuras se observa que las variables físicas y mecánicas como la velocidad y la potencia descienden a cero abruptamente. Asimismo se observa, que las variables de temperaturas (que suelen mantenerse constantes) descienden levemente y, en tiempos prolongados llegan a cero. Además, se observa que las variables como SetPoints o volúmenes de mineral permanecen constantes y, en tiempos prolongados descienden lentamente a cero. Todo esto ratifica, que en estos instantes de tiempo donde hay 33 o más variables en cero, efectivamente el molino estaría detenido. Este análisis respalda la elección tomada para el umbral definido para el filtro.



(a) SP Flujo Mineral en el tiempo SAG 16.



(b) SP Flujo Mineral en el tiempo SAG 17.

Figura 3.25: Gráficas en un instante con 33 o más variables en cero.

3.3. SEÑALES EN CERO

De la memoria [2], se rescata el Heatmap de las 61 variables durante todo el año 2019, observado en la Figura 3.26. En este gráfico, se observa dos grupos principales de señales: Hay algunas variables con peso en las primeras 41 componentes principales, y otras con peso en las útimas 20 componentes principales.

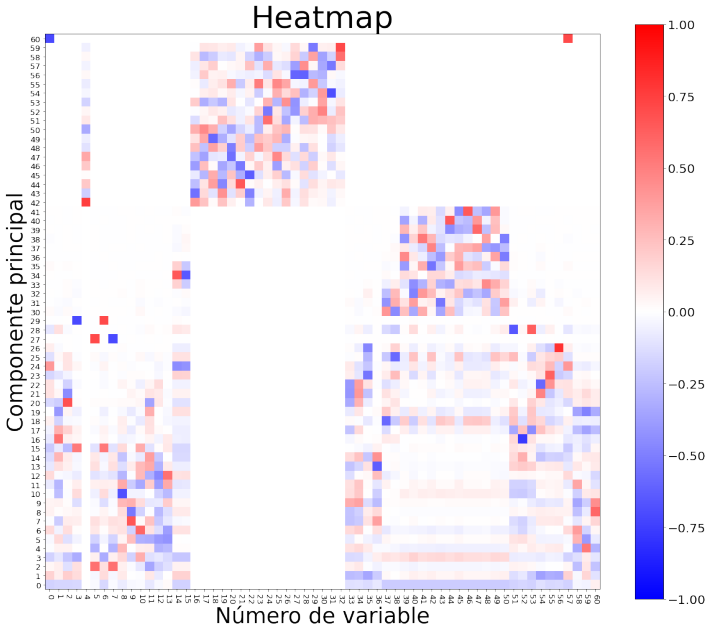


Figura 3.26: Heatmap análisis [2].

Luego, según el Scree Plot de la Figura 3.27, se tiene el ratio de la varianza explicada por

cada componente principal. Esto indica que las últimas 20 componentes casi no aportan información. Por lo tanto, en estricto rigor, se concluye que las variables con peso en las primeras componentes tienen especial relevancia en la nube de puntos, aportando la mayor cantidad de información. Y las otras variables, al estar al final del orden de las componentes, pueden ser descartadas.

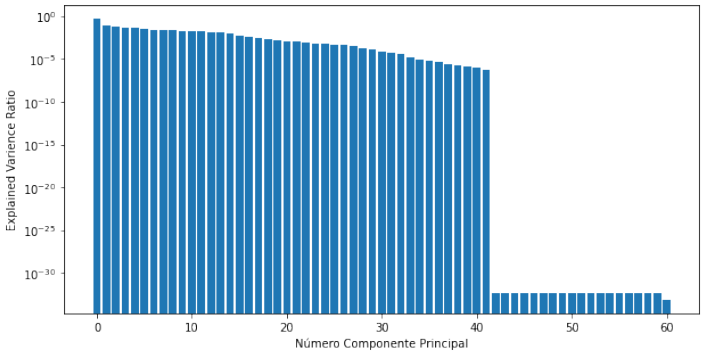


Figura 3.27: Scree Plot análisis [2].

Lo anteriormente descrito resulta algo extraño, pues la discriminación entre los grupos de variables en el heatmap de la figura 3.26, resulta demasiado abrupta. Y, con los análisis realizados hasta ahora, sobre todo en la sección 2 de este capítulo, se sabe que las variables que solo tienen peso en las componentes que no aportan información, tienen poca variabilidad. Es por esto, que se decide hacer un análisis de las variable utilizada en [2] en función del tiempo. De las 61 variables se observan particularmente las 18 que se estarían descartando. Luego, todas las variables que el gráfico de la Figura 3.27 descarta, son como las de la Figura 3.28.

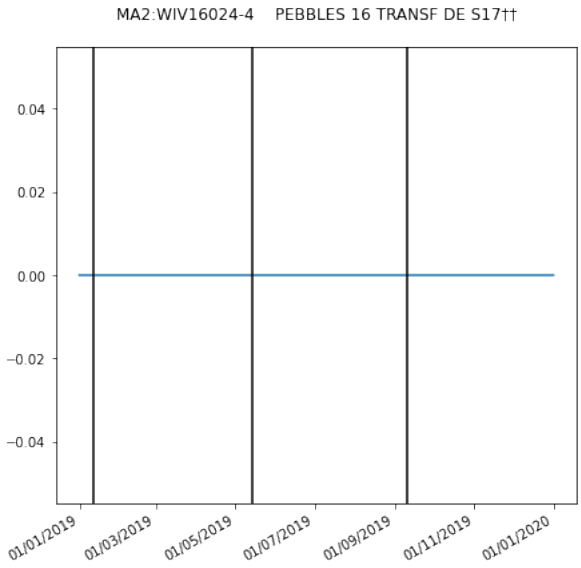


Figura 3.28

Entonces, se deduce que la discriminación abrupta observada en el Heatmap y Scree Plot, se debe a la inexistencia de variabilidad. Las señales descartadas no están aportando ningún tipo de información relevante, puesto que son variables estáticas. Esto hace necesario, poner especial énfasis en el comportamiento de las variables en el tiempo. Además, crea la necesidad de hacer otros análisis que permitan en conjunto, generar decisiones y una discriminación de variables más representativa.

CAPÍTULO 4

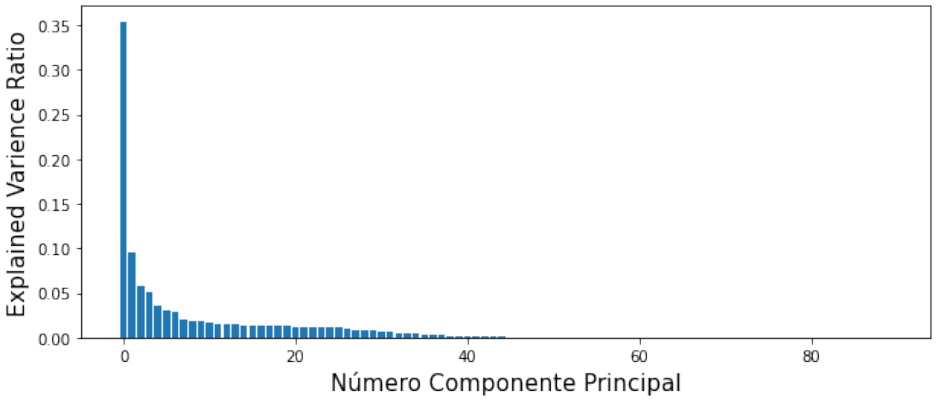
ANÁLISIS CON ALGORITMOS Y OBTENCIÓN MODOS DE OPERACIÓN

En esta sección se detalla el análisis utilizando los algoritmos PCA, UMAP y k-means. Se crean indicadores con PCA que permiten una discriminación de variables que no aporten información relevante para reducir dimensinalidad. Posteriormente se aplica el algoritmo UMAP, se busca algún mapa que represente diferentes modos de operación de los molinos SAG, para luego clasificarlos utilizando k-means.

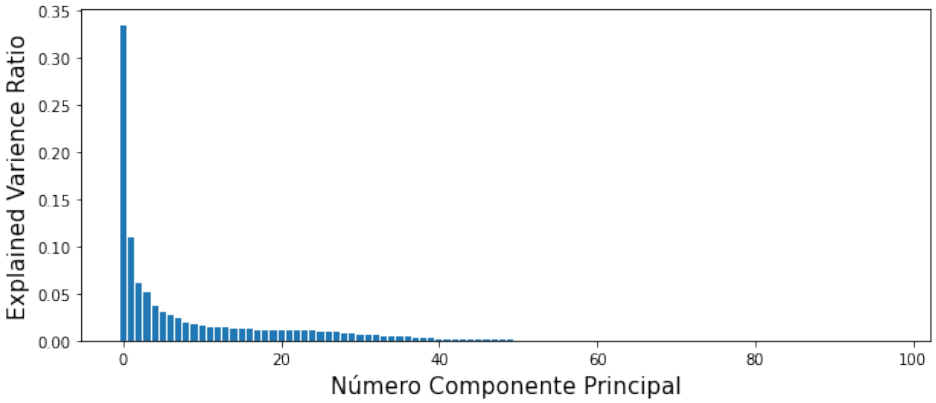
4.1. ANÁLISIS PCA

Al aplicar análisis PCA a los datos disponibles, se tienen los heatmaps de la Figura 3.18, con 90 variables para el molino SAG16 y con 98 variables para el SAG17. Surge la necesidad, como ya se ha mencionado, de reducir la cantidad de datos distinguiendo las variables más relevantes, antes de obtener los modos de operación de los molinos con UMAP.

