



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

**MODELOS PREDICTIVOS BASADOS EN TÉCNICAS DE APRENDIZAJE DE  
MÁQUINAS PARA EL ANÁLISIS DE PROCESOS DE DESGASTE EN  
NEUMÁTICOS DE CAMIONES MINEROS**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

RODRIGO IGNACIO ALBORNOZ CORNEJO

PROFESOR GUÍA:  
MARCOS ORCHARD CONCHA

MIEMBROS DE LA COMISIÓN:  
FRANCISCO RIVERA SERRANO  
LUIS MONTECINOS MOLINA

Este trabajo ha sido parcialmente financiado por:  
CONTAC INGENIEROS

SANTIAGO DE CHILE  
2022

RESUMEN DE LA MEMORIA PARA OPTAR  
AL TÍTULO DE: INGENIERO CIVIL ELÉCTRICO  
POR: RODRIGO IGNACIO ALBORNOZ CORNEJO  
FECHA: 2022  
PROF. GUÍA: MARCOS ORCHARD CONCHA

## **MODELOS PREDICTIVOS BASADOS EN TÉCNICAS DE APRENDIZAJE DE MÁQUINAS PARA EL ANÁLISIS DE PROCESOS DE DESGASTE EN NEUMÁTICOS DE CAMIONES MINEROS**

Los neumáticos de camiones de extracción (CAEX) son propensos a fallar inesperadamente, inhabilitando por completo el camión afectado. La infraestructura actual de la minera en estudio establece realizar inspecciones preventivas a intervalos no especificados de todos los neumáticos de la flota, las que implican altos costos en horas hombre. Optimizaciones respecto de cantidad de inspecciones y decisiones tomadas en cuanto a neumáticos de CAEX se hacen necesarias de implementar. Además, la disponibilidad de extensos sets de datos de inspecciones y variables operativas de neumáticos de la minera, vuelve atractiva la propuesta de buscar soluciones que aprovechen la vasta base de conocimiento ya compilada. Se propone el desarrollo de un sistema de monitoreo, donde el objetivo inicial y que valida la solución, es obtener un modelo capaz de predecir el desgaste de caucho ocurrido en los hombros del neumático desde la última inspección realizada. Mediante técnicas de aprendizaje de máquinas, en particular árboles de decisión, se crean tres módulos de dos etapas que conforman el modelo computacional. Se obtiene un modelo predictivo capaz de estimar el desgaste experimentado por los neumáticos en estudio, a partir de los datos de operación comprendidos en el intervalo última inspección-actualidad y registros históricos, con errores menores a 5 [mm] en 60 mil kilómetros recorridos y un MSE general de 1.44. Se logra entonces desarrollar un modelo predictivo que recoge la fenomenología subyacente de la minera en cuestión y permite monitorear el desgaste ocurrido en neumáticos de CAEX, lo que a su vez dispone una base para futuros estudios de optimización del uso de neumáticos en operaciones mineras.

*A mi hermosa familia,  
y a todos quienes me formaron.*

***Gracias***

# Agradecimientos

Especial agradecimiento a CONTAC Ingenieros por permitirme trabajar en este proyecto. Ha sido una experiencia enriquecedora y me ha permitido aplicar *machine learning* en un contexto real y desafiante. Este trabajo no habría sido posible sin su guía experta y disposición. Agradecimientos igualmente a Marcos Orchard, por ser un gran mentor durante el desarrollo de este trabajo de título, ofreciendo siempre excelentes consejos y claras nociones técnicas. Las decisiones tomadas durante este trabajo fueron siempre originadas en las largas conversaciones con ellos, por lo que nada de esto habría sido posible sin su ayuda.

Agradecimientos igualmente a la Universidad de Chile, que durante este periodo de pandemia ha tenido que adaptarse a las condiciones adversas y me ha permitido llevar a cabo mi trabajo de título sin mayores complicaciones.

Por último, agradezco a mi familia por darme el tiempo y espacio necesarios para llevar este proyecto a cabo, así como la necesidad de destacarme en mis actividades y funciones, para cumplir mis propias metas y anhelos.

# Tabla de Contenido

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>1</b>  |
| 1.1. Motivación y Antecedentes del Problema . . . . .   | 1         |
| 1.2. Alcance . . . . .  | 2         |
| 1.3. Formulación del Problema . . . . .   | 2         |
| 1.4. Objetivos . . . . .  | 3         |
| 1.5. Estructura del informe . . . . .   | 4         |
| <b>2. Marco Teórico</b>   | <b>5</b>  |
| 2.1. Conceptos . . . . .  | 5         |
| 2.2. Neumáticos de CAEX . . . . .   | 6         |
| 2.2.1. Características Fundamentales . . . . .  | 6         |
| 2.2.1.1. Materiales . . . . .   | 6         |
| 2.2.1.2. Partes . . . . .   | 6         |
| 2.2.1.3. Posición en el camión . . . . .  | 7         |
| 2.2.2. Adquisición de Datos . . . . .   | 7         |
| 2.2.2.1. Sensores de Presión y Temperatura . . . . .  | 8         |
| 2.2.3. Formato de la Data . . . . .   | 9         |
| 2.3. Algoritmos de Regresión . . . . .  | 11        |
| 2.3.1. Regresión Lineal . . . . .   | 11        |
| 2.3.1.1. Método <i>Least Squares</i> . . . . .  | 12        |
| 2.3.2. Árboles de Regresión . . . . .   | 13        |
| 2.3.2.1. Partición Recursiva . . . . .  | 13        |
| 2.3.2.2. Podado ( <i>Pruning</i> ) de Árboles de Regresión . . . . .                                      | 14        |
| 2.3.3. <i>Boosting</i> . . . . .  | 15        |
| 2.3.4. <i>Adaboost</i> . . . . .  | 16        |
| 2.3.5. <i>Gradient Boost</i> . . . . .  | 17        |
| 2.3.6. <i>Extreme Gradient Boost (XGBoost)</i> . . . . .  | 18        |
| 2.4. Análisis de modelos . . . . .  | 19        |
| 2.4.1. <i>Mean Squared Error</i> . . . . .  | 20        |
| 2.4.2. <i>GridSearchCV</i> . . . . .  | 20        |
| 2.4.3. <i>Feature Importances</i> . . . . .   | 21        |
| 2.4.4. <i>SHAP Values</i> . . . . .   | 21        |
| <b>3. Estado del Arte</b>   | <b>23</b> |
| 3.1. “ <i>Haul truck tyre dynamics due to tire condition</i> ” [21]. . . . .                              | 23        |
| 3.2. “ <i>Reducing GHG emissions through efficient tire consumption in open pit mines</i> ” [22]. . . . . | 24        |

|           |   |           |
|-----------|---|-----------|
| 3.3.      | <i>"Predicting the effect of operating parameters on the radial tire treadwear using response surface method" [23]. . . . .</i> | 25        |
| 3.4.      | Aumento de vida útil de los neumáticos para la flota de transporte de Minera Centinela [24]. . . . .                            | 26        |
| 3.5.      | Disminución de los costos por hora en neumáticos de camiones mineros [25] .   | 27        |
| <b>4.</b> | <b>Trabajo Realizado</b>  | <b>29</b> |
| 4.1.      | Pre-procesamiento de la Data . . . . .  | 29        |
| 4.2.      | División del <i>Dataset</i> . . . . .   | 30        |
| 4.3.      | Prueba de Concepto . . . . .  | 31        |
| 4.4.      | Incorporación de Fenomenología de la Mina . . . . .   | 33        |
| 4.5.      | Inclusión de Data de Presión y Temperatura . . . . .  | 35        |
| <b>5.</b> | <b>Resultados</b>   | <b>40</b> |
| <b>6.</b> | <b>Conclusiones</b>   | <b>44</b> |
|           | <b>Bibliografía</b>   | <b>45</b> |
|           | <b>Anexo A. Tasa de Desgaste en Base a Horas Entre Inspecciones</b>   | <b>47</b> |
| A.1.      | MSE . . . . .   | 47        |
| A.2.      | Seguimiento de Desgaste Real . . . . .  | 48        |
| A.3.      | Importancia de Variables . . . . .  | 52        |
| A.4.      | Conclusiones . . . . .  | 53        |

# Índice de Tablas

|      |   |    |
|------|---|----|
| 4.1. | Valor de MSE de modelos predictivos según posición en el camión y tipo de algoritmo usado. . . . .  | 32 |
| 4.2. | Valor de MSE de modelos predictivos según posición en el camión, al ser entrenados con las variables iniciales y al agregar data obtenida de presión y temperatura. . . . .                                   | 39 |
| 5.1. | Valor de MSE de modelos predictivos según posición en el camión y tipo de algoritmo usado, para la arquitectura de 2 etapas. . . . .  | 40 |
| 5.2. | Valor de MSE de todos los modelos predictivos entrenados, midiendo el error del seguimiento de desgaste real acumulado $y_i$ . . . . .  | 40 |
| A.1. | Valor de MSE de modelos predictivos según posición en el camión y tipo de algoritmo usado para la arquitectura de 2 etapas, usando la variable $\Delta Horas$ para el cálculo de la tasa de desgaste. . . . . | 48 |

# Índice de Ilustraciones

|      |   |    |
|------|---|----|
| 1.1. | Esquema ilustrativo de la formulación del problema. . . . .   | 3  |
| 2.1. | Posibles posiciones de un neumático de CAEX y sus descriptores [8]. . . . .   | 7  |
| 2.2. | Sensor MEMS® 4 [9]. . . . .   | 9  |
| 2.3. | Interpretación Geométrica de <i>least squares</i> [11]. . . . .   | 12 |
| 2.4. | Partición del espacio de predicción de un árbol de regresión [11]. . . . .  | 13 |
| 2.5. | Ejemplo de un árbol antes de podar y después de podar, entrenados en un <i>dataset</i> de rendimiento de golpeadores de <i>baseball</i> [12]. . . . .   | 15 |
| 2.6. | Ejemplo de visualización de resultados de la librería de <i>SHAP Values</i> de Python. Las variables aumentan en importancia mientras más arriba se encuentran en el gráfico, cada punto representa una observación de la variable y su color se relaciona con el valor de la misma. Mientras más a la izquierda se encuentran los puntos, más restan al resultado y mientras más a la derecha estén, más suman [20]. . . . . | 22 |
| 3.1. | Desplazamiento vertical de un modelo basado en la física de neumáticos de CAEX, al considerar un neumático en buen estado y otro con falla, en un camino pedregoso [21]. . . . .  | 24 |
| 3.2. | Progreso y disminución de residuos generados en una mina de Brasil, medido como WTR ( <i>Waste Tire Rubber</i> ), es decir, la masa de neumáticos descartados dividida por la producción de la mina [22]. . . . .   | 25 |
| 3.3. | Plataforma experimental fabricada para medir desgaste ocurrido en un neumático y relacionarlo con variables operativas como presión, distancia y velocidad [23]. . . . .  | 26 |
| 3.4. | Distribución correcta de la carga en un camión de extracción, del “ <i>Manual de concientización de neumáticos</i> ” de la tesis de grado de Juan Ignacio Morales Navarro [24]. . . . .   | 27 |
| 3.5. | Evolución de los costos por hora entre dos faenas mineras, usando el acercamiento de <i>frecuencia óptima</i> de recambio de neumáticos [25]. . . . .   | 28 |
| 4.1. | Medianas de tasas de desgaste mensuales obtenidas con $\Delta y_j / \Delta Odom$ según posición en el camión, agrupadas en delanteras, traseras externas y traseras internas. . . . .   | 31 |
| 4.2. | Acumulación de desgaste en mm y predicciones de algoritmos regresión lineal, <i>Adaboost</i> y <i>XGBoost</i> , para 3 neumáticos, en función de las horas que el neumático se mantuvo instalado en el camión. El área destacada representa el espacio de tiempo que el neumático se encontró en posiciones delanteras. . . . .   | 32 |
| 4.3. | Histogramas de $\Delta y_j$ para la flota de neumáticos y predicciones de <i>Adaboost</i> para estos valores. . . . .   | 33 |
| 4.4. | Esquema del algoritmo computacional de 2 etapas generado. . . . .   | 34 |