El primer análisis para la reducción de datos se realiza utilizando los Scree Plot de cada molino, que se observan en la Figura 4.29. Se ve claramente en las imágenes un decaimiento considerable de varianza explicada desde la primera hasta la última componente principal. Esto tiene sentido puesto que PCA crea las componentes principales de esta manera. Para visualizar de mejor forma estos gráficos, se le aplica una escala logarítmica a los valores de varianza. Esto se observa en la Figura 4.30.

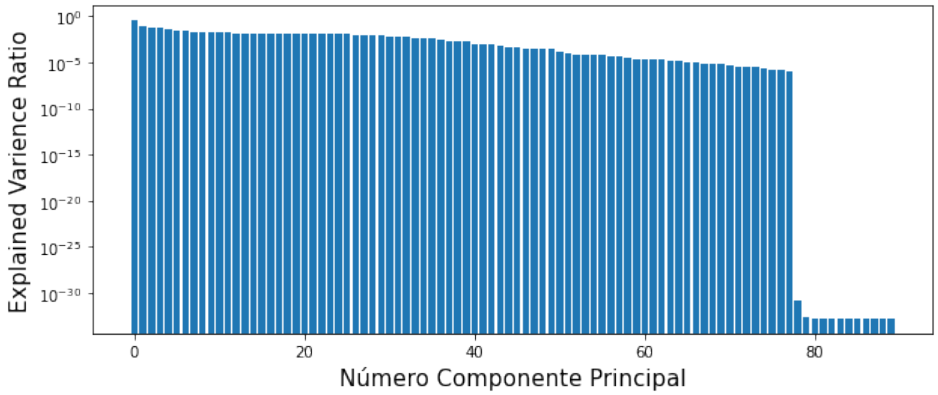


(a) Scree Plot SAG16.

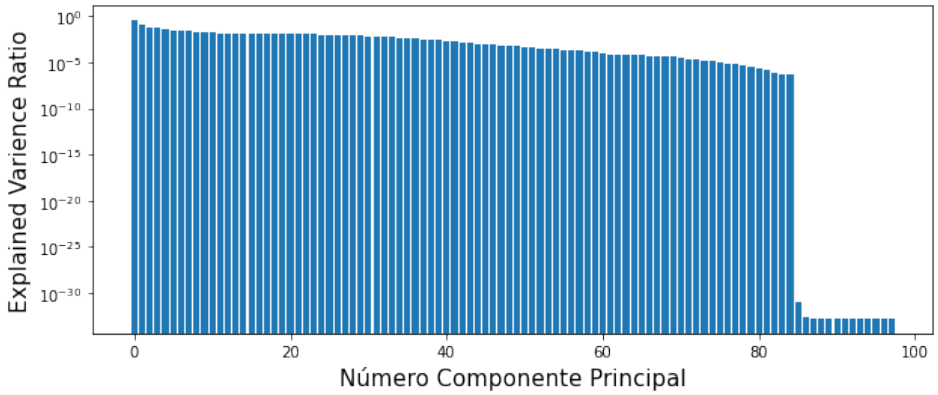


(b) Scree Plot SAG17.

Figura 4.29



(a) Scree Plot SAG16 con escala logarítmica.



(b) Scree Plot SAG17 con escala lograrítmica.

Figura 4.30

Se observa según la figura 4.30, que las últimas 12 componentes en el SAG16 y 13 en el SAG17, practicamente no aportan información. Debido a esto, y a lo concluido en la sección 3.3, se infiere que las variables con peso en esas componentes, podrían quedar fuera del análisis por su poca variabilidad. En la Figura 4.31 se observa el comportamiento en el tiempo que tienen las variables con peso en las componentes principales con poca varianza explicada.

Al ver que las variables efectivamente no varían en lo absoluto, se decide dejar las variables de la Tabla 4.5 fuera del análisis debido a su nula variabilidad. Luego, el molino SAG16 queda con 80 variables y el molino SAG17 con 88 variables. La inexistente variabilidad de los operadores, se debe a que no están operando la maquinaria en la

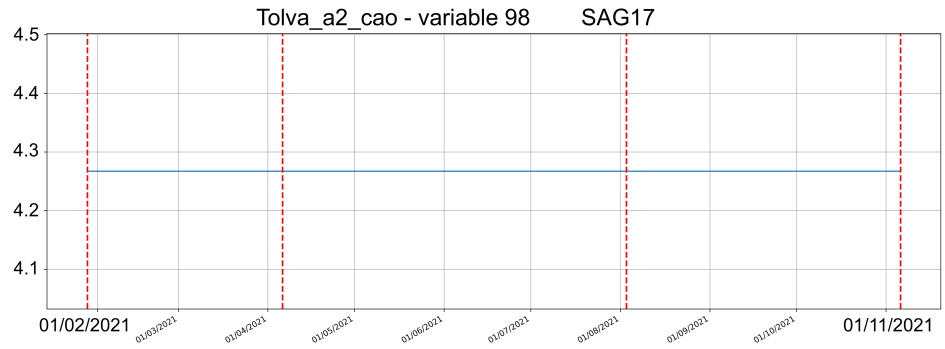


Figura 4.31: Ejemplo de variable descartada.

ventana de tiempo utilizada. El comportamiento de las otras variables excluidas, podría deberse a que los sensores no estén funcionando en la ventana de tiempo utilizada o, incluso a un sensor sin uso. Con esto, se explica el análisis del algoritmo sobre no considerar estas variables para el análisis posterior de los datos.

Tags descartadas SAG16	**Tags descartadas SAG17**
49. Pebbles 17 Transferidos de Sag16	32. SetPoint Presion Descansos Sag-17
53. Correa Mendoza Camila	61. Correa Mendoza Camila
54. Jefes de Turno Concentradora	62. Jefes de Turno Concentradora
55. Escobar Marambio Osvaldo	63. Escobar Marambio Osvaldo
56. Pacheco Araya Oscar	64. Pacheco Araya Oscar
65. Jeraldo Campaña Eduardo	73. Jeraldo Campaña Eduardo
66. Valenzuela Alfaro Alvaro	74. Valenzuela Alfaro Alvaro
72. Peralta Castillo Luciano	80. Peralta Castillo Luciano
88. Indicador tolva esm	96. Indicador tolva esm
89. Indicador tolva cao	97. Indicador tolva cao

Tabla 4.5: Listado de variables descartadas

La reducción de datos utilizando el ratio de la varianza explicada elimina las variables que no aportan información. Sin embargo, se necesita buscar otro método adicional que también descarte variables, pero que considere la existencia de señales que sí podrían estar variando y aún así, no estén aportando información de interés.

4.1.1. INDICADOR PESO

Ante la necesidad de hacer un análisis cuantitativo adicional al análisis visual de los heatmap, en esta sección se crea un indicador numérico. Este indicador debe representar el peso de cada una de las variables, considerando su correlación con cada una de las componentes principales.

El algoritmo propuesto, multiplica la varianza explicada de cada componente por la magnitud del coeficiente en que participa la señal. Luego, se realiza una sumatoria por variable, en donde se suma cada valor multiplicado anteriormente. Esto se replica para cada una de las variables, obteniendo un indicador para cada señal. De esta manera, se tiene una métrica escalar que permite conocer el peso total que tiene cada variable. Para entender el algoritmo de manera gráfica, en el heatmap de la Figura 4.32, cada valor de una columna se refiere al peso de la variable en cada componente. Y cada fila, contiene valores que indican la magnitud del coeficiente en que participa la señal. Cada valor de cada fila, se multiplica por la varianza explicada de la respectiva componente principal, y luego se suman todos los valores pertenecientes a la variable en cuestión. De esta manera, se crea el indicador.

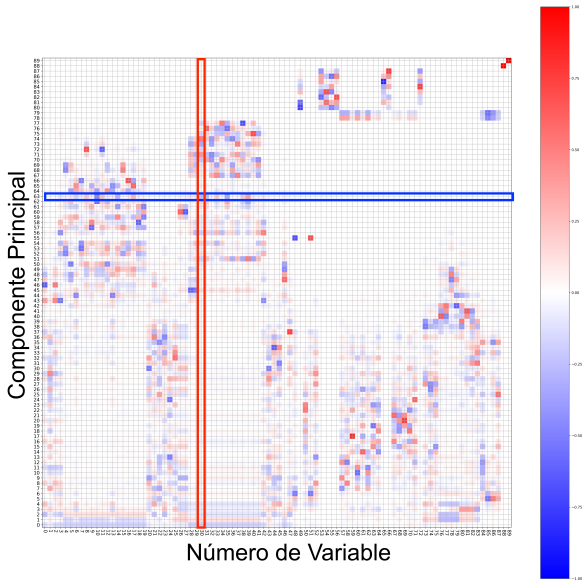
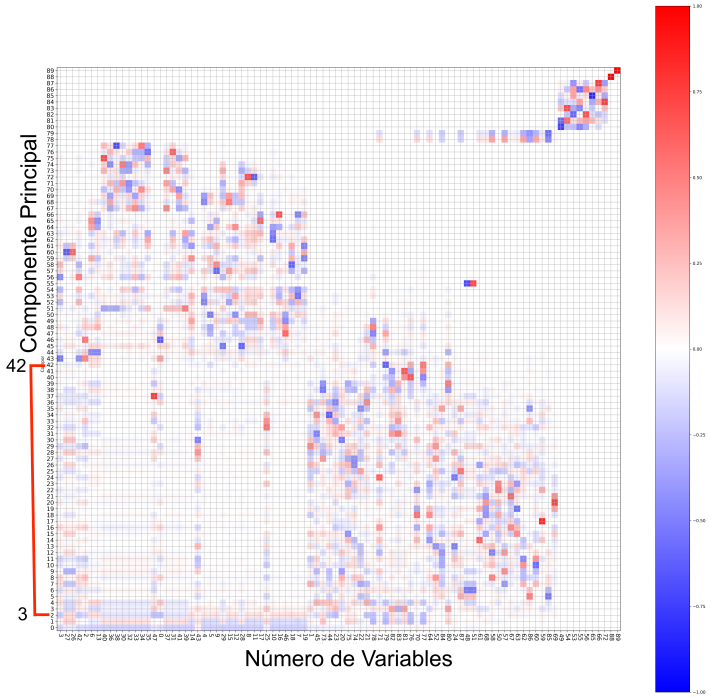
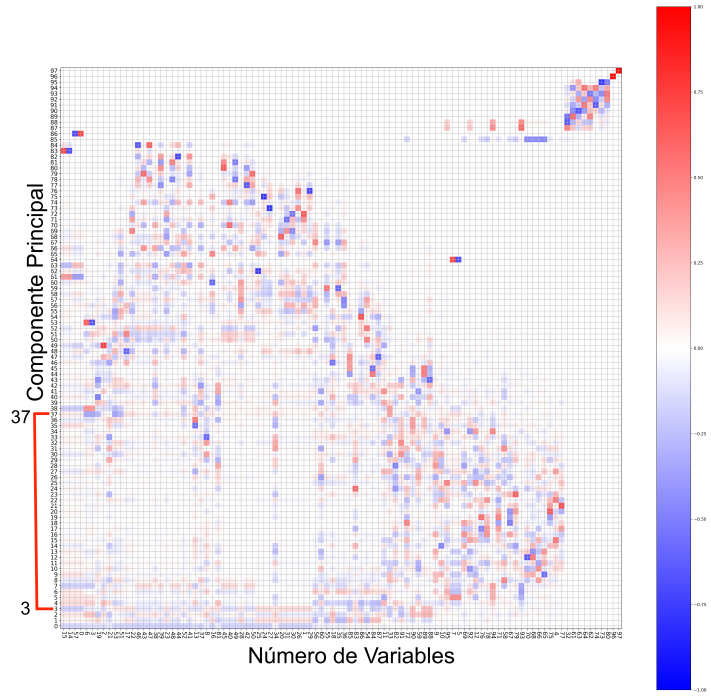


Figura 4.32: Ejemplo del indicador.

Cabe destacar, que lo relatado se aplica tanto al molino SAG16 como al SAG17. Luego de calcular el indicador a cada una de las variables del sistema, se reordenan las variables del heatmap en función del indicador, tal como se ilustra en la Figura 4.33.



(a) Heatmap SAG16 ordenado con indicador.



(b) Heatmap SAG17 ordenado con indicador.

Figura 4.33

De los gráficos de la Figura 4.33, se evidencia que las variables con mayor indicador coinciden con las variables que tienen peso en, aproximadamente las primeras 3 componentes principales. De esto se concluye que estas componentes, no pueden ser eliminadas del proceso. Ya que, independientemente de que tienen, en su mayoría, correlación con las últimas componentes principales, son las variables con mayor indicador debido a la alta varianza explicada de las primeras componentes.

De los gráficos de la figura 4.33 también se observa, que las variables con peso entre las componentes 3 y 42 para el caso del molino SAG16 y, peso entre las componentes 3 y 37 en el caso del SAG17, tienen un indicador más bajo. Esto preocupa por el hecho de que en su mayoría son variables que indican al operador de los molinos, información añadida a modo de mejorar la metodología. Además, se observa claramente de los gráficos, que las últimas variables con menor indicador, son las que ya se habían dejado fuera del análisis pues no tienen variabilidad.

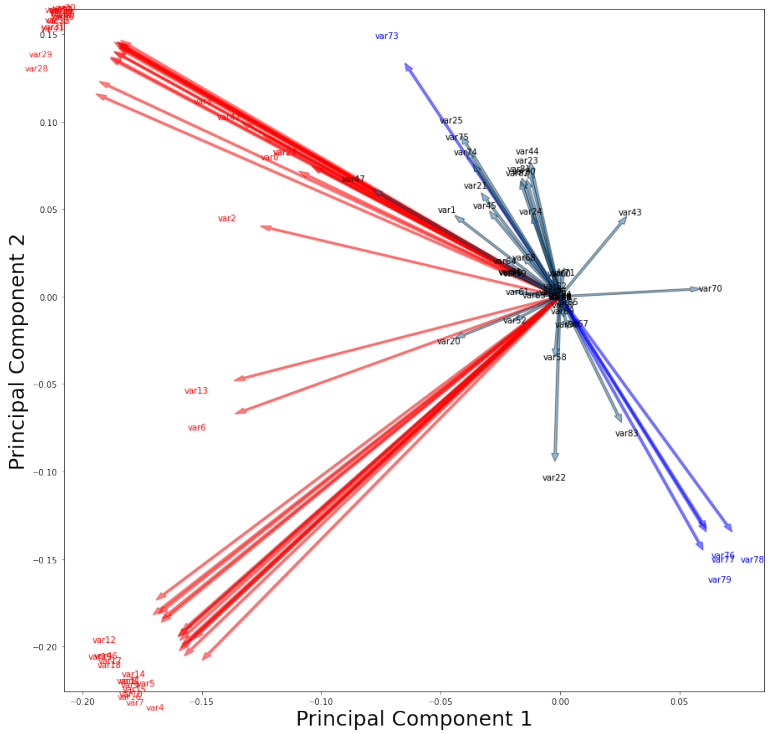
Si bien, este indicador arroja resultados significativos y un criterio importante para decidir sobre las variables a descartar, se siguen necesitando más análisis que permitan determinar las variables a eliminar antes de continuar el análisis con UMAP.

4.1.2. LOADING PLOT

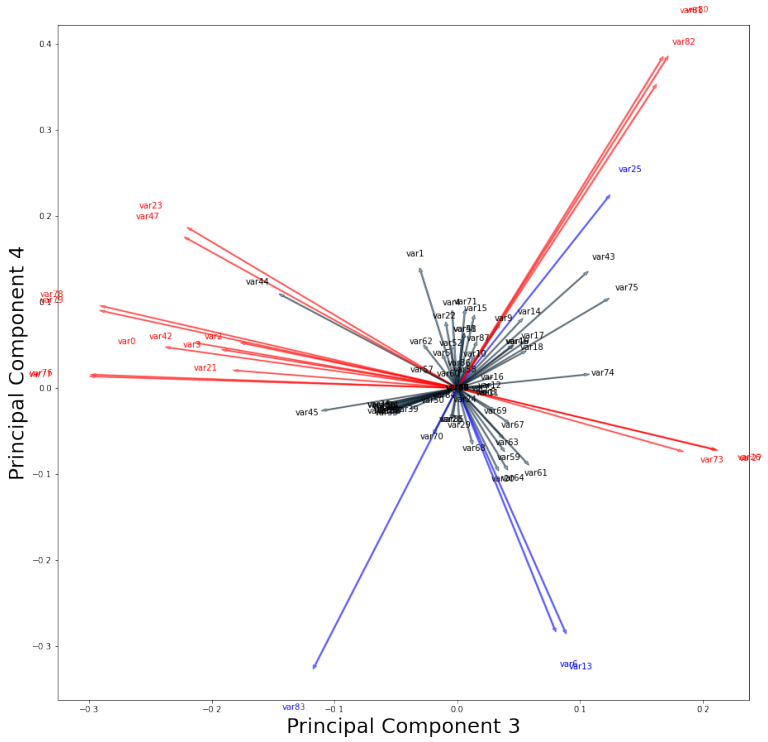
En los Scree Plot de la Figura 4.29, se observa el porcentaje de varianza explicada de cada una de las componentes principales. Para el activo SAG16, se utilizan 4 componentes principales para describir el 54 % de la varianza explicada. Mientras que para el activo SAG17, se utilizan 6 componentes principales para describir el 58 % de la varianza explicada. Se utilizan tal cantidad de componentes, ya que estas logran describir más de la mitad de la varianza explicada total, y por lo tanto, resulta representativo usarlas. Cabe destacar que el hecho de que los molinos necesiten diferentes cantidades de componentes principales para representar un mismo porcentaje de varianza explicada, deja la interrogante de qué tan replicable son los análisis en ambos activos.