|      |   |    |
|------|---|----|
| 4.5. | Gráfico de <i>feature importance</i> y <i>SHAP Values</i> para el algoritmo de 2 etapas generado, en posiciones delanteras, usando XGBoost para entrenar el modelo predictivo. . . . .  | 36 |
| 4.6. | Gráfico de <i>feature importances</i> y <i>SHAP Values</i> para el algoritmo de 2 etapas generado, usando <i>XGBoost</i> como modelo predictivo en posiciones delanteras, con nuevas variables de presión y temperatura. . . . .  | 38 |
| 5.1. | Curvas de desgaste predicho (usando solo tasa mensual y, por otra parte, agregando ajuste de XGBoost) y real, para los 9 neumáticos más problemáticos de la flota (mayor error de seguimiento de desgaste real). El área azul representa el periodo que el neumático estuvo en posiciones delanteras. . . . . | 42 |
| 5.2. | Curvas de desgaste predicho (usando solo tasa mensual y, por otra parte, agregando ajuste de XGBoost) y real, para 9 neumáticos escogidos al azar de la flota. El área azul representa el periodo que el neumático estuvo en posiciones delanteras. . . . .   | 43 |
| A.1. | Curvas de desgaste predicho (usando por un lado sólo tasa mensual y por otro agregando los ajustes de algoritmos predictivos) y real, para los 9 neumáticos más problemáticos de la flota. El área azul representa el periodo que el neumático estuvo en posiciones delanteras. . . . .                       | 49 |
| A.2. | Histograma de kilometraje alcanzable para un paso fijo de 250 horas entre mediciones, usando todas las tasas de desgaste disponibles en el <i>dataset</i> ( $250 \frac{\Delta Odom}{\Delta Horas}$ para todos los neumáticos). . . . .  | 50 |
| A.3. | Curvas de desgaste predicho (usando por un lado sólo tasa mensual y por otro agregando los ajustes de algoritmos predictivos) y real, para 9 neumáticos de la flota escogidos al azar. El área azul representa el periodo que el neumático estuvo en posiciones delanteras. . . . .                           | 51 |
| A.4. | <i>SHAP Values</i> para el experimento de tasa de desgaste en base a horas en neumáticos traseros externos. . . . .   | 52 |
| A.5. | <i>Feature importances</i> para el experimento de tasa de desgaste en base a horas en neumáticos delanteros. . . . .  | 53 |

# Capítulo 1

## Introducción

### 1.1. Motivación y Antecedentes del Problema

Los camiones de extracción (CAEX) cumplen un rol crucial en la actividad minera, por lo que resulta imprescindible asegurar altos estándares de disponibilidad y continuidad operacional en dichos activos. En este contexto, el problema de monitoreo de la condición de neumáticos es particularmente relevante, por cuanto dichos elementos presentan condiciones de falla que pueden ser difíciles de pronosticar, además de representar un porcentaje significativo del costo operacional de CAEX, superando los 5 mil dólares por unidad. En 2015, la empresa Otraco implementó medidas para prevenir fallas inesperadas de neumáticos de CAEX en la mina Radomiro Tomic, ahorrando 4.9 millones de dólares gracias a la mejora de rendimiento obtenida [1]. Las inspecciones regulares de neumáticos de CAEX incluyen la medición de milimetraje de hombros, así como la inspección visual de las distintas partes del activo [2]. Este es un proceso laborioso y que requiere de mantener un conjunto de los camiones detenidos, representando altos costos en las mineras que usan este recurso con regularidad.

El mercado de neumáticos de CAEX está dominado por empresas como Bridgestone y Michelin, quienes recomiendan que estos activos pasen  $\frac{1}{3}$  de su vida en posiciones delanteras y el resto en posiciones traseras [3], situación constatada en la mina estudiada. Esto, sin embargo, puede no ser óptimo y ofrecer una oportunidad de mejora según las condiciones individuales de uso de los activos. Las mineras ponen especial atención en las condiciones de temperatura de los neumáticos de CAEX, pero esta variable no es particularmente representativa de la tasa de desgaste del activo en su conjunto, ya que tiende a variar considerablemente de minuto a minuto y depende de múltiples factores [4].

Nuevos métodos para el monitoreo de desgaste de neumáticos han sido de interés para la industria los últimos años, viéndose iniciativas con múltiples propuestas. Recientemente, CODELCO publicó un desafío para la creación de un sistema de monitoreo de desgaste de neumáticos que no se base únicamente en datos de presión y temperatura y que, preferentemente, ocupe algoritmos de *machine learning* [5]. Conceptos tales como Internet de las Cosas ofrecen una oportunidad concreta para replantear la forma en la que se implementan esquemas de supervisión, diagnóstico y pronóstico de eventos catastróficos. Dadas estas circunstancias, CONTAC Ingenieros se encuentra desarrollando múltiples proyectos de minería relacionados con aprendizaje de máquinas y ciencia de datos. Como parte de estos esfuer-

zos, se ha planteado la idea de aprovechar los recursos disponibles de data de neumáticos de CAEX, para mejorar las operaciones del cliente en términos de eficiencia del desgaste de caucho.

El trabajo realizado muestra la aplicación de herramientas de *machine learning* en un contexto real, sirviendo como caso de estudio en la implementación de algoritmos inteligentes para la detección del estado de salud de activos (en este caso, desgaste de caucho en los hombros de neumáticos). Además, da un primer paso a la solución de gestión de neumáticos que será implementada en la minera en estudio y que, eventualmente, permitirá reducir costos en mantención y adquisición de neumáticos de CAEX.

## 1.2. Alcance

Una herramienta capaz de optimizar el desgaste de neumáticos requeriría de múltiples algoritmos computacionales, debido a la complejidad del problema mismo. En primer lugar, se requiere de un modelo predictivo capaz de detectar el desgaste de caucho ocurrido en el neumático a partir de sus variables de operación, lo que determina una primera capa de abstracción para el problema, además de ofrecer un sensor virtual que evitaría el constante inspeccionamiento de neumáticos. Luego, este modelo puede ser utilizado en modelos de pronóstico, simulando escenarios futuros a partir de probabilidades de uso del neumático y, así, optimizando el desgaste de la goma según las decisiones tomadas en la actualidad.

Este proyecto estará enfocado en la primera etapa, por cuanto no se considera como parte del mismo programar la optimización del uso de neumáticos en la minera, así como la generación de buenas prácticas y aumento de la vida útil si bien el código escrito será retomado luego para estos fines. El trabajo de título en cuestión considera el desarrollo de un modelo predictivo capaz de monitorear el desgaste de los neumáticos en el instante actual a partir de sus datos operacionales, quitando la necesidad de inspección constante de estos activos.

## 1.3. Formulación del Problema

En las tarjetas de vida de los neumáticos se lleva registro de la medición de milimetraje de sus hombros, lo que refiere a los bordes externos del neumático, que pierden material progresivamente con el uso. Estas mediciones son promediadas para cada neumático y el resultado es usado como indicador del estado de salud, siendo, efectivamente, la variable a predecir en este problema. En la figura 1.1 se observa un esquema ilustrativo del problema a resolver.

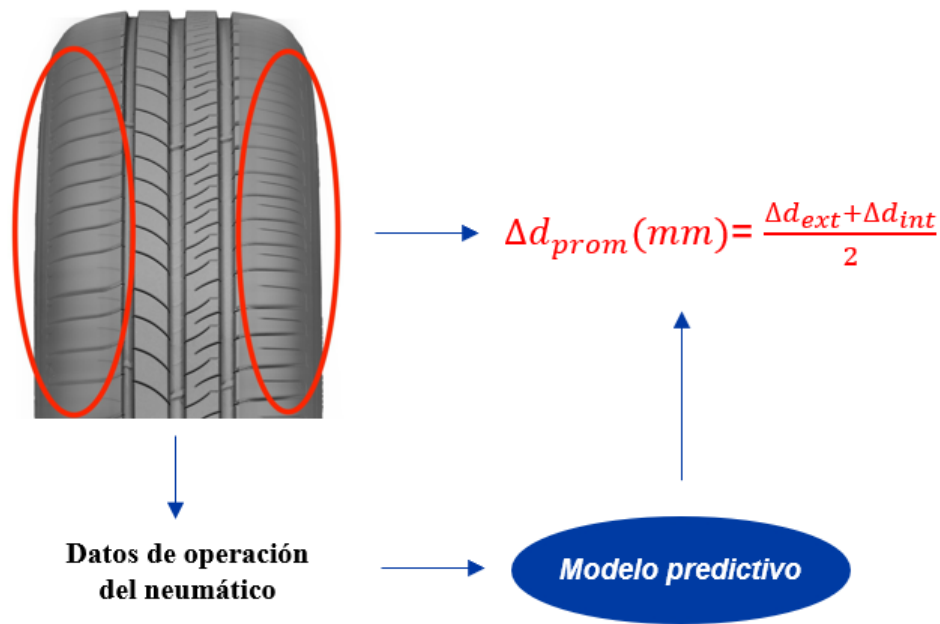


Figura 1.1: Esquema ilustrativo de la formulación del problema.

En este proyecto se buscan predecir las diferencias de milimetraje entre inspecciones. Esto permite usar el modelo en cualquier instante de operación del neumático, sin necesidad de informar al mismo del momento en el que se realiza la predicción (útil para los modelos de pronóstico). La variable a predecir, entonces, está representada por  $\Delta d_{prom}(mm)$ . Esto implica que los modelos predictivos a entrenar corresponden a algoritmos de regresión, donde la variable objetivo es continua.

## 1.4. Objetivos

El objetivo principal de este proyecto, es **desarrollar un modelo predictivo capaz de detectar el desgaste que ha sufrido un neumático de CAEX desde la última medición, a partir de variables operacionales (*Sensor virtual*)**. Parte del pre-procesamiento de la información fue ya ejecutado por el Laboratorio de Diagnóstico y Pronóstico de Fallas de la Universidad de Chile, quienes desarrollaron un odómetro a partir de datos de ubicación satelital de los camiones, además de cruzar los datos de tarjetas de vida con esta variable.

Los objetivos específicos del proyecto en cuestión, son:

- Procesar data disponible para ser usada en entrenamiento de algoritmos regresivos.
- Dividir el *dataset* según las tasas de desgaste subyacentes.
- Realizar una prueba de concepto con regresión lineal, Adaboost y XGBoost.

- Incorporar aspectos fenomenológicos de la mina en estudio en el entrenamiento de modelos.
- Procesar data de presión y temperatura obtenida con sensores de presión y temperatura, además de reconocer su importancia en el proceso estudiado.

Se usará *Mean Squared Error* (MSE), histogramas de error, comparación de gráficos de desgaste sobre kilometraje e histogramas de predicciones para determinar qué tan bueno es el desempeño de los modelos entrenados.

## 1.5. Estructura del informe

El informe a continuación se compone de los siguientes capítulos: En la siguiente sección se detalla el **marco teórico** en el cual se basa el trabajo, donde se definen dos partes: en primer lugar se detallan los conceptos y algoritmos usados a lo largo del proyecto y que componen la base de conocimiento necesario para entender lo tratado en este documento; y, en segundo lugar, se describe el estado del arte en el área de estudio, donde se analizan otras soluciones similares al problema desgaste de neumáticos de CAEX y que serán útiles para el desarrollo de este trabajo. La siguiente sección detalla la **metodología** utilizada para lograr efectivamente solucionar el problema de estudio, así como las diversas ideas que se pusieron a prueba a lo largo de este esfuerzo. Seguido de esta sección, se presenta el **trabajo realizado** junto a los resultados obtenidos, mostrando métricas de rendimiento de los algoritmos entrenados, con el objetivo de demostrar su eficacia. Por último, se muestran las **conclusiones** del trabajo realizado, haciendo una proyección a futuro de cómo será utilizado luego para mejorar el rendimiento de los neumáticos de CAEX en la minera objetivo.

# Capítulo 2

## Marco Teórico

En esta sección se detallarán las bases teóricas necesarias para entender el trabajo de título realizado, así como conceptos clave y el estado del arte. El marco teórico se dividirá en: neumáticos de CAEX, donde se especificará cómo se adquieren los datos usados en el entrenamiento de los modelos, así como características fundamentales de estos activos; algoritmos de regresión, donde se tratarán los métodos usados para entrenar modelos predictivos; y análisis de modelos, sección en la cual se detallarán las herramientas utilizadas para comprobar el rendimiento de los modelos predictivos entrenados.