A continuación, se utiliza la herramienta Loading Plot. Esta herramienta permite obtener un segundo indicador que establezca otro criterio para la discriminación de variables. Se crean los gráficos Loading Plot de la Figura 4.34 para el molino SAG16 y las Figuras 4.35 y 4.36 para el molino SAG17, utilizando la cantidad de componentes mencionadas anteriormente.

Los vectores coloreados de gris, representan las variables que no se consideran en el análisis, ya que son muy cortos. Los vectores rojos indican las variables que influyen en la componente principal 1, 3 o 5 dependiendo del gráfico. Por su parte, los vectores azules representan las variables que más impactan en las componentes 2, 4 o 6 dependiendo del gráfico.

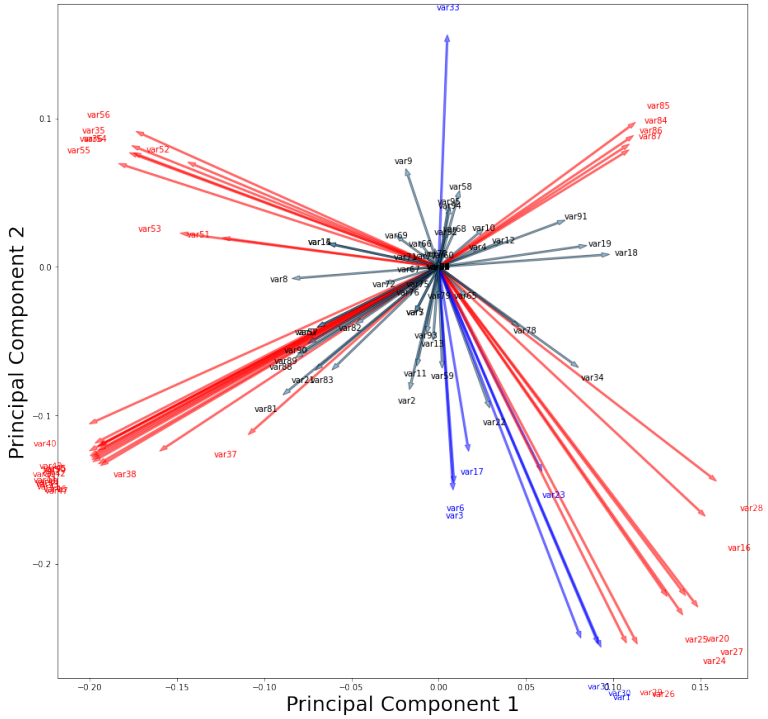


(a) PC1 vs PC2.

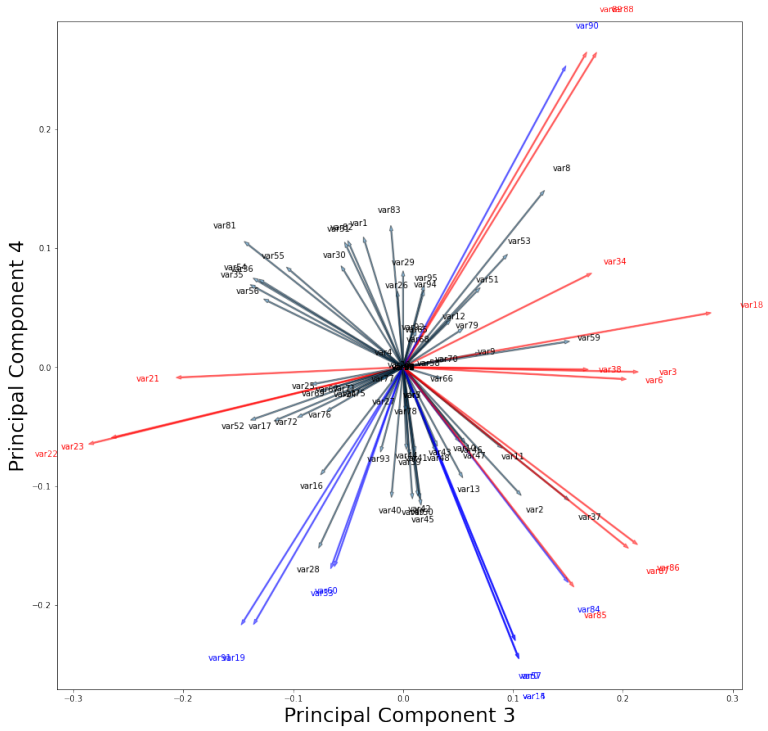


(b) PC3 vs PC4.

Figura 4.34: Loading Plots molino SAG 16.

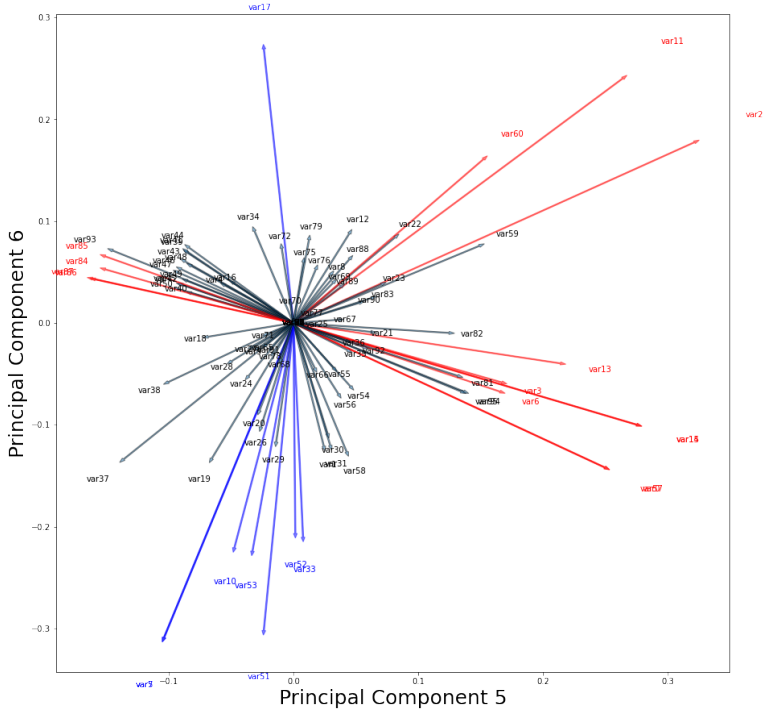


(a) PC1 vs PC2.



(b) PC3 vs PC4.

Figura 4.35: Loading Plots molino SAG 17.



(a) PC5 vs PC6.

Figura 4.36: Loading Plots molino SAG 17.

Luego, es necesario ordenar la información de las variables que afectan a las componentes del análisis, en conjunto con el indicador de peso explicado en la sección anterior. Para esto, se crea un archivo Excel con ambos indicadores organizados. En la Figura 4.37, se observa un extracto de la planilla Excel creada. Esto ayudará a crear un criterio de selección de variables.

Variable	Descripción	Indicador	PC1	PC2	PC3	PC4	PC5	PC6
variable 3 - MA2:3JI155	POTENCIA MOL. SAG 16	0,096573813						
variable 42 - MA2:3SI154	VELOC MOLINO SAG16	0,095756601						
variable 26 - MA2:3DT113	Temp Agua Enfriamiento Motor Entrada 1 SAG 16	0,095661804						
variable 27 - MA2:3DT115	Temp Agua Enfriamiento Motor Entrada 2 SAG 16	0,095578767						
variable 2 - MA2:3KIC118	Porcentaje de solidos Molino SAG-16	0,09397962						
variable 0 - MA2:3WIC110	FLUJO MINERAL A SAG16	0,089297075						
variable 73 - tolva_a2_rajo	Origen mineral Rajo	0,087957271						
variable 6 - MA2:3Y1PISX08	PRESION ACEITE DESCANSO ALIMENTACION (SALA DE ...	0,086681139						
variable 47 - MA2:3WI116	FLUJO MINERAL Pebbles Generados SAG 16.	0,08652472						
variable 14 - MA2:3Y1PISX18	PRESION ACEITE DESCANSO ALIMENTACION LADO NORT...	0,086126899						
variable 13 - MA2:3Y1PISX16	PRESION ACEITE DESCANSO ALIMENTACION LADO NORT...	0,085593133						
variable 5 - MA2:3Y1PISX07	PRESION ACEITE DESCANSO ALIMENTACION (SALA DE ...	0,085313626						
variable 12 - MA2:3Y1PISX14	PRESION ACEITE DESCANSO ALIMENTACION LADO NORT...	0,085015163						
variable 25 - MA2:3TIC165_Notify	TEMP ACEITE DESCAN SAG16	0,084989731						

Figura 4.37: Extracto planilla Excel con indicadores.

La planilla contiene el número, tag y descripción de cada variable en el análisis. También

contiene explícitamente el indicador de peso. Además, tiene una columna por cada componente principal utilizada en los Loading Plot. Estas columnas están coloreadas dependiendo de si la variables influye en la componente. De esta forma, y considerando la información que aporta cada uno de los indicadores, se establace un criterio de selección. Tal criterio se reduce a elegir hasta la última variable coloreada de las planillas excel creadas. Las variables seleccionadas tienen los indicadores de peso más altos y, la mayoría afectan a las primeras componentes principales.

De esta manera, la cantidad de variables a considerar para el análisis UMAP, se reduce a 62 para el molino SAG16 y a 74 para el SAG17. Analizando y comparando estas variables una a una, se obtiene que 61 de ellas son compartidas por ambos molinos. Por lo tanto, existe una variable para el activo SAG16 que solo afecta a este molino y, 13 variables que solo afectan al activo SAG17. Cabe destacar que, de las 13 variables, existen dos propias del molino SAG17 que no tienen equivalencia en el SAG16. De esto se deduce que los molinos si pueden usar la misma metodología y por lo tanto, pueden ser comparables hasta ahora con la metodología propuesta.

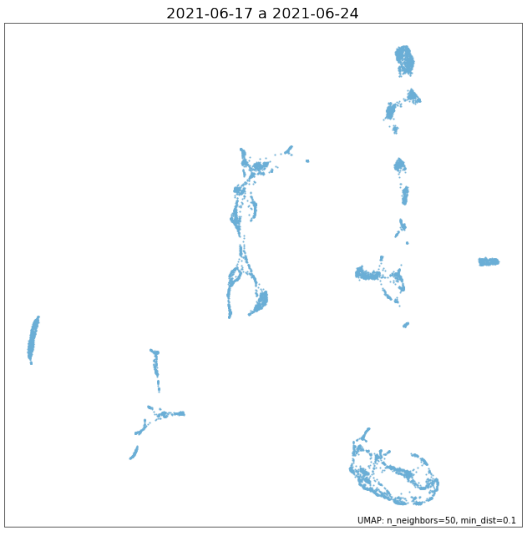
4.2. ANÁLISIS UMAP

En esta sección, se utiliza el algoritmo UMAP para caracterizar modos de operación de los molinos y poder explicar parte de la fenomenología del proceso, utilizando las variables escogidas por el criterio creado luego del análisis PCA.

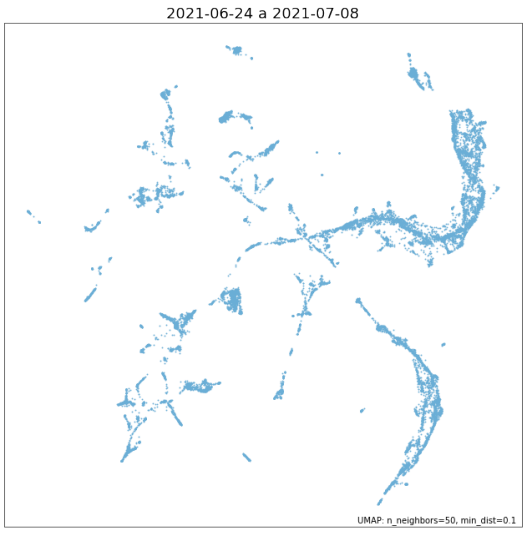
Primero, se debe definir una ventana de tiempo adecuada para entrenar UMAP. Idealmente, se debe usar la totalidad de los datos, ya que al contemplar la mayor cantidad de datos, se podrá caracterizar los modos de operación del molino de manera más representativa. Sin embargo, esto podría ser difícil de implementar, debido a la gran carga en la CPU y memoria RAM del equipo a utilizar. La ventana de tiempo para entrenamiento tampoco puede ser muy pequeña, pues se necesita información suficiente para caracterizar el comportamiento del molino lo mejor posible. Además, se busca una ventana de

tiempo que proyecte grupos de puntos lo mejor definidos posible.

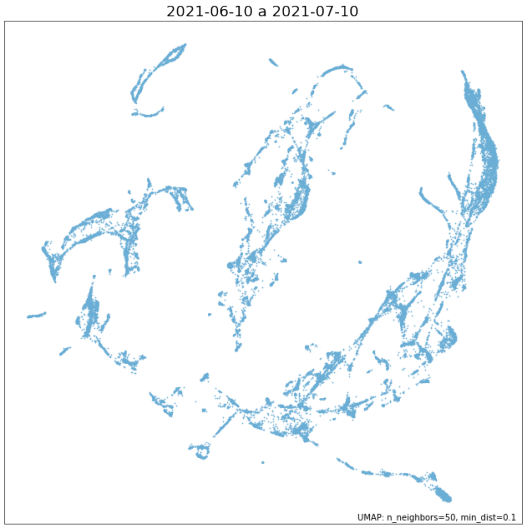
Al igual que en [2], se propone considerar ventanas de tiempo semanales, bisemanales, mensuales y bimensuales. Se muestran ejemplos de este análisis en la Figura 4.38. Los mapas semanales, como el de la Figura 4.38a contienen muy pocos datos. Con una ventana mensual como la de la figura 4.38c, se tienen una gran cantidad de datos que dificultan la formación de grupos. Para el caso de las ventanas bimensuales como las de la Figura 4.38d, se proyectan tantos datos que es difícil apreciar la nube de puntos creada. Por estas razones (y al igual que en [2]), se decide trabajar con ventanas de tiempo bisemanales. Resultan ser ventanas con grupos definidos y con una cantidad de datos aceptable. Por lo tanto, se consideran suficientemente representativas para el análisis.



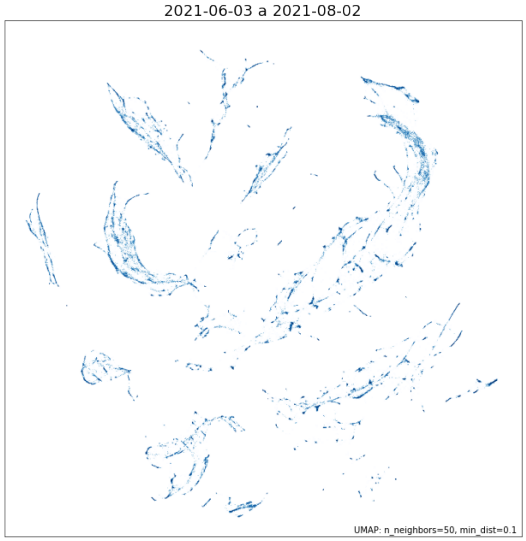
(a) Semanal.



(b) Bisemanal.



(c) Mensual.

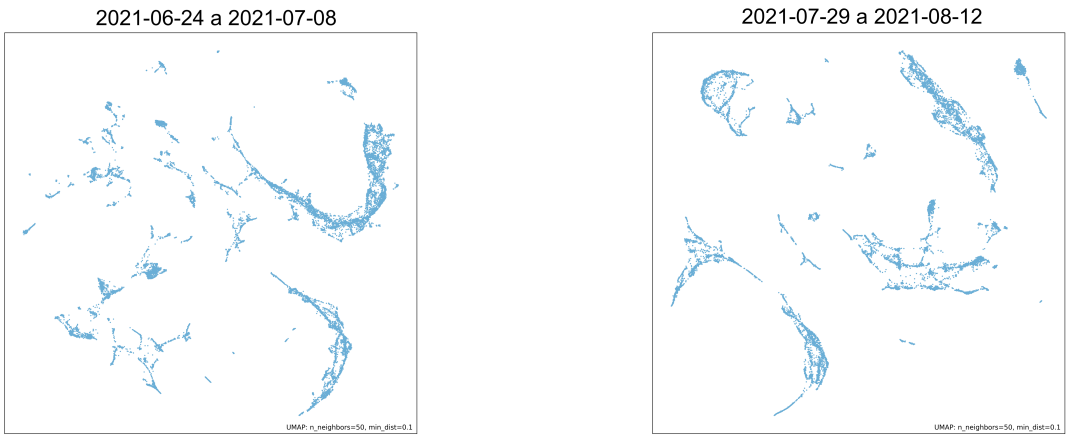


(d) Bimensual.

Figura 4.38: Ejemplo ventanas de tiempo.