### 2.1. Conceptos

- **Dataset:** Conjunto de datos relacionados entre sí, recopilados en una tabla.
- **Odometría:** Obtención de la distancia recorrida por un vehículo a partir de datos de movimiento. En este trabajo se le denomina odometría a la medida de kilometraje recorrido por el camión de extracción, estimada a partir de datos de posición satelital (GPS).
- **Aprendizaje Supervisado:** Uso de datos estocásticos para deducir una función. Su objetivo es crear una función capaz de predecir un valor a partir de una serie de datos de entrada, luego de haberle presentado ejemplos o datos de entrenamiento.
- **Overfitting:** Rendimiento insuficiente de un clasificador o regresor producto de una generalización sesgada de la muestra de entrenamiento. Cuando el modelo se ajusta a los casos particulares del conjunto de entrenamiento y no es capaz de reconocer nuevas muestras correctamente, puede estar considerando como válidos solo los datos idénticos a los del conjunto de entrenamiento.
- **Cross-validation:** Método de validación de modelos predictivos, donde se divide el *dataset* de entrenamiento en sub-conjuntos complementarios y se entrena al algoritmo de predicción en parte de ellos, para luego usar el sub-conjunto sobrante como conjunto de validación. Esto es realizado con múltiples particiones del espacio, probando el algoritmo entrenado en distintos sub-conjuntos cada vez.

## 2.2. Neumáticos de CAEX

En esta sección se presentarán las características fundamentales de los neumáticos de camiones de extracción, así como los sistemas de adquisición de datos y el formato de la data obtenida.

### 2.2.1. Características Fundamentales

A continuación se describen a grandes rasgos las características más importantes de los neumáticos de CAEX, para efectos de este trabajo.

#### 2.2.1.1. Materiales

Los neumáticos de camiones mineros se componen principalmente de goma y materiales de refuerzo. Estos se pueden dividir en [6]:

- Partículas de refuerzo: estos materiales son usados como refuerzo y relleno y, usualmente, se incluyen entre ellos silica, negro de carbón y resinas. Estos mejoran la estructura de la goma vulcanizada, aumentando su resistencia mecánica.
- Cordones de refuerzo: una parte de estos se componen de fibras poliméricas, usados para reforzar la carcasa del neumático y, entre ellos, se cuentan rayón, algodón, poliéster y nailon. Otra parte son usados en cinturones, entre los que se pueden encontrar filamentos de acero, fibra de vidrio y aramida.
- Goma: esta es formulada, específicamente, para su uso en minería y los camiones de extracción usan *goma natural* (NR de *natural rubber*), debido a su elasticidad y resistencia a cortes y a temperatura.

#### 2.2.1.2. Partes

A pesar de ser neumáticos especialmente fabricados para ser operados en minería, comparten características con sus pares de calle, en particular en sus partes fundamentales. Las partes de un neumático son [7]:

- Capa interior o aislamiento interno: capa de caucho que evita filtración del aire.
- Carcasa: aguanta la presión del aire y la tracción.
- Talón: asegura el neumático en la llanta, evitando que patine, además de evitar la filtración del aire.
- Flanco o Costado: protege la carcasa y se encuentra entre el hombro y la llanta.
- Hombro: sección ubicada entre la banda de rodadura y el flanco. Es la capa más gruesa y permite la evacuación de temperatura.

- Banda de rodadura: parte externa del neumático que mantiene contacto con el suelo. Se compone a su vez de:
  - Surcos: encargados de la evacuación de agua.
  - Costillas: secciones protuberantes situadas entre los surcos.
  - Tacos: permiten la tracción del vehículo y se encuentran en los cantos externos.
  - *Sipes*: cortes transversales que ayudan en la eliminación de agua.

### 2.2.1.3. Posición en el camión

Los neumáticos estudiados pueden ocupar una de seis posiciones disponibles en un camión de extracción: dos delanteras y cuatro traseras. La posición que ocupan afecta directamente su tasa de desgaste promedio y varía dependiendo del mes de estudio. Las recomendaciones de los fabricantes indican que estos activos deben pasar 1/3 de su vida útil en posiciones delanteras (de mayor tasa de desgaste) y el resto en posiciones traseras. En la figura 2.1 se presenta una ilustración de las posibles posiciones que puede tomar un neumático en CAEX y sus respectivos descriptores.

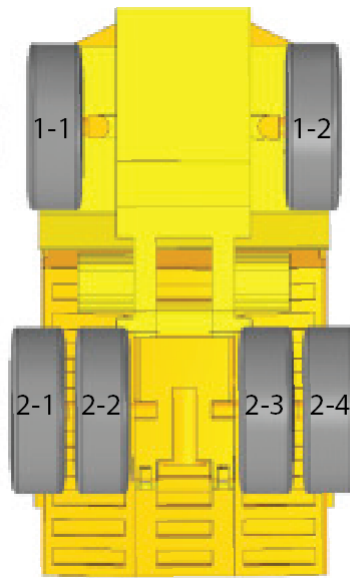


Figura 2.1: Posibles posiciones de un neumático de CAEX y sus descriptores [8].

### 2.2.2. Adquisición de Datos

Los camiones de extracción son máquinas complejas y presentan numerosos sistemas de seguimiento. Los neumáticos de estos activos, por su parte, comúnmente incluyen sensores de presión y temperatura, que envían datos en tiempo real a los sistemas de control de la minera. Para compensar la falta de información disponible en cuanto a neumáticos, estos son inspeccionados continuamente para controlar su estado de salud, donde el operador usa una



linterna para detectar posibles cortes en el caucho, los cuales, dependiendo del lugar, pueden ser indicio de una pronta falla del activo. Además de esto, se usa un medidor de profundidad de la banda de rodadura, para tomar registro del milimetraje de los hombros del neumático [2]. Esta variable puede ser característica del perfil de uso del neumático en el último tiempo y representativa de la vida útil restante.

Los datos operacionales de los neumáticos tienen su origen en 3 fuentes:

- **Tarjetas de vida:** acopio de información recopilada en inspecciones de neumáticos, donde se lleva registro según número de serie, fecha y hora, de las medidas de milimetraje de los hombros de los neumáticos, así como anotaciones basadas en observación superficial y la posición del camión en la que se encuentra el neumático.
- **GPS:** ubicación satelital del camión en tiempo real.
- **Sensores de presión y temperatura:** sensores que reportan variables de estado al operador, entre las que se encuentran presión y temperatura, con un intervalo aproximado de 1 minuto.

Las 3 fuentes pueden ser combinadas en un solo dataset para luego entrenar los modelos predictivos. En particular, se toma como base las tarjetas de vida y, sobre esto, se armará el *dataset* conjunto.

Las bases de datos de las que se dispone corresponden a todos los neumáticos de la flota de la minera en estudio que fueron operados entre los años 2018 y 2019.

#### 2.2.2.1. Sensores de Presión y Temperatura

En los neumáticos de CAEX actuales es posible incorporar la tecnología de sensores de presión y temperatura para el monitoreo del estado del producto, los cuales, en ciertos casos, incluyen tecnología GPS. Estos sensores tienen la capacidad de enviar alarmas de temperatura y presión para modificar la operación del camión en tiempo real. En la figura 2.2 se muestra una imagen de un MEMS 4 instalado en un neumático de CAEX, sensor que dispone de las capacidades previamente mencionadas.



Figura 2.2: Sensor MEMS® 4 [9].

Los sensores de presión y temperatura usados en minería son capaces de enviar data con el número serial específico del neumático, lo que es útil para incorporar los datos recopilados a las bases de datos de la mina. En el caso de los sensores MEMS 4, estos usan una aplicación web para conectarse con los sistemas de monitoreo [10].

### 2.2.3. Formato de la Data

Dado que parte de los datos se encuentran pre-procesados, ya se dispone de un odómetro combinado con las tarjetas de vida de los neumáticos, además de las variables de presión y temperatura instantáneas de las bases de datos de sensores de presión y temperatura. Debido a esto, durante este proyecto se dispuso de dos principales *datasets*, ambos en formato *.csv*: *acum\_prom* (tarjetas de vida, odometría y parte de las variables de presión y temperatura) y *acumulados\_G* (data de temperatura y presión actualizada a intervalos de aproximadamente 1 minuto). Estas se componen de las siguientes variables:

- *acum\_prom*:
  1. *tyreserialnumber* (*string*): número de serie del neumático.
  2. *wheelpositionname* (*string*): posición del neumático en el camión.
  3. *td\_press* (*float*): presión instantánea.
  4. *td\_targetpress* (*float*): presión objetivo instantánea.
  5. *td\_temp* (*float*): temperatura instantánea.
  6. *td\_talert* (*bool*): alerta de sobre-temperatura.
  7. *td\_palert* (*bool*): alerta de diferencia de presión.

8. DateTime (*Datetime*): fecha y hora de la medida.
  9. dif\_press (*float*): diferencia entre presión medida y presión objetivo.
  10. PosChange (*bool*): indicador de cambio de posición.
  11. Equipo (*string*): identificador de camión.
  12. Odom (*float*): kilometraje recorrido desde instalación.
  13. neg\_dif\_ac (*float*): suma de diferencias de presión negativas desde instalación.
  14. pos\_dif\_ac (*float*): suma de diferencias de presión positivas desde instalación.
  15. acu\_1-1 (*float*): kilometraje recorrido en posición 1-1 desde la instalación.
  16. acu\_1-2 (*float*): kilometraje recorrido en posición 1-2 desde la instalación.
  17. acu\_2-1 (*float*): kilometraje recorrido en posición 2-1 desde la instalación.
  18. acu\_2-2 (*float*): kilometraje recorrido en posición 2-2 desde la instalación.
  19. acu\_2-3 (*float*): kilometraje recorrido en posición 2-3 desde la instalación.
  20. acu\_2-4 (*float*): kilometraje recorrido en posición 2-4 desde la instalación.
  21. Horas (*int*): cantidad de horas en que el neumático ha estado en uso desde la instalación.
  22. Exterior (*float*): medida de milimetraje del hombro exterior.
  23. Interior (*float*): medida de milimetraje del hombro interior.
  24. prom (*float*): promedio de medidas de milimetraje interior y exterior.
  25. dif\_prom (*float*): diferencia de la variable *prom* entre filas (desgaste ocurrido desde la última medición, variable objetivo *y*).
  26. dif\_hrs (*int*): diferencia de la variable *Horas* entre filas (horas de uso transcurridas desde la última medición).
- *acumulados\_G*:
    1. tyreserialnumber (*string*): número de serie del neumático.
    2. wheelpositionname (*string*): posición del neumático en el camión.
    3. td\_press (*float*): presión instantánea.
    4. td\_targetpress (*float*): presión objetivo instantánea.
    5. td\_temp (*float*): temperatura instantánea.
    6. td\_talert (*bool*): alerta de sobre-temperatura.
    7. td\_palert (*bool*): alerta de diferencia de presión.
    8. DateTime (*Datetime*): fecha y hora de la medida.
    9. dif\_press (*float*): diferencia entre presión medida y presión objetivo.
    10. PosChange (*bool*): indicador de cambio de posición.
    11. Equipo (*string*): identificador de camión.
    12. Odom (*float*): kilometraje recorrido desde instalación.
    13. neg\_dif\_ac (*float*): suma de diferencias de presión negativas desde instalación.

14. `pos_dif_ac` (*float*): suma de diferencias de presión positivas desde instalación.
15. `acu_1-1` (*float*): kilometraje recorrido en posición 1-1 desde la instalación.
16. `acu_1-2` (*float*): kilometraje recorrido en posición 1-2 desde la instalación.
17. `acu_2-1` (*float*): kilometraje recorrido en posición 2-1 desde la instalación.
18. `acu_2-2` (*float*): kilometraje recorrido en posición 2-2 desde la instalación.
19. `acu_2-3` (*float*): kilometraje recorrido en posición 2-3 desde la instalación.
20. `acu_2-4` (*float*): kilometraje recorrido en posición 2-4 desde la instalación.

Como se puede observar, el dataset `acum_prom` ya dispone de parte de las variables presentes en los dataset `acumulados_G` pero, al haber distintos intervalos de actualización, la gran mayoría de las filas presentes en `acumulados_G` no son transferidas (sólo se ha realizado una combinación en base al índice serial-fecha/hora). Sumado a esto, la información que se obtiene de variables instantáneas no es representativa de un proceso de desgaste lento como el estudiado.

## 2.3. Algoritmos de Regresión

Los algoritmos de regresión son métodos usados para estimar la relación subyacente entre dos o más variables cuantitativas, a partir de datos observados. Estas variables se dividirán en variable dependiente ( $y$ ) y variables independientes ( $x$ ), siendo  $y$  continua en el tiempo. Los algoritmos de regresión plantean la hipótesis de que existe una función  $f()$  tal que  $f(x) = y$  y buscan aproximar su estructura y factores.

A lo largo de este trabajo se experimentó con múltiples algoritmos de regresión. Dado su particular rendimiento en las etapas iniciales, los árboles de regresión tuvieron singular énfasis.