Se busca a través de toda la data disponible, alguna nube de puntos con ventana de tiempo bisemanal, que cumpla con tener sus datos agrupados. Se encuentran los mapas de la Figura 4.39. Estos mapas serán usados en UMAP para el entrenamiento del Manifold. Esto permite, de cierta manera, forzar la forma de las nubes de puntos al hacer la proyección en otros datos. Por esta razón, la ventana escogida para crear el Manifold, debe tener las características requeridas comentadas anteriormente: suficiente

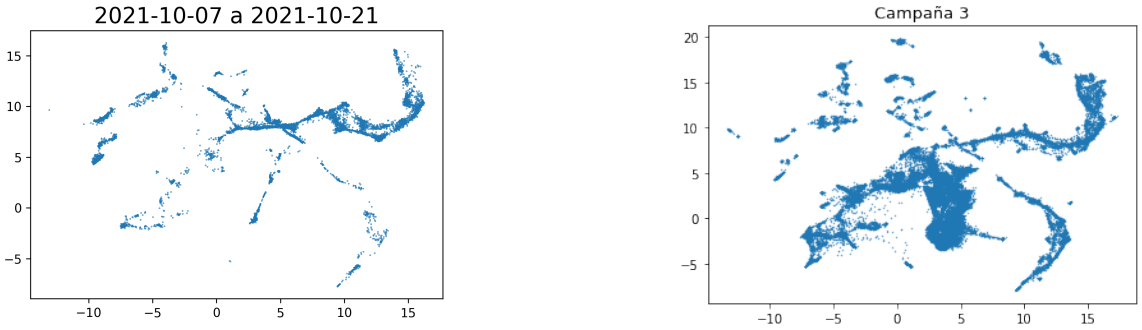
diversidad de puntos y grupos definidos.



(a) Molino SAG16. (b) Molinos SAG17.

Figura 4.39: Espacios bidimensional de entrenamiento.

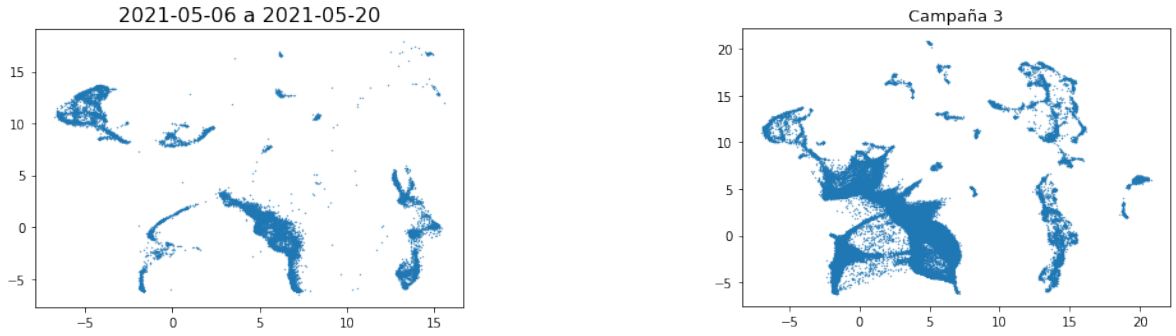
Luego, se proyectan mapas con otros periodos de tiempo utilizando el manifold creado, obteniéndose nuevos mapas. Para la selección de esta nueva nube de puntos, también es importante que se formen conjuntos de puntos lo mejor definidos posible, pues se busca que los clusters creados, caracterizan los modos de operación de los molinos. Se consideran para el análisis las gráficas de las Figuras 4.40 y 4.41.



(a) (b)

Figura 4.40: Ventanas de tiempo utilizando el Manifold de entrenamiento en molino SAG16.

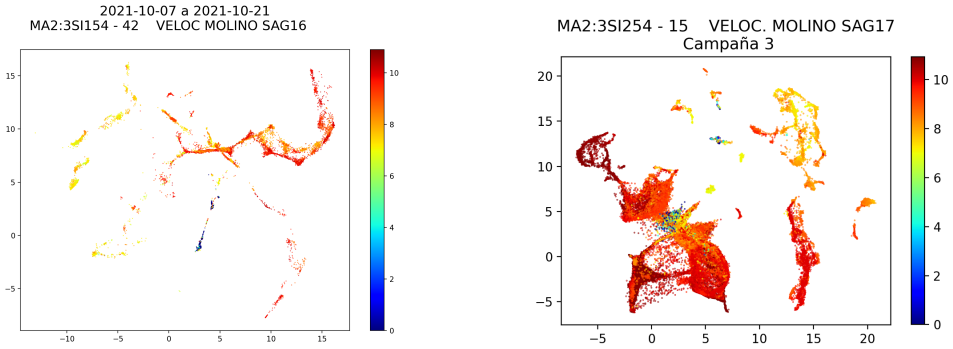
Posteriormente, se realiza un análisis con las proyecciones UMAP escogidas de las Fi-



(a) (b)

Figura 4.41: Ventanas de tiempo utilizando el Manifold de entrenamiento en molino SAG17.

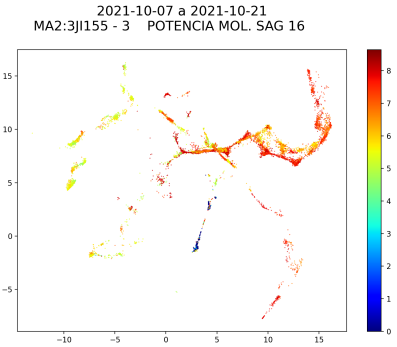
guras 4.40 y 4.41. Los mapas se colorean en función de algunas variables que permitan entender mejor su relación con la operación del molino. Se crean mapas coloreados en función de las variables que indican mayor variación. Esto permite conocer patrones y obtener conclusiones acerca de fenómenos producidos en el proceso minero. Cabe destacar que el comportamiento de algunas variables se aprecia mejor en periodos más extensos de tiempo. Por lo que el análisis también se hace en algunas nubes que representan campañas completas de los molinos SAG.



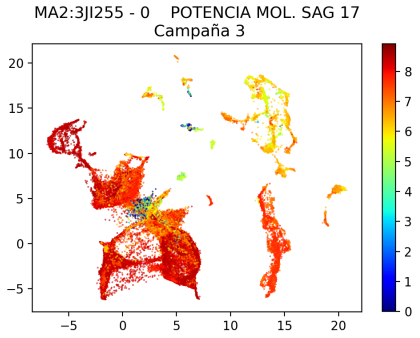
(a) Ventana bisemanal SAG16. (b) Ventana por campaña SAG17.

Figura 4.42: Proyecciones UMAP coloreadas en función de la velocidad.

En las imágenes de las Figuras 4.42 y 4.43, se observa que se tiene un comportamiento similar para velocidad y potencia. Las dos variables aumentan y disminuyen sus valores en los mismos instantes de tiempo. Esto tiene sentido, pues mientras a más potencia esté

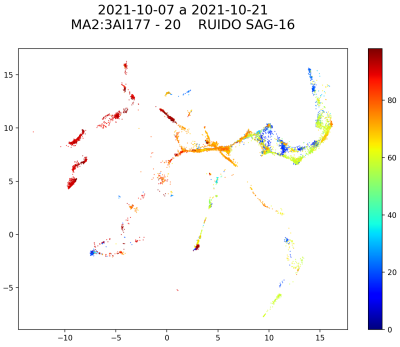


(a) Ventana bisemanal SAG16.

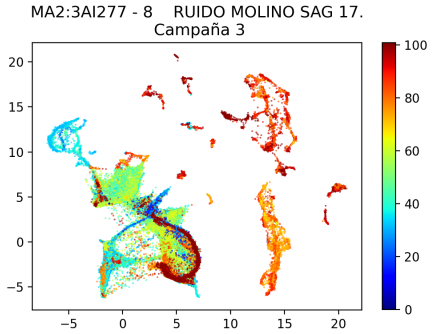


(b) Ventana por campaña SAG17.

Figura 4.43: Proyecciones UMAP coloreadas en función de la potencia.



(a) Ventana bisemanal SAG16.

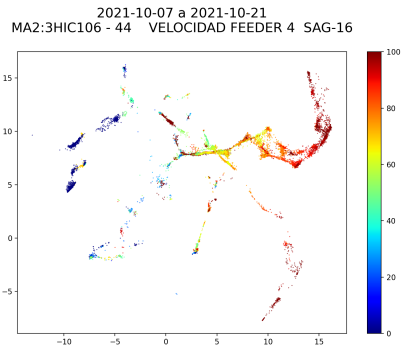


(b) Ventana por campaña SAG17.

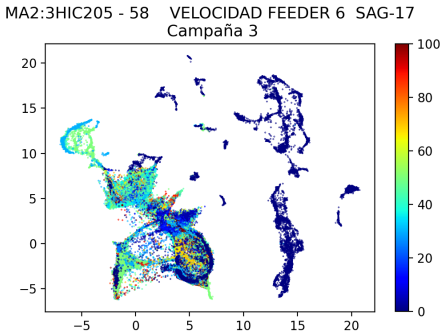
Figura 4.44: Proyecciones UMAP coloreadas en función del ruido.

funcionando el molino, este se moverá más rápido, y viceversa. Luego, al comparar los gráficos de velocidad y potencia con los de la Figura 4.44, se aprecia que en los instantes de tiempo en que el ruido es bajo la velocidad y la potencia son altas, y viceversa. Este fenoméno se explicaría debido a que el molino funciona a una mayor potencia y por lo tanto velocidad, cuando necesita moler mineral grueso. El trabajar con mineral grueso implica rocas grandes, pero menor cantidad de ellas para completar el peso para llenar el molino. Por ende, en este caso, los molinos estarían produciendo menor cantidad de ruido que si se muele mineral fino y pequeño donde hay más cantidad de rocas.

La velocidad del alimentador de la Figura 4.45, está aumentando a medida que también aumenta la potencia y la velocidad, eso tiene sentido pues esta variable tiene relación



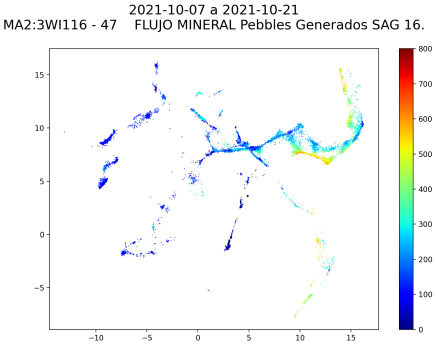
(a) Ventana bisemanal SAG16.



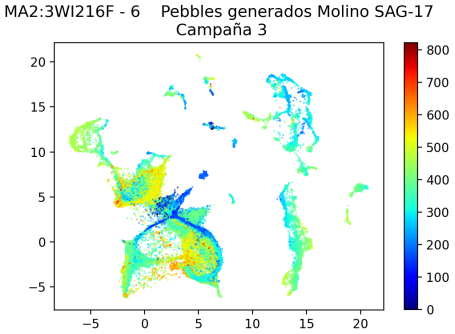
(b) Ventana por campaña SAG17.

Figura 4.45: Proyecciones UMAP coloreadas en función de la velocidad de un alimentador.

con la cantidad de mineral que se ingresa al molino.



(a) Ventana bisemanal SAG16.

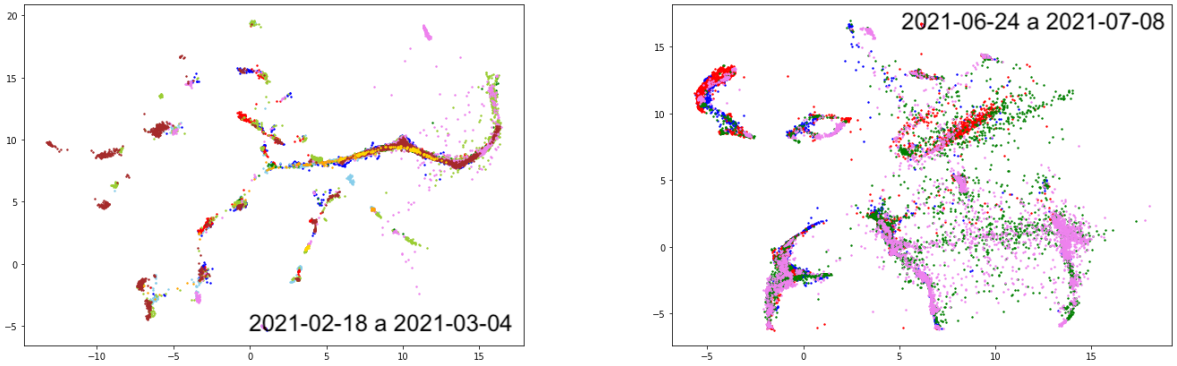


(b) Ventana por campaña SAG16.

Figura 4.46: Proyecciones UMAP coloreadas en función de los peebles generados.

En la Figura 4.46, se observa una nueva relación: Las zonas del mapa donde hay mayor cantidad de pebbles generados, coincide nuevamente con las zonas de alta alimentación, velocidad y potencia. En general, no es útil tener gran cantidad de pebbles, pues hay que decidir que hacer con ellos. Pueden ir a chancado o ser recirculados, por lo que el ideal sería solo obtener pulpa del proceso de molienda. Por lo tanto, las zonas donde hay mayor cantidad de pebbles podrían ser condiciones no deseadas para el proceso. Además, en estos instantes la potencia del molino es muy alta, por lo que podría significar un aumento en la agresividad del molino al añadir más mineral en los momentos donde hay muchos pebbles dentro del molino para por ejemplo, cumplir alguna meta.

Adicionalmente, se realiza un análisis con respecto a los operadores y su influencia en los molinos. Los resultados gráficos de este análisis se observan en la Figura 4.47. Se sabe que cada punto en el mapa se refiere a un instante de tiempo, y también que en cada instante hay un operador manipulando el molino. El análisis se enfoca en observar si la manipulación de los operadores llevan al proceso a moverse de cierta manera. Si bien, para la generación del mapa se utilizan los datos escogidos luego del análisis PCA. Para colorear el gráfico, se utiliza la data anterior al descarte de variables. La figura muestra los instantes de manipulación de los molinos de cada operador. Y, debido a que PCA deja fuera a la mayoría de variables con información de operadores, resulta necesario ocupar la información descartada.



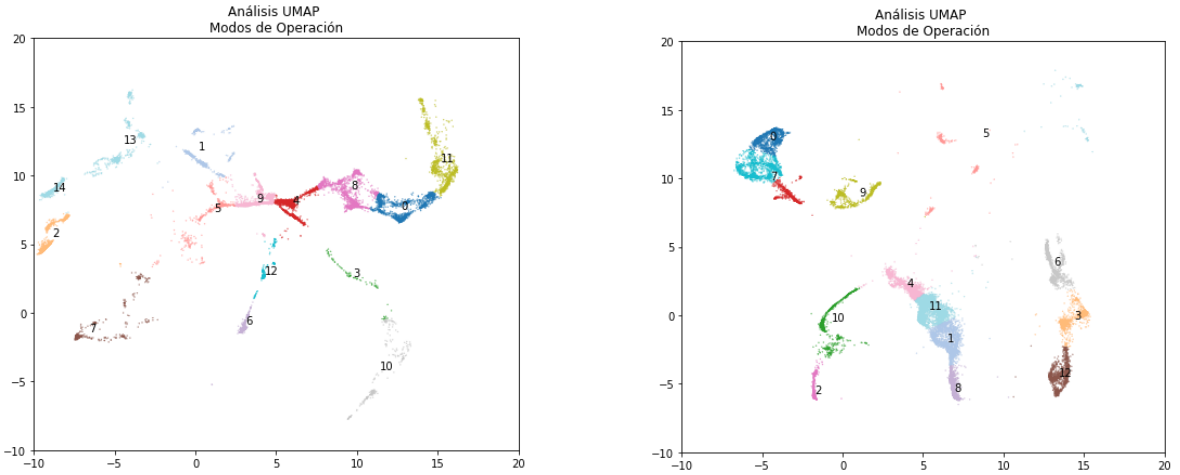
(a) Molino SAG16. (b) Molino SAG17.

Figura 4.47: Proyecciones UMAP coloreadas en función de los distintos operadores.

Cabe destacar, que cada color señala a un operador diferente. Entonces, según la Figura 4.47a, los operadores están en todas partes del mapa. Por lo tanto, no se observa una diferencia sustancial en el funcionamiento de los molinos en base a los operadores. En general, las variables de operadores no afectan de manera significativa al proceso según los análisis realizados hasta ahora. Esto podría estar señalando, que en realidad los operadores no influyen tanto en el proceso. Al menos con los análisis que se han hecho hasta el momento. Sin embargo, en la Figura 4.47b se observa que el operador indicado por el color rosado, pareciera estar afectando de mayor manera en algunos clusters. Esto podría indicar que en particular el operador rosado si afecta en algunos modos de operación del molino o, que los momentos en los que opera, el molino por ciertos

motivos podría estar actuando de manera diferente.

Finalmente, se utilizan los mapas UMAP de las Figuras 4.40a y 4.41a para definir clusters usando k-means. Además de que estos gráficos reflejan adecuadamente los modos de operación del molino, también se prefieren pues en los mapas coloreados son los que más varianza tienen. A diferencia de [2] donde se encontraron 14 clusters, se observa en la figura 4.48 que para el molino SAG16 se consideran 15 modos de operación, mientras que para el SAG17 13 modos de operación. La cantidad de clusters escogidos y coloreados, depende de la cantidad natural de grupos creados por el algoritmo UMAP, y se seleccionan por tener en el centro una cantidad considerable de puntos agrupados. Esto permite obtener conclusiones para un eventual análisis. Se dejan propuestos los modos de operación agrupados por el algoritmo para seguir el proceso de creación de un índice de degradación en posteriores investigaciones.



(a) Molino SAG16. (b) Molino SAG17.

Figura 4.48: Asignación de clusters, categorizados por k-means.