### 2.3.1. Regresión Lineal

La regresión lineal es un algoritmo muy comúnmente usado debido a su simpleza de implementación y facilidad de optimización. La regresión lineal busca aproximar la relación entre  $x$  e  $y$  a partir de  $N$  puntos  $(x_i, y_i)_{i=1}^N$ , usando una función lineal. El problema de regresión general se representa con la ecuación:

$$y_i = f(x_i) + \epsilon_i, \quad (2.1)$$

donde  $y_i \in \mathbb{R}$  corresponde al *output* obtenido a partir de la variable de entrada  $x_i \in \mathbb{R}^d$  y  $\epsilon_i$  es una variable aleatoria que representa el ruido de medición (comúnmente se le da una distribución normal  $N(0, \sigma^2)$ ). Para llegar a esta función, la regresión lineal requiere de la parametrización de la función de regresión  $f()$ . A este parámetro le llamaremos  $w \in \mathbb{R}^H$ , con  $H$  la cantidad de funciones  $\{\phi_h(x_i)\}_{h=1}^H$  que definen la parametrización  $f(x_i) := \langle \phi(x_i), w \rangle$ .

### 2.3.1.1. Método *Least Squares*

Considerando que  $X$  es una matriz de  $N \times d$  entradas, e  $y$  es un vector de  $N$  variables de salida, este método busca inferir el vector  $w$  a partir de la minimización de la función

$$E(w) = (Xw - y)^T(Xw - y), \quad (2.2)$$

donde se genera una ecuación cuadrática para  $w$  con una sola solución. Luego, para encontrar esta solución, es posible derivar con respecto a  $w$  e igualar a cero, de la forma:

$$\frac{\delta E(w)}{\delta w} = 2X^T(Xw - y) = 0 \quad (2.3)$$

$$\implies w^* = (X^T X)^{-1} X^T y, \quad (2.4)$$

obteniendo una solución óptima para el vector de pesos  $w$ . En la figura 2.3 se incluye una interpretación geométrica del método de *least squares*.

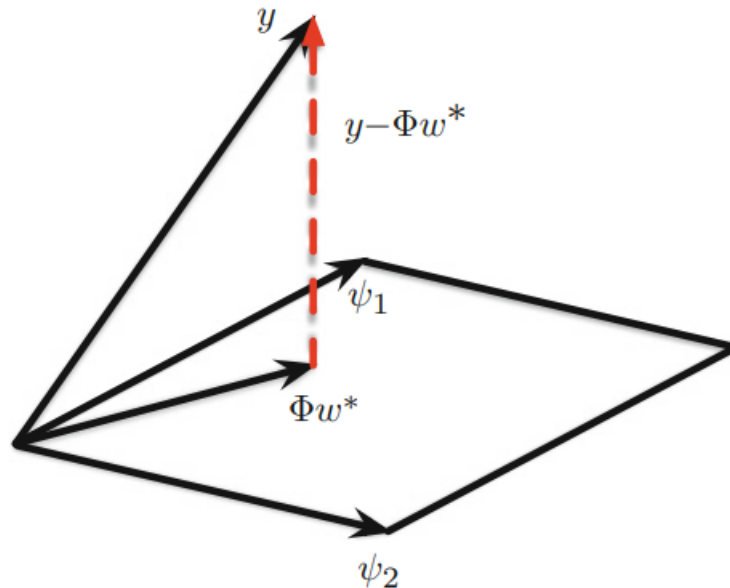


Figura 2.3: Interpretación Geométrica de *least squares* [11].

La solución óptima  $w^*$  usando el método de *least squares* es la proyección ortogonal de la función  $y = f(x)$  en el espacio lineal formado por las funciones paramétricas  $\{\phi_h(x_i)\}_{h=1}^H$ .

## 2.3.2. Árboles de Regresión

Los árboles de regresión son métodos de aprendizaje supervisado, aplicables en múltiples problemas de regresión. Estos modelos igualmente toman como referencia la ecuación (2.1), pero proponen como aproximación un árbol (comúnmente binario)  $\hat{f}$ . El árbol creado presenta una jerarquía de nodos lógicos, donde se definen umbrales de decisión  $\alpha$ . Estos nodos presentan una pregunta del tipo  $x_j < \alpha$  para alguna observación  $x_j$ , según lo cual se avanza en la estructura hasta llegar a una hoja, que contiene la predicción numérica de la variable  $y$ . En la figura 2.4 se muestra un ejemplo de árbol de regresión y su respectiva partición del espacio, donde cada partición representa una predicción para el valor de  $y$ .

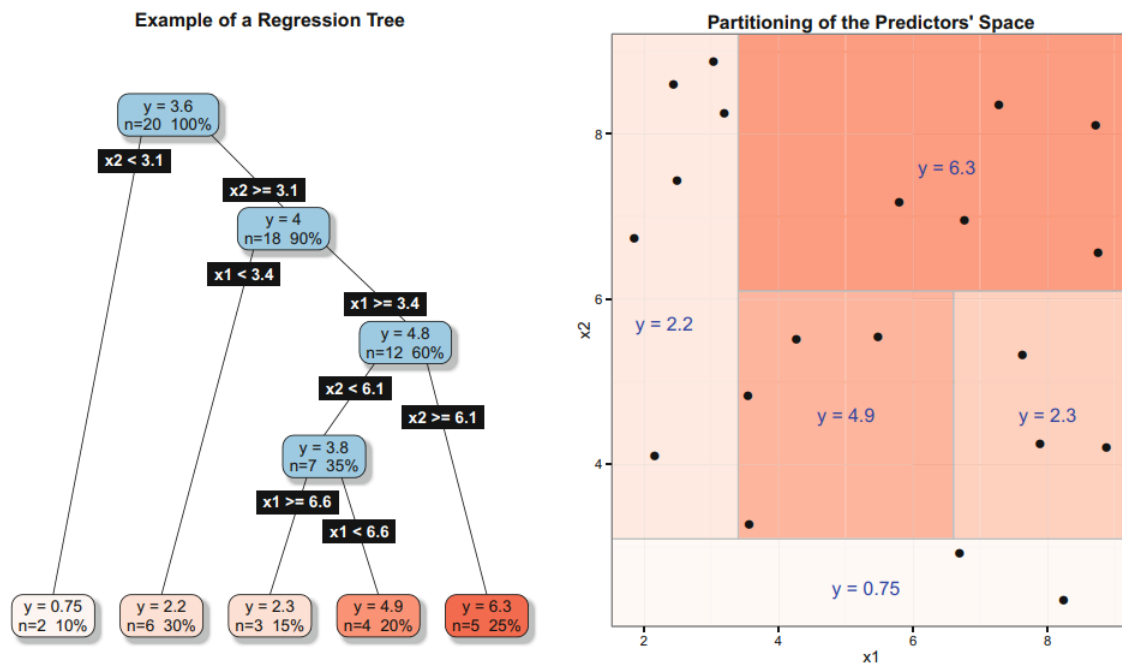


Figura 2.4: Partición del espacio de predicción de un árbol de regresión [11].

Los árboles de regresión dividen el espacio de predicción en el que se encuentra la variable  $y$  de una manera mutuamente excluyente, por lo que cada camino desde la raíz es único y llega a un resultado distinto de predicción para  $y$ . Como este método ofrece un espacio de predicción tan grande como la cantidad de hojas del árbol, puede resultar impreciso para ciertas aplicaciones.

### 2.3.2.1. Partición Recursiva

El algoritmo de entrenamiento de árboles de regresión tiene por nombre Partición Recursiva. En este algoritmo se define un criterio de término  $y$ , mientras este no se cumpla, el espacio de predicción será separado de acuerdo a una pregunta lógica realizada en alguna de las variables de entrada, dividiendo la muestra en dos. Este proceso se repite recursivamente para cada rama, hasta cumplir con el criterio de término del algoritmo.

Para elegir la pregunta que se agregará a continuación, usualmente se selecciona aquella que minimiza en mayor medida la suma de los errores de predicción al cuadrado, conocido como *Residual Sum of Squares* o RSS, similar a lo previamente visto en *least squares*. El cálculo de RSS para  $J$  particiones del espacio de predicción está dado por:

$$\sum_{j=1}^J \sum_{i \in P_j} (y_i - \hat{y}_{P_j})^2, \quad (2.5)$$

donde  $P_j$  es la  $j$ -ésima partición del espacio, e  $\hat{y}_{P_j}$  es la predicción numérica asociada a tal partición. Para el caso del criterio de término, este puede elegirse de tal manera que el árbol tenga una gran cantidad de hojas (mejorando la precisión pero arriesgando *overfitting*), ya que luego este será podado. Por ejemplo, se puede continuar el proceso hasta que cada partición del espacio no tenga más de 10 observaciones.

### 2.3.2.2. Podado (*Pruning*) de Árboles de Regresión

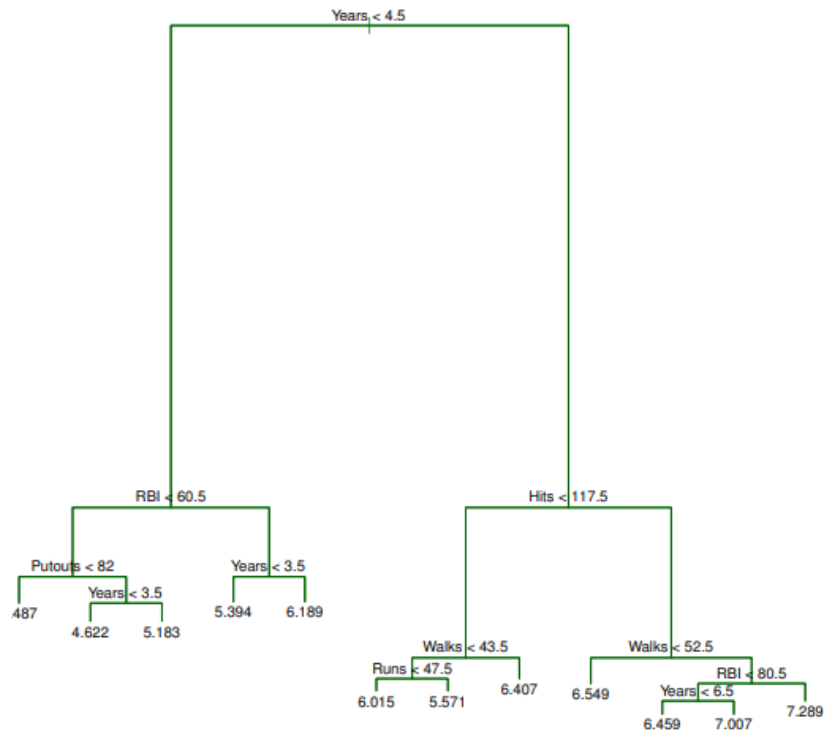
La confección de árboles con gran cantidad de hojas puede aumentar la precisión del modelo en el conjunto de entrenamiento, pero puede llevar a *overfit* y poca precisión en el conjunto de prueba. Ahora, una opción para mejorar el rendimiento de los árboles de regresión es evitar que crezcan demasiado, por ejemplo, terminando la iteración del algoritmo cuando se llega a cierto umbral de ganancia en RSS. Esto, sin embargo, podría reducir las posibilidades de encontrar preguntas que reduzcan en gran medida la medida de RSS, ya que estas pueden suscitarse en cualquier momento de la confección del árbol.