CAPÍTULO 5

CONCLUSIONES

5.1. APORTES A LA METODOLOGÍA PREVIA

Uno de los objetivos del presente trabajo de título, es la revisión de la metodología propuesta en [2]. Durante este proceso, se encontraron situaciones y decisiones que pudieron cambiar o mejorarse.

Dentro de lo posible, se añadió información adicional a la ya existente. Especificamente, 20 variables que señalan información acerca de los operadores de los molinos SAG y 17 variables que indican información sobre el mineral del proceso. El añadir tal información, aporta en el sentido de la fenomenología considerada para los análisis, puesto que la información incorporada en el desgaste de los molinos según conversaciones con CODELCO y el Departamento de Ingeniería Metalúrgica y de Materiales de la UTSFM.

Además, se realizó una correción en el filtro utilizado para detectar cuando el molino no opera. Se aplicó un filtro previo al analisis PCA, que permitió descartar los instantes de tiempo en que los molinos se encuentren detenidos. El haber añadido este filtro al análisis, permitió utilizar la información solo en los momentos en los cuales el revestimiento de los molinos se esté degradando, haciendo el análisis más preciso y enfocado en la operación de los molinos.

Los análisis propuestos en esta memoria, se basan en el algoritmo PCA. Los Scree Plot, Heatmap, el indicador de peso creado y los Loading Plot permitieron obtener una gran cantidad de resultados útiles para seguir avanzando en la investigación. El estudio profundo de PCA complementa la metodología previa y, podría servir de base para una futura investigación o desarrollo de un modelo predictivo.

5.2. LIMITACIONES

Si bien hubo considerables hallazgos en el presente trabajo, también hubieron complicaciones que limitaron algunos avances. El estudiar y comprender las áreas de minería y data science significó una gran inversión de tiempo y dificultad en la interpretación de los análisis creados. Luego de algunos meses, se comienzan a realizar todos los procedimientos que se requerían para obtener resultados. Debido a esto, no hubo suficiente tiempo para lograr desarrollar un índice de degradación.

Otras limitaciones ocurren respecto a los sensores. Estos presentan algunos desfases con otros debido a sus naturalezas. Además, en la data entregada por CODELCO, muchos sensores no tenían información en varios instantes de tiempo. Estos datos incompletos fueron sustituidos por ceros, lo que podría alterar el análisis con el molino detenido. Esto podría hacer que la metodología no fuera tan precisa, ya que se podría estar considerando el molino detenido cuando en verdad no lo estaría.

Otra situación que ocasionó algunas dificultades en las conclusiones de los análisis, fueron las estimaciones en los datos utilizados. En particular los datos empleados sobre origen y propiedades del mineral que llega a los molinos no provienen de sensores, si no que de modelos desarrollados por la empresa. Sin embargo, se demuestra a lo largo del proyecto, que independientemente de esto, se pueden obtener resultados útiles.

5.3. RESULTADOS OBTENIDOS

En la revisión de la metodología previa, se hizo un análisis de las señales en el tiempo, lo que explicó la separación de señales en el heatmap en [2]. Esto ratifica la importancia de estudiar las variables en el tiempo, para conocer su naturaleza y comportamiento. Por lo tanto, si se tiene la opción de decidir la ventana de tiempo de la data disponible para trabajar, es importante considerar que las variables tengan cierta varibilidad. En este mismo sentido, se cree que el criterio de selección utilizando solo el análisis de

varianza explicada no es suficiente para tomar una decisión respecto a la discriminación de variables.

Es asi como, de los análisis con el indicador de peso y los loading plots, se tiene un descarte de variables significativo. Existen 61 variables comunes entre los dos activos (SAG 16 y 17), que se dejan fuera del análisis. Esto demuestra que la metodología puede aplicaarse a ambos molinos entregando un análisis comparable. Esto podría servir como una base común para desarrollar un modelo predictivo, y con ello emplear estos análisis no solo en un tipo de molino, si no que para molinos de caracteristicas similares, aún cuando tengan un uso diverso.

En el análisis de UMAP coloreado, se logra dar un sentido al comportamiento de las variables observadas en estas nubes de puntos. El análisis permite establecer hipótesis sobre la fenomenología del proceso, visibilizando los instantes de tiempo en los que aportan las variables. Además, al hacer este análisis con los operadores, se deduce que no tienen gran relevancia en el proceso.

5.4. TRABAJO A FUTURO

A lo largo del trabajo, se obtuvieron resultados con muchas posibles decisiones que tomar. A pesar del exhaustivo y largo trabajo con PCA, podrían haberse tomado algunas decisiones diferentes como por ejemplo, el criterio de descarte de variables en la planilla Excel 4.37. El establecer criterios diferentes podría encaminar el análisis desde otro punto de vista y obtener nuevos resultados.

PCA es un algoritmo que considera la variabilidad y correlación de los datos, no explica causalidad. Debido a esto, se cree muy importante tener una buena primera elección de variables que expliquen de la mejor manera posible la fenomenología de los molinos pues el algoritmo se basará en la variabilidad para escoger variables y no en la influencia de estas sobre el molino. Por otra parte, PCA podría no ser la mejor herramienta para

discriminar las variables de los molinos SAG. Se recomienda estar abiertos a utilizar nuevos algoritmos como T-SNE (T-distributed Stochastic Neighbor Embedding) o, LDA (Latent Dirichlet Allocation) que pudieran obtener diferentes resultados en caso de realizar nuevos análisis, junto con poder corroborar lo obtenido en esta trabajo.

Además, en la medida de lo posible, se recomienda emplear sets de datos con la mayor ventana de tiempo posible, ya que utilizar datos en periodos de tiempo extensos, creará análisis más representativos. Se recomienda que la data contega por lo menos tres campañas completas del activo. Esto podría ayudar a comprender de mejor manera el comportamiento de las variables durante una campaña del molino.

REFERENCIAS BIBLIOGRÁFICAS

[1] CODELCO, "Minería en constante innovación," 2021. [Online]. Available: https://www.codelcoeduca.cl/codelcoeduca/site/edic/base/port/innovacion.html

[2] E. R. Grendi Molina, *Desarrollo de un modelo de degradación de un molino SAG para mantenimiento predictivo.* Universidad Técnica Federico Santa María, 2021.

[3] CODELCO, "Kairos mining s.a." 2011. [Online]. Available: https://www.codelco.com/kairos-mining-s-a/prontus_codelco/2011-03-03/111405.html

[4] J. A. Rodrigo, "Análisis de componentes principales," 2017. [Online]. Available: https://www.cienciadedatos.net/documentos/35_principal_component_analysis

[5] "How UMAP works." [Online]. Available: https://umap-learn.readthedocs.io/en/latest/how_umap_works.htmls

[6] Scikit-Learn, "Clustering k-means." [Online]. Available: https://scikit-learn.org/stable/modules/clustering.html

[7] Wikipedia, "Ciencia de datos," 2022. [Online]. Available: https://es.wikipedia.org/wiki/Ciencia_de_datos

[8] J. F. V. Rueda, "Aprendizaje supervisado y no supervisado," 2020. [Online]. Available: https://healthdataminer.com/data-mining/aprendizaje-supervisado-y-no-supervisado/

[9] IArtificial, "Análisis descriptivo, predictivo y prescriptivo de datos," 2020. [Online]. Available: https://www.iartificial.net/analisis-predictivo-y-prescriptivo-con-machine-learning/#Analisis_Prescriptivo

[10] P. U. C. de Chile, "¿Por qué estudiar estadística?" 2022. [Online]. Available: http://datascience.uc.cl/porque-estudiar-estadistica

[11] Crehana, "Mejores librerías de python," 2021. [Online]. Available: https://www.crehana.com/cl/blog/desarrollo-web/librerias-python/

[12] J. Arroyo-Hernández, "Métodos de reducción de dimensionalidad: Análisis comparativo de los métodos APC, ACPP y ACPK," *Uniciencia*, vol. 30, 2016.

[13] E. Gil, "Reducción de dimensionalidad de un dataset." [Online]. Available: https://profesordata.com/2020/09/01/reduccion-de-la-dimensionalidad-analisis-de-componentes-principales-pca/

[14] E. B. Turing, "Cómo leer biplots PCA y scree plots," 2018. [Online]. Available: https://bioturing.medium.com/how-to-read-pca-biplots-and-scree-plots-186246aae063

[15] "Interpret all statistics and graphs for principal components analysis." [Online]. Available: https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/all-statistics-and-graphs/

[16] "Principal component analysis in R," 2017. [Online]. Available: http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/118-principal-component-analysis-in-r-prcomp-vs-princomp/

[17] GraphPad, "Graphs for principal component analysis," 2017. [Online]. Available: https://www.graphpad.com/guides/prism/latest/statistics/stat_pca_graphs_tab.htm

[18] L. Ngo, "How to read PCA biplots and scree plots," 2018. [Online]. Available: https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/

[19] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[20] F. S. Caparrini, "Algoritmos de clustering," 2020. [Online]. Available: http://www.cs.us.es/~fsancho/?e=230