Un acercamiento más útil y por lo tanto más usado, es el de podar árboles de regresión. En este método, se crea en primer lugar un árbol de gran tamaño, que podría causar *overfitting* en el conjunto de prueba y, luego, se reduce la cantidad de hojas, analizando los sub-árboles y sus rendimientos. Comparar los rendimientos de todos los sub-árboles  $T$  posibles dentro de un mismo árbol  $T_0$  sería muy costoso computacionalmente, por lo que se usa un algoritmo llamado *Cost Complexity Pruning*. Para evitar comparar todos los sub-árboles posibles, se considera una secuencia de árboles ordenados por un parámetro  $\alpha$ . Para cada valor de  $\alpha$  existe un sub-árbol  $T \subset T_0$  que minimiza el valor de:

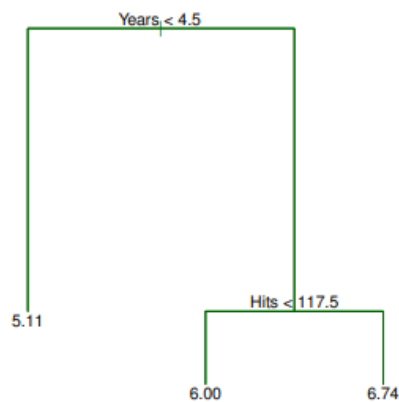
$$\sum_{m=1}^{|T|} \sum_{i: x_i \in P_m} (y_i - \hat{y}_{P_m})^2 + \alpha |T|, \quad (2.6)$$

donde  $|T|$  representa la cantidad de hojas del sub-árbol  $T \subset T_0$ ,  $P_m$  es la partición del espacio de predicción  $m$ -ésima, e  $\hat{y}_{P_m}$  es la predicción numérica asociada a tal partición. Así, el parámetro  $\alpha$  asume un compromiso entre precisión y complejidad del árbol final, al considerar la cantidad de hojas en el árbol como un factor en el cálculo de precisión y así reducir la posibilidad de *overfitting* en el conjunto de entrenamiento. Ahora, para encontrar el mejor valor del parámetro  $\alpha$  se puede hacer uso de *K-fold cross-validation*, es decir, dividir la muestra en  $K$  pliegues y para cada pliegue  $k$  generado, construir un sub-árbol que no lo considere dentro del conjunto de entrenamiento. Seguido de esto, se calcula el MSE de predicción en el pliegue  $k$  previamente ignorado y los  $K$  resultados de MSE son promediados

y guardados como función del parámetro  $\alpha$ , para así encontrar el sub-árbol de menor error cuadrático medio. En la figura 2.5 se presenta un ejemplo de árbol previo a la poda y después de la poda.



(a) Árbol sin podar



(b) Árbol podado

Figura 2.5: Ejemplo de un árbol antes de podar y después de podar, entrenados en un *dataset* de rendimiento de golpeadores de *baseball* [12].

### 2.3.3. Boosting

La técnica de *Boosting* permite mejorar el rendimiento de árboles de decisión y regresión. Este consiste en usar errores de los árboles previamente entrenados para entrenar el árbol



actual. Para ello, se crea un primer árbol usando métodos comunes y, luego, se entrena un nuevo árbol en los errores del anterior, en lugar de entrenarlo en la salida  $y$ . Luego la salida de este árbol se agrega a la función de predicción, ponderada por un coeficiente de aprendizaje  $\lambda$  para ajustar el árbol a los errores. Este es un método de aprendizaje más lento que el entrenamiento normal de un árbol de regresión y permite mejorar los árboles donde tienen mayores problemas.

El algoritmo de *Boosting* requiere de 3 parámetros:

- Cantidad de árboles  $B$ . Este parámetro puede encontrarse usando *cross-validation*, un valor muy grande de  $B$  puede causar *overfitting*.
- Coeficiente de aprendizaje  $\lambda$ . Este parámetro maneja la velocidad de aprendizaje del algoritmo y tiende a ser un valor pequeño como 0.1 o 0.001.
- Número de nodos  $d$  en el árbol. Este parámetro controla la complejidad de los árboles entrenados. Se debe tomar en cuenta que  $d$  nodos pueden considerar hasta  $d$  variables en el aprendizaje.

Cada nuevo árbol creado se suma al modelo de predicción y la importancia de cada uno en la salida se relaciona con el valor de  $\lambda$ . Este método, usualmente, usa *weak-learners*, es decir, árboles de baja profundidad. Siendo  $\hat{f}_b$  el árbol  $b$ -ésimo generado usando este método, el modelo entrenado tendrá la fórmula:

$$\hat{F}(x) = \sum_{b=1}^B \lambda \hat{f}_b(x) \quad (2.7)$$

### 2.3.4. *Adaboost*

Este algoritmo fue el primer algoritmo de *boosting* entrenado exitosamente y consiste en usar los principios de este método en *weak-learners*, es decir, árboles de baja profundidad. En particular, *Adaboost* se enfoca en *stumps*, o árboles de un nodo y dos hojas.

A diferencia del método de *boosting* detallado anteriormente, donde se utilizaba un parámetro  $\lambda$  como la medida de importancia de todos los árboles entrenados, en este caso cada árbol entrenado tiene distinta ponderación en la decisión final, intrínsecamente relacionada con su error de predicción.

Al entrenar un modelo *Adaboost*, se inicia dándole el mismo peso a cada observación  $D_1(i) = \frac{1}{m}$ , con  $m$  la cantidad de observaciones. Luego de esto, se crea un *stump* para cada característica de la muestra, usando los métodos de entrenamiento de árboles de regresión (se elige un umbral  $\alpha$  que minimice el MSE) y se selecciona aquel con un menor error. Este primer árbol presentará errores de predicción en el conjunto de entrenamiento, los que serán usados para darle un parámetro ponderador o capacidad de voto  $\alpha_t$ . En efecto, este parámetro ponderador estará dado por [13]:

$$\alpha_t = \frac{1}{2} \log \left( \frac{1 - \epsilon_t}{\epsilon_t} \right), \quad (2.8)$$

donde  $\epsilon_t$  será la suma de los pesos de las observaciones incorrectamente clasificadas por el árbol. Ahora, para que cada árbol tome en cuenta los errores de sus predecesores, se modifican los pesos  $D_t(i)$  asociados a cada observación. Para ello, se usa la ecuación:

$$D_{t+1}(i) = \frac{D_t(i) e^{-\alpha_t y_i \hat{f}_t(x_i)}}{Z_t}, \quad (2.9)$$

siendo  $\hat{f}_t$  el árbol actual (*stump*) y  $Z_t$  un factor de normalización elegido de tal forma que  $D_{t+1}(i)$  sea una distribución. Esto se repite  $T$  veces, entrenando árboles de regresión con distintas capacidades de voto en la decisión final del modelo. *Adaboost* entrega un modelo:

$$\hat{F}(x) = \text{signo} \left( \sum_{t=1}^T \alpha_t \hat{f}_t(x) \right), \quad (2.10)$$

donde  $T$  es la cantidad de árboles entrenados y  $\hat{f}_t(x)$  es la predicción del árbol  $t$ -ésimo para la entrada  $x$ .

### 2.3.5. Gradient Boost

El algoritmo de *Gradient Boost* consiste en obtener una función  $F(x) = y$  que defina una relación entre las variables  $x$  e  $y$  de un conjunto de datos  $\{(x_i, y_i)\}_{i=1}^n$ , a partir del entrenamiento de árboles de regresión que progresivamente mejoran en rendimiento al minimizar una función de pérdida (*loss function*)  $\Psi(y, F(x))$  diferenciable. Una función de pérdida comúnmente usada es:

$$\frac{1}{2} (y_i - F(x_i))^2, \quad (2.11)$$

que es fácilmente diferenciable y minimiza el error de predicción. Luego de establecer una función de pérdida, el algoritmo requiere de los siguientes pasos [14]:

1. Iniciar el modelo con un valor constante  $\gamma$  que minimice la función de pérdida:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n \Psi(y_i, \gamma) \quad (2.12)$$

2. Este paso se enfoca en el entrenamiento de nuevos árboles de regresión y se iterará  $M$  veces (cantidad de árboles, usualmente  $M = 100$ ).

a) Calcular los pseudo-residuales  $y_{lm}$  a partir del cálculo del gradiente

$$y_{lm} = - \left[ \frac{\delta \Psi(y_i, F(x_i))}{\delta F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad (2.13)$$

- b) Entrenar un árbol de regresión con los valores de  $y_{lm}$ , anteriormente calculados y crear regiones terminales  $R_{lm}$  para  $l = 1, \dots, L$ , con  $L$  la cantidad de hojas del árbol.
- c) Determinar valores de salida para las  $L$  hojas del árbol de regresión. Para ello, se calcula un valor  $\gamma_{lm}$  asociado al árbol  $m$  y la hoja  $l$ , que minimiza el valor de la función de pérdida. Es decir:

$$\gamma_{lm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{lm}} \Psi(y_i, F_{m-1}(x_i) + \gamma), \quad (2.14)$$

donde  $F_{m-1}(x_i)$  corresponde a la predicción del árbol  $m - 1$  para  $x_i$ . De esta forma, el valor de  $\gamma_{lm}$  minimiza la distancia de la predicción  $F(x)$  al valor de  $y_i$ .

- d) Guardar el valor de la predicción  $m$ , que considera las predicciones de los árboles anteriores. Esto se calcula con la ecuación:

$$F_m(x) = F_{m-1}(x) + \nu \sum_{l=1}^L \gamma_{lm} I(x \in R_{lm}), \quad (2.15)$$

donde  $\nu$  es un coeficiente de aprendizaje que define la velocidad de entrenamiento del modelo y busca reducir las posibilidades de *overfitting*. Usualmente es un valor pequeño como 0.001 o 0.1.

3. Entregar el resultado de  $F_M(x)$ , es decir, la predicción de la función una vez ajustada por los  $M$  árboles de regresión entrenados.

### 2.3.6. *Extreme Gradient Boost (XGBoost)*

*XGBoost* es un algoritmo de regresión basado en el previamente visto *Gradient Boost*, pero ocupa un tipo especial (y único) de árboles de regresión. Al igual que en el método de *Gradient Boost*, se elige una función de pérdida  $\Psi(y_i, F(x_i))$ , que regularmente sigue la ecuación ya vista en 2.11. Similar a los anteriores modelos, *XGBoost* realiza predicciones a partir de la suma de  $K$  funciones [15]:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(x_i), \quad (2.16)$$

donde  $f_k \in \mathcal{F}$ , siendo  $\mathcal{F}$  el espacio de los árboles de regresión. Ahora, para encontrar las funciones  $f_k$ , se minimiza la función objetivo:

$$\mathcal{L}(\phi) = \sum_i \Psi(y_i, F(x_i)) + \sum_k \Omega(f_k) \quad (2.17)$$

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2, \quad (2.18)$$

donde  $T$  es la cantidad de hojas del árbol de regresión,  $\gamma$  un parámetro dirigido a penalizar la cantidad de hojas del árbol (incentiva el podado) y  $\lambda$  un parámetro que penaliza la complejidad del árbol al multiplicarse por el término  $\|w\|^2$ , donde  $w$  es el *output* de la hoja. De esta forma, se toma en cuenta tanto la precisión del árbol con la función de pérdida, como la complejidad del mismo con el término de regularización  $\Omega(f)$ . Ahora, para lograr minimizar la función  $\mathcal{L}(\phi)$ , se agrega la función del árbol actual  $f_t$ :

$$\mathcal{L}(\phi) = \sum_{i=1}^n \Psi(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t), \quad (2.19)$$

siendo  $\hat{y}_i^{(t-1)}$  la predicción del árbol  $t-1$ . Seguido de esto, se realiza una expansión polinomial en serie de Taylor de segundo orden. Como esta serie presenta términos constantes, es posible simplificarla, resultando la función objetivo:

$$\bar{\mathcal{L}}(\phi) \simeq \sum_i g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) + \Omega(f_t), \quad (2.20)$$

donde  $g_i = \delta_{\hat{y}^{(t-1)}} \Psi(y_i, \hat{y}_i^{(t-1)})$  y  $h_i = \delta_{\hat{y}^{(t-1)}}^2 \Psi(y_i, \hat{y}_i^{(t-1)})$ . Al derivar e igualar a cero esta ecuación, se obtiene un valor óptimo para el peso de la hoja  $j$ -ésima  $w_j$ :

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}, \quad (2.21)$$

siendo  $I_j = \{i | q(x_i) = j\}$  y  $q$  la estructura del árbol ( $q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T$ ). Como se puede observar en esta ecuación, el peso óptimo de la hoja corresponde (usando la función de pérdida especificada anteriormente en 2.11), a la suma de los errores de predicción (*residuals*), dividida por la cantidad de valores de la hoja, más el parámetro  $\lambda$ . Ahora, se obtiene un puntaje para la estructura  $q$  del árbol creado con la ecuación:

$$\tilde{\mathcal{L}}^t(q) = -\frac{1}{2} \left( \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} \right) + \gamma T, \quad (2.22)$$

lo que permite comparar el rendimiento de múltiples árboles candidatos y elegir aquel de mayor puntaje. Luego de haber entrenado  $K$  árboles de regresión, se realizan predicciones mediante la ecuación planteada en 2.16.

## 2.4. Análisis de modelos

En esta sección se mostrarán las métricas y técnicas usadas para evaluar el rendimiento de los algoritmos de regresión entrenados, así como herramientas útiles para reconocer la importancia de las variables de entrada  $x_i$  en la predicción.

### 2.4.1. Mean Squared Error

El error cuadrático medio (*Mean Squared Error*) es una métrica comúnmente usada para evaluar algoritmos de regresión y permite comparar las predicciones realizadas en el conjunto de test  $\hat{y}_i$  con los valores observados  $y_i$  [16]. El cálculo de MSE se lleva a cabo mediante la ecuación:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}, \quad (2.23)$$

con  $n$  la cantidad de predicciones realizadas.

### 2.4.2. GridSearchCV

*GridSearchCV* es una herramienta libremente disponible en la librería *scikit-learn* y permite realizar *cross-validation* en un modelo con distintos parámetros para encontrar la mejor configuración de entrenamiento. La función *GridSearchCV* recibe una red de parámetros con distintas alternativas para cada hiperparámetro del modelo y luego entrena el algoritmo especificado con todas las combinaciones posibles de la red. Se obtiene un puntaje para cada posibilidad de combinación de hiperparámetros y se entrega como resultado la mejor opción (en el caso de regresión, es aquella de menor MSE). Esto permite encontrar, fácilmente, la mejor configuración para el algoritmo de regresión utilizado. En el código 2.1 se muestra un ejemplo de resultado posible de obtener usando *GridSearchCV*.

Código 2.1: Ejemplo de resultado que puede ser obtenido usando *GridSearchCV*, donde se busca encontrar los mejores valores para parámetros “C”, “gamma” y “kernel” [17].

```
1 Best parameters set found on development set:
2
3 {'C': 10, 'gamma': 0.001, 'kernel': 'rbf'}
4
5 Grid scores on development set:
6
7 0.986 (+/-0.016) for {'C': 1, 'gamma': 0.001, 'kernel': 'rbf'}
8 0.959 (+/-0.028) for {'C': 1, 'gamma': 0.0001, 'kernel': 'rbf'}
9 0.988 (+/-0.017) for {'C': 10, 'gamma': 0.001, 'kernel': 'rbf'}
10 0.982 (+/-0.026) for {'C': 10, 'gamma': 0.0001, 'kernel': 'rbf'}
11 0.988 (+/-0.017) for {'C': 100, 'gamma': 0.001, 'kernel': 'rbf'}
12 0.983 (+/-0.026) for {'C': 100, 'gamma': 0.0001, 'kernel': 'rbf'}
13 0.988 (+/-0.017) for {'C': 1000, 'gamma': 0.001, 'kernel': 'rbf'}
14 0.983 (+/-0.026) for {'C': 1000, 'gamma': 0.0001, 'kernel': 'rbf'}
15 0.974 (+/-0.012) for {'C': 1, 'kernel': 'linear'}
16 0.974 (+/-0.012) for {'C': 10, 'kernel': 'linear'}
17 0.974 (+/-0.012) for {'C': 100, 'kernel': 'linear'}
18 0.974 (+/-0.012) for {'C': 1000, 'kernel': 'linear'}
```

### 2.4.3. *Feature Importances*

La librería *Scikit-Learn* dispone de múltiples funciones para el reconocimiento de importancia de las variables de entrada en un modelo. En particular, para árboles de regresión, se cuenta con la propiedad `feature_importances_`. Esta se basa en el algoritmo *CART*, que asigna puntaje a las variables según la reducción en el criterio utilizado para puntuar los nodos del árbol, es decir, aquellas variables que reducen en mayor medida el error del modelo. En el código 2.2 se muestra un ejemplo de la sintaxis usada para ejecutar código de la librería `feature_importance_`.

Código 2.2: Ejemplo de código usado para cálculo de `feature_importance_` [18].

```
1 model.fit(X, y)
2 # get importance
3 importance = model.feature_importances_
4 # summarize feature importance
5 for i,v in enumerate(importance):
6     print('Feature: %0d, Score: %.5f' % (i,v))
```

### 2.4.4. *SHAP Values*

*SHAP Values* es una estructura de métricas que facilita la interpretación del efecto de las variables de entrada en el *output* del modelo entrenado. *SHAP* le asigna a cada variable un valor de importancia en una predicción particular y ofrece múltiples opciones de visualización de los resultados. Este método de interpretación ofrece tres principales ventajas:

1. Interpretabilidad global, ya que muestra la contribución de cada variable, además de indicar si tienen una relación positiva o negativa con la variable predicha.
2. Interpretabilidad local, porque cada variable tiene su propio conjunto de valores *SHAP*.
3. Aplicable en cualquier modelo basado en árboles de regresión.

El cálculo de *SHAP Values* sigue la ecuación [19]:

$$\phi_i(v) = \sum_{S \subset N \setminus \{i\}} \frac{|S|!(N - |S| - 1)!}{N!} (v(S \cup \{i\}) - v(S)), \quad (2.24)$$

siendo  $N$  la cantidad de variables,  $S$  un sub-conjunto de las mismas y  $v(S)$  el valor total de las variables  $S$ . Cuando una variable  $i$  se agrega al sub-conjunto  $S$ , su contribución marginal corresponde a  $v(S \cup \{i\}) - v(S)$ , por lo que esta ecuación permite calcular la importancia de cada variable al agregarla al conjunto de entrenamiento. La librería de *SHAP Values* de Python usada, ofrece múltiples opciones de visualización de los resultados, como la jerarquía de importancia de variables y su efecto en el resultado según su valor. En la figura 2.6 se muestra un ejemplo de resultado posible de obtener usando la librería de *SHAP Values* de Python.

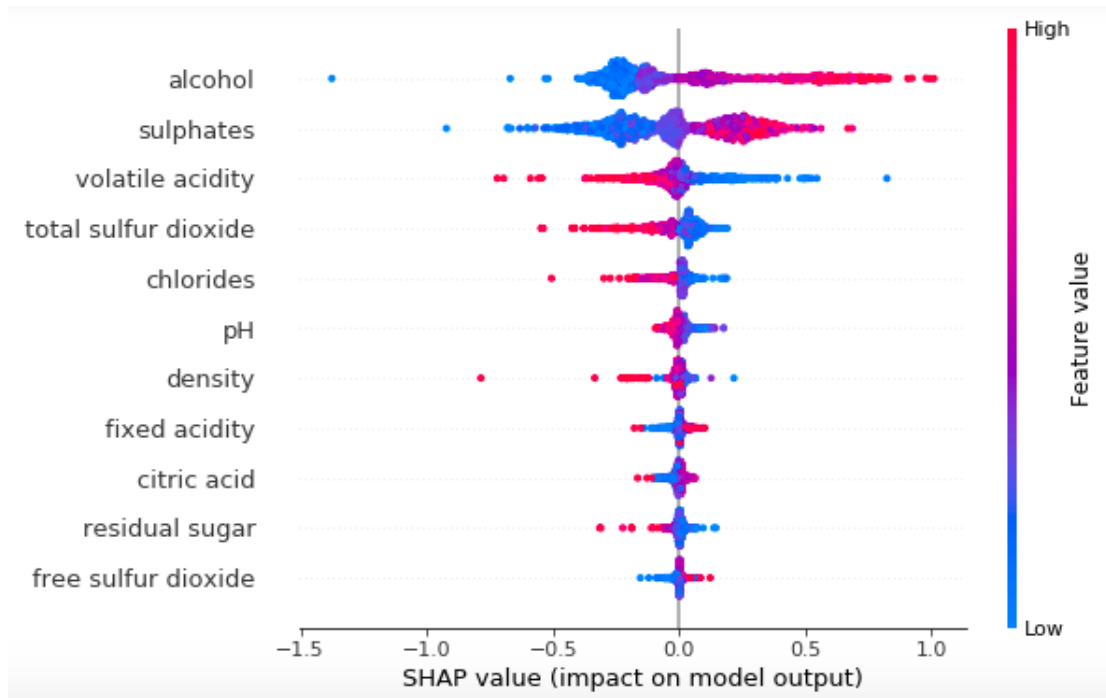


Figura 2.6: Ejemplo de visualización de resultados de la librería de *SHAP Values* de Python. Las variables aumentan en importancia mientras más arriba se encuentran en el gráfico, cada punto representa una observación de la variable y su color se relaciona con el valor de la misma. Mientras más a la izquierda se encuentran los puntos, más restan al resultado y mientras más a la derecha estén, más suman [20].

# Capítulo 3

## Estado del Arte

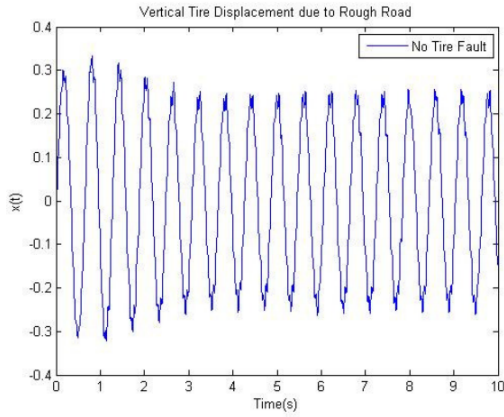
Este trabajo de título se enfoca en un problema particular, relacionado con el aprovechamiento de data disponible de neumáticos de CAEX de una minera, para el desarrollo de un "sensor virtual". En este contexto, soluciones que involucran un estudio a fondo de los caminos u otro análisis en terreno son descartadas, a pesar de los posibles beneficios que puedan ofrecer. Por otro lado, dado que se busca reducir al mínimo posible la cantidad de inspecciones de neumáticos, se descartan soluciones que requieran poner operarios en terreno para hacer mediciones. En este sentido, una solución como la propuesta en este trabajo no presenta precedentes significativos en la industria de la minería chilena y, dadas sus restricciones y bajo costo de implementación, no se compara directamente en rendimiento con otras soluciones propuestas.

A continuación, se enumeran algunos de los estudios más importantes desarrollados en el tópico de aprovechamiento de neumáticos en minería.

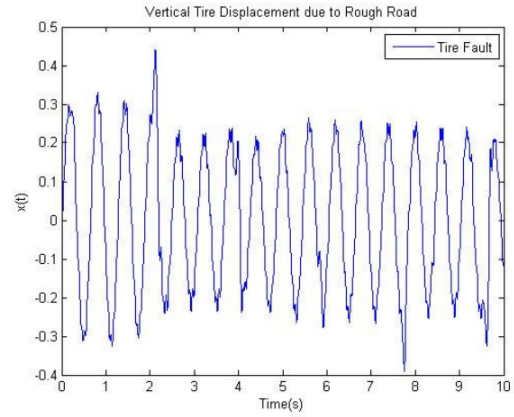
### 3.1. *“Haul truck tyre dynamics due to tire condition”* [21].

En este estudio, se explora la condición de neumáticos y posibilidad de fallas, a partir del análisis de señales de movimiento en respuesta a perturbaciones (rocas/hoyos). Se propone que los neumáticos en buen estado responden de forma distinta que los neumáticos que presentan fallas, lo que es posteriormente demostrado. Para ello, se desarrolla un modelo de la dinámica de los neumáticos de camiones mineros, basados en la física del proceso, que muestra una diferencia en el desplazamiento vertical de los neumáticos con falla y aquellos en buen estado al ser perturbados. Este documento se enfoca en la detección de fallas y se aleja del objetivo particular del proyecto en cuestión. Además, la implementación de este algoritmo en terreno requeriría de la instalación de sensores de movimiento en neumáticos, lo que aumentaría la inversión necesaria para llevar el proyecto a ejecución. En la figura 3.1 se muestran resultados obtenidos en este estudio, al comparar perturbaciones en neumáticos con y sin falla.





(a) tire without fault



(b) tire with fault

Figura 3.1: Desplazamiento vertical de un modelo basado en la física de neumáticos de CAEX, al considerar un neumático en buen estado y otro con falla, en un camino pedregoso [21].

### 3.2. *“Reducing GHG emissions through efficient tire consumption in open pit mines” [22].*

En este paper se plantea la posibilidad de reducir emisiones en una mina a partir del uso eficiente de neumáticos. Bajo esta premisa, plantea la posibilidad de normalizar caminos usando material removido previamente, con el objetivo de desgastar neumáticos más lentamente. Los autores de este documento ponen a prueba esta teoría en una mina en Brasil, obteniendo los resultados esperados: las emisiones efectivamente disminuyeron luego de normalizar los caminos de la mina, como consecuencia de la mejora de rendimiento de los neumáticos. A pesar de que este documento se acerca más a lo buscado en el proyecto actual, ya que se enfoca en la mejora de rendimiento de los neumáticos y no en la detección de fallas, involucra un estudio de caminos de la mina, además del trabajo agregado en la normalización de los mismos. Esto la posiciona como una alternativa infactible en el trabajo de título presente. En la figura 3.2 se muestra la reducción efectiva de emisiones de la mina gracias a las mejoras propuestas por este paper.

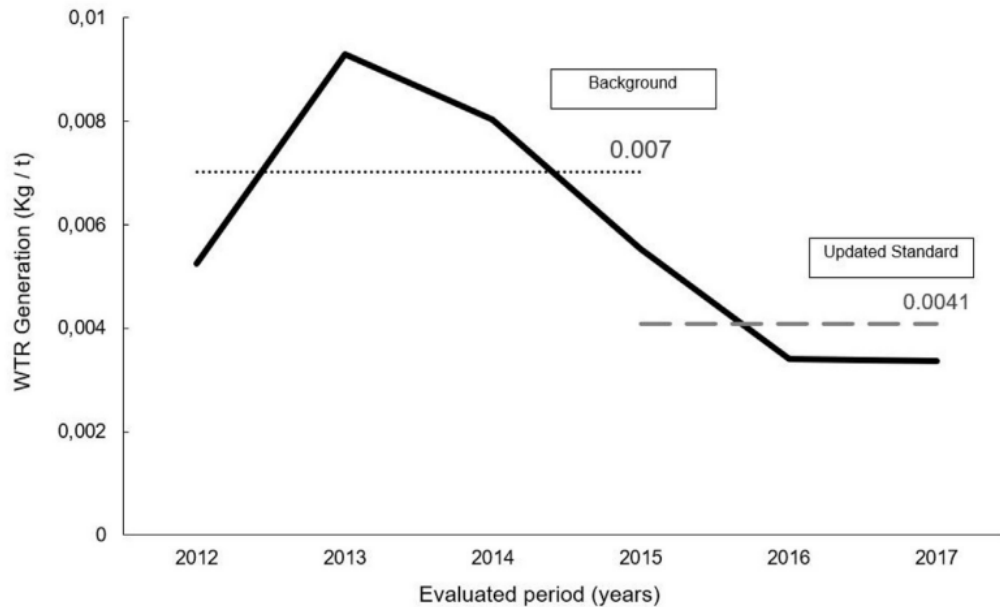


Figura 3.2: Progreso y disminución de residuos generados en una mina de Brasil, medido como WTR (*Waste Tire Rubber*), es decir, la masa de neumáticos descartados dividida por la producción de la mina [22].

### 3.3. *"Predicting the effect of operating parameters on the radial tire treadwear using response surface method" [23].*

Este estudio busca determinar la influencia de variables operacionales en el desgaste de goma ocurrido en la banda de rodadura de un neumático de calle. En particular, se seleccionaron las variables presión de inflado, velocidad y distancia recorrida; y se usaron neumáticos de una misma marca y modelo. Para analizar los efectos de estos parámetros en el desgaste ocurrido en la banda de rodadura, se fabricó una plataforma experimental que reduce la influencia de otras variables y que permite operar un neumático a distintas velocidades y distancias de manera exacta. Se determinó que la variable que más afecta el desgaste del neumático es la velocidad, posiblemente por el mayor roce con el asfalto que deviene de una mayor velocidad. A pesar de la pertinencia de este estudio con el problema de este trabajo, la plataforma experimental fabricada por los investigadores requiere de mayor desarrollo para ser usada aquí, ya que no es directamente aplicable a neumáticos del tamaño de los que han sido usados en este trabajo. Sumado a esto, el desgaste medido en el estudio aludido corresponde al de la banda de rodadura, lo que igualmente difiere de la variable que se busca predecir en este caso. Sin embargo, las variables usadas para detectar el desgaste ocurrido sí son del interés de este trabajo, por cuanto pueden dar pistas de qué relación guardan con este y su importancia. En la figura 3.3 se muestra cómo se configuró el experimento realizado en este estudio.



Figura 3.3: Plataforma experimental fabricada para medir desgaste ocurrido en un neumático y relacionarlo con variables operativas como presión, distancia y velocidad [23].

### 3.4. Aumento de vida útil de los neumáticos para la flota de transporte de Minera Centinela [24].

Este estudio corresponde a la tesis de grado de Juan Ignacio Morales Navarro, de la Universidad de Chile. En él, se buscó mejorar la eficiencia en el uso de neumáticos de CAEX en la minera Centinela, en el año 2020, debido a la gran cantidad de bajas de neumáticos ocurrida en 2018. Para ello, se proponen mejoras en 4 ejes: caminos, prácticas operacionales, operación del neumático y modelo de negocio. El trabajo aludido hace uso de métricas como TKPH (Tonelada–Kilómetro por Hora) y las obtenidas del sistema *RACK–BIAS–PITCH* (RAC), que se encuentra en suspensiones Caterpillar y permite inferir malas prácticas operacionales del conductor de la máquina. Este trabajo es extensivo y considera múltiples áreas de influencia en el desgaste de neumáticos de CAEX, por lo que es una base de conocimiento de gran utilidad. Sin embargo, las medidas aplicables en este trabajo son escasas ya que, como fue mencionado previamente, el estudio y normalización de caminos escapa al presupuesto y tiempo del que se dispone y es en lo que se basan los resultados de la tesis aquí mencionada. En la figura 3.4 se muestra uno de los lineamientos propuestos a la mina en este estudio, con el objetivo de mejorar su uso de neumáticos.



Figura 3.4: Distribución correcta de la carga en un camión de extracción, del “Manual de concientización de neumáticos” de la tesis de grado de Juan Ignacio Morales Navarro [24].

### 3.5. Disminución de los costos por hora en neumáticos de camiones mineros [25]

Este artículo busca resolver el problema de optimizar la vida útil de neumáticos de camiones mineros, para lo cual propone la búsqueda de *insights*, o situaciones donde: se revelan conductas profundas no develadas previamente, se sugiere un potencial para disrupción, se desafían las convenciones actuales de la industria y/o el mercado, se inspira a unos y se crea incomodidad en otros y se crea un momento de sorpresa en el cual se piensa diferente. A partir de esta metodología, se cuestiona la sugerencia técnica de usar los neumáticos durante 1/3 de su vida en posiciones delanteras, para luego ser colocados en posiciones traseras, aludiendo a que esto responde a necesidades de disponibilidad de inventario, más que a una acumulación de esfuerzo. Se plantea el problema de minimizar la cantidad de neumáticos retirados de bodega, con el objetivo de disminuir los costos hora de uso de neumáticos en una mina, para lo cual se calcula una *frecuencia óptima* de recambio de neumáticos. Este estudio busca correlaciones entre costos hora y variables como: la frecuencia de recambio, tiempos de logística por reparación y niveles de inventario de neumáticos reutilizables. A pesar de que el objetivo final de este trabajo de título es obtener una mejora en las operaciones de neumáticos de CAEX en una mina, el estudio aludido será de mayor utilidad en la segunda etapa (creación de modelos de pronóstico para optimización de recursos) y, como fue mencionado en la introducción de este documento, este proyecto se enfoca en la creación de un sensor virtual y no considera la optimización de las prácticas operacionales. En la figura 3.5 se muestra la mejora en costos de operación de neumáticos obtenida en este artículo.

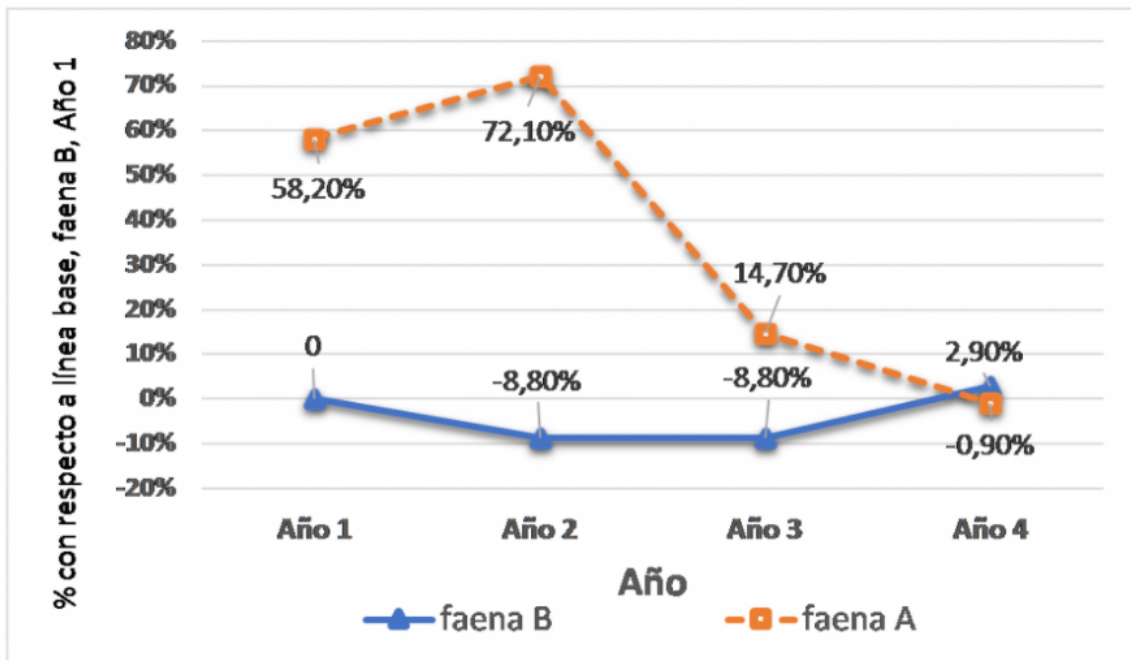


Figura 3.5: Evolución de los costos por hora entre dos faenas mineras, usando el acercamiento de *frecuencia óptima* de recambio de neumáticos [25].

# Capítulo 4

## Trabajo Realizado

### 4.1. Pre-procesamiento de la Data

Previo al entrenamiento de modelos, se debe adaptar la data disponible al problema previamente formulado (1.1). Es decir, el *dataset* usado debe disponer de variables  $x_k$  e  $y_k$  tal que sea posible establecer una relación:

$$y_k = y_{k-1} + f(x_k), \quad (4.1)$$

es decir, la función  $f()$  (modelo predictivo) sea capaz de explicar el desgaste ocurrido entre inspecciones, a partir de variables  $x_k$  que representen este intervalo de tiempo. Para efectos del entrenamiento de modelos, el problema se simplifica de la forma:

$$\Delta y_j = \hat{f}(x_j), \quad (4.2)$$

de manera que  $x_j$  es un vector de variables de operación y  $\Delta y_j$  es la variación de desgaste entre inspecciones. De esta forma, el modelo puede usarse en cualquier instante temporal, sin información del desgaste en  $y_{k-1}$  (una vez implementada la solución).

Se crean las variables:

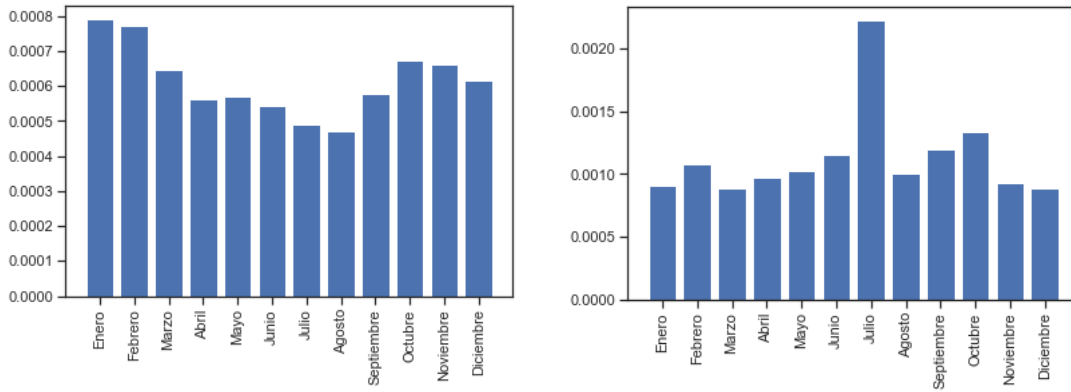
- *dif\_prom*:  $\Delta y_j$ , diferencia de desgaste entre inspecciones.
- *dif\_odom*:  $\Delta Odom$ , kilometraje recorrido desde última inspección.
- *tasa\_desgaste*:  $\Delta y_j / \Delta Odom$ , tasa de desgaste del neumático en función del kilometraje recorrido entre inspecciones.
- *dif\_hrs\_med*:  $\Delta Horas$ , cantidad de horas transcurridas desde la última inspección.
- *explotacion* (en algunos gráficos aparece como *freq\_uso*):  $\Delta Odom / \Delta Horas$ , medida de explotación del neumático, obtenida como kilometraje por hora.
- *press\_odom*:  $P \cdot \Delta Odom$ , presión instantánea del neumático multiplicada por el kilometraje recorrido desde última inspección.

- *acum\_press*:  $\sum \Delta P$ , suma de diferencias de presión respecto al *target* acumuladas desde la instalación del neumático.
- *mes\_actual*: mes en que se toma la medición  $j$ , se separa en 12 variables de tipo *bool*, de forma *one-hot-encoding*, para evitar aprendizaje erróneo de valores numéricos (enero está al lado de diciembre, pero al usar números del 1 al 12, esto no es deducible por el modelo predictivo).
- *mes\_inst*: mes en que el neumático fue instalado, igualmente codificado en *one-hot-encoding*.
- *q1*: cuartil 1 de las tasas de desgaste del mes de la medición  $j$ .
- *q3*: cuartil 3 de las tasas de desgaste del mes de la medición  $j$ .

Se eliminan las filas de datos que contienen NaN luego de este proceso y se elimina la primera fila de datos de cada neumático, ya que, al usar  $\Delta$  de variables, esta primera fila no tiene un antecesor con el cual restarse (y no es posible asumir que el neumático siempre parte con 112[mm], ya que no necesariamente se cuenta con su vida completa), lo que evita anomalías de predicción asociadas al ruido ocasionado por valores nulos. Sumado a esto, se eliminan filas que contengan una tasa de desgaste mayor a 0.2, ya que se consideran como anomalías del *dataset* imposibles de ocurrir en la realidad y confunden a los modelos predictivos (esto ocurre sólo en 3 filas del *dataset*).

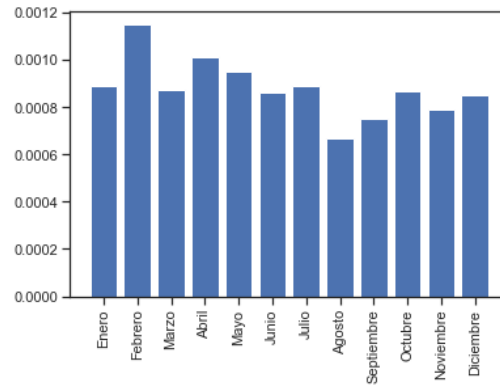
## 4.2. División del *Dataset*

Al analizar el *dataset*, se distinguió una clara diferencia en la tasa de desgaste mensual promedio. Debido a esto, se decidió balancear la data de entrenamiento según el mes de instalación del neumático de manera de obtener un modelo más completo. Es decir, se entrena al modelo con un 60 % de los neumáticos instalados en enero, un 60 % de los neumáticos instalados en febrero, etc. Luego, se realizan predicciones en el 40 % restante de los neumáticos.



(a) 1-1, 1-2

(b) 2-1, 2-4



(c) 2-2, 2-3

Figura 4.1: Medianas de tasas de desgaste mensuales obtenidas con  $\Delta y_j / \Delta Odom$  según posición en el camión, agrupadas en delanteras, traseras externas y traseras internas.

Como se observa en los gráficos de la figura 4.1, las distintas posiciones en el camión y los meses de operación afectan el resultado. Debido a esto, se decide separar el *dataset* en 3: posiciones delanteras, posiciones traseras externas y posiciones traseras internas, por lo que se entrenan 3 modelos predictivos que conforman el modelo general. Esto tiene el objetivo de mejorar la precisión de los algoritmos, ya que al entrenar con data homogénea, estos son capaces de captar de mejor manera los detalles de cada agrupación.

### 4.3. Prueba de Concepto

Para comenzar a iterar sobre este problema, se entrenaron modelos predictivos con la data previamente detallada. Los algoritmos usados son regresión lineal, *Adaboost* y *XGBoost*; y se entrenan 3 modelos en cada caso (delantero, trasero externo y trasero interno). Para mejor visualización del seguimiento del desgaste real del neumático, se crea la variable de desgaste acumulado:



$$\hat{y}_i = \sum_{j=1}^i \Delta y_j, \quad (4.3)$$

donde  $i$  es el indicador del instante de medición. De esta manera, se obtiene una predicción del desgaste que ha sufrido el neumático desde que fue instalado hasta el momento actual. De forma análoga, se obtiene una medida del desgaste real acumulado, como curva de referencia. Usando este acercamiento, se obtuvieron los resultados presentes en la figura 4.2.

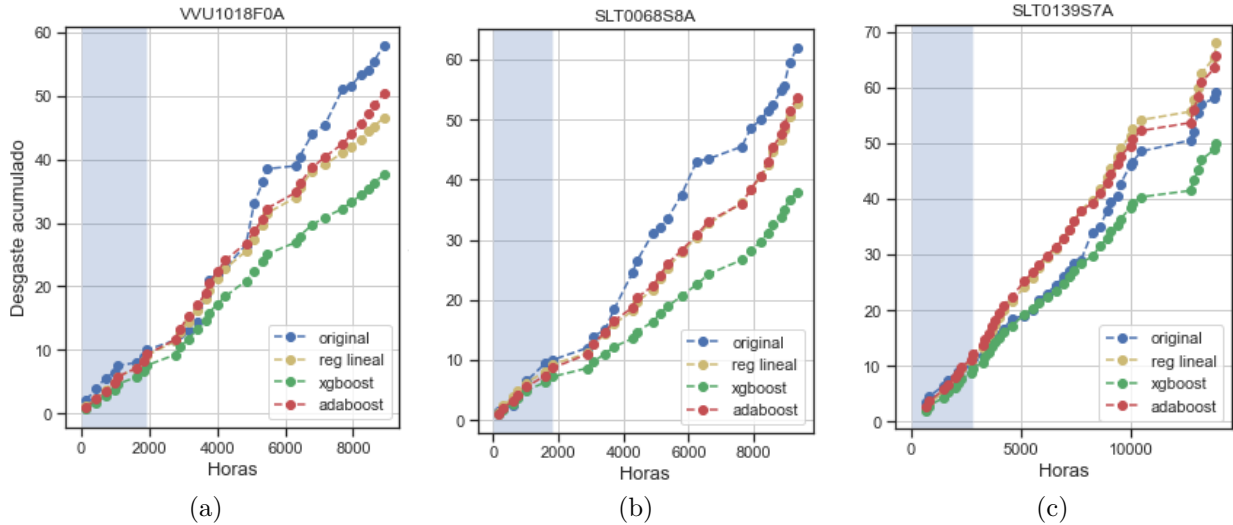


Figura 4.2: Acumulación de desgaste en mm y predicciones de algoritmos regresión lineal, *Adaboost* y *XGBoost*, para 3 neumáticos, en función de las horas que el neumático se mantuvo instalado en el camión. El área destacada representa el espacio de tiempo que el neumático se encontró en posiciones delanteras.

Mientras más se parecen las curvas predichas a las originales (línea azul), mejor es el rendimiento de los algoritmos de predicción. Como se puede observar, tanto la regresión lineal como *Adaboost* presentaron un buen rendimiento a lo largo de este experimento, lo que se puede confirmar al observar los MSE de cada algoritmo, presentes en la tabla 4.1.

Tabla 4.1: Valor de MSE de modelos predictivos según posición en el camión y tipo de algoritmo usado.

| Posición        | XGBoost | Adaboost | Regresión Lineal |
|-----------------|---------|----------|------------------|
| Delantero       | 0.730   | 0.611    | 0.606            |
| Trasero externo | 2.723   | 2.251    | 2.225            |
| Trasero interno | 1.377   | 1.178    | 1.179            |
| General         | 1.619   | 1.354    | 1.343            |

La regresión lineal y *Adaboost* tienen resultados muy similares en términos de error y mantienen un buen rendimiento. A pesar de esto, el seguimiento de referencia no es aún

satisfactorio. Por otro lado, en los resultados se observa que *XGBoost* subestima constantemente el desgaste ocurrido, pero mantiene un error bajo a pesar de esto.

Por otro lado, se detectó una anomalía en el algoritmo *Adaboost*, al analizar sus histogramas de predicción y compararlos con el histograma de valores reales de  $y$ , lo que puede observarse en la figura 4.3.

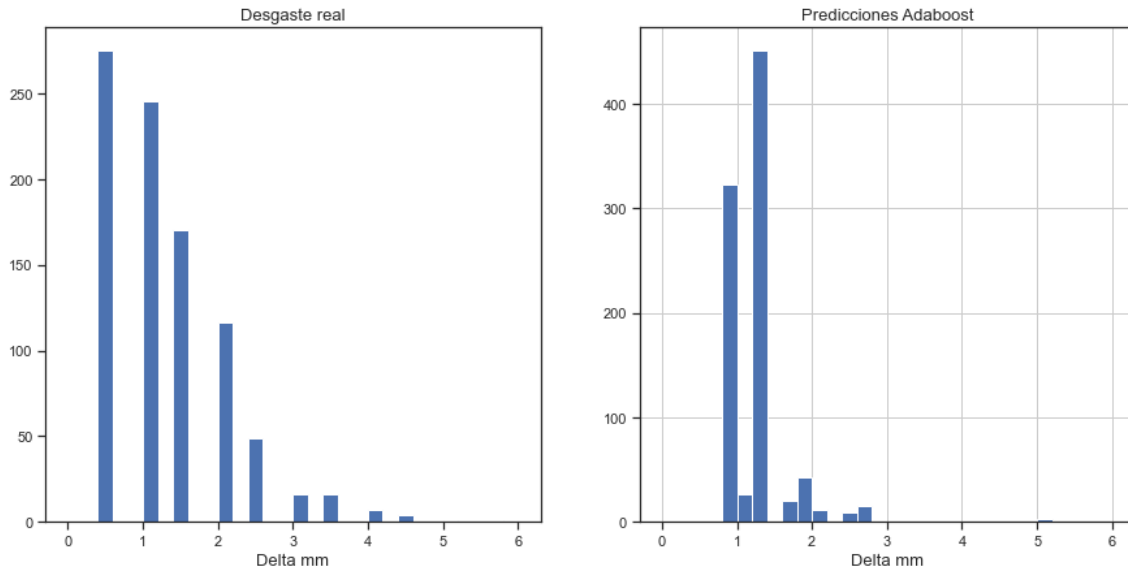


Figura 4.3: Histogramas de  $\Delta y_j$  para la flota de neumáticos y predicciones de *Adaboost* para estos valores.

Al observar estos gráficos, es claro que *Adaboost* es incapaz de captar correctamente las características fundamentales de este proceso y su buen rendimiento se debe a que predice promedios generales de variación de desgaste aleatoriamente, centrándose principalmente en dos valores cercanos a 1. Esto ofrece un error pequeño pero no explica la fenomenología de la mina y es tan eficaz como lanzar una moneda para predecir el desgaste de neumáticos. En el caso de regresión lineal, el error es pequeño pero el seguimiento de referencia es insatisfactorio.

## 4.4. Incorporación de Fenomenología de la Mina

Dado que es posible obtener un buen rendimiento a partir de los promedios de desgaste, se plantea hacer un uso inteligente de estos promedios. Como se observó en la figura 4.1, los neumáticos presentan comportamientos variantes en el tiempo y estos no pueden ser fácilmente explicados con las variables operacionales de las que se disponen. Por ejemplo, en el mes de julio se instalan cadenas en los neumáticos traseros externos, lo que aumenta significativamente la tasa de desgaste de estos y no se encuentra documentado en el *dataset* (ninguna variable indica instalación de cadenas). Situaciones como estas pueden ser múltiples, como la explotación de la mina, temperatura del suelo, condiciones climáticas, entre otras, y la mejor manera de integrarlas al estudio es a través de sus efectos. Es decir, dado que cada

mes de operación representa una situación característica medible solo mediante sus efectos en las tasas de desgaste de los neumáticos, usar estos efectos en el cálculo ofrecería un mejor acercamiento de los modelos al desgaste real medido.

Se divide entonces el modelo en 2 etapas: en una primera etapa, se multiplica el kilometraje recorrido desde la última inspección por la mediana de las tasas de desgaste del mes de operación (la dispersión de las tasas de desgaste evita que el promedio sea la variable más eficaz), obteniendo así una medida de desgaste inicial basada en la mediana de los datos de la flota de neumáticos; y en una segunda etapa, se explica el error de esta estimación respecto del desgaste real, entrenando algoritmos regresivos con las demás variables operacionales (se descarta el odómetro ya usado en la primera etapa, para evitar *overfitting*). De esta forma, el algoritmo predictivo entrenado es capaz de explicar la situación particular de cada neumático y su desviación respecto de la mediana de los datos.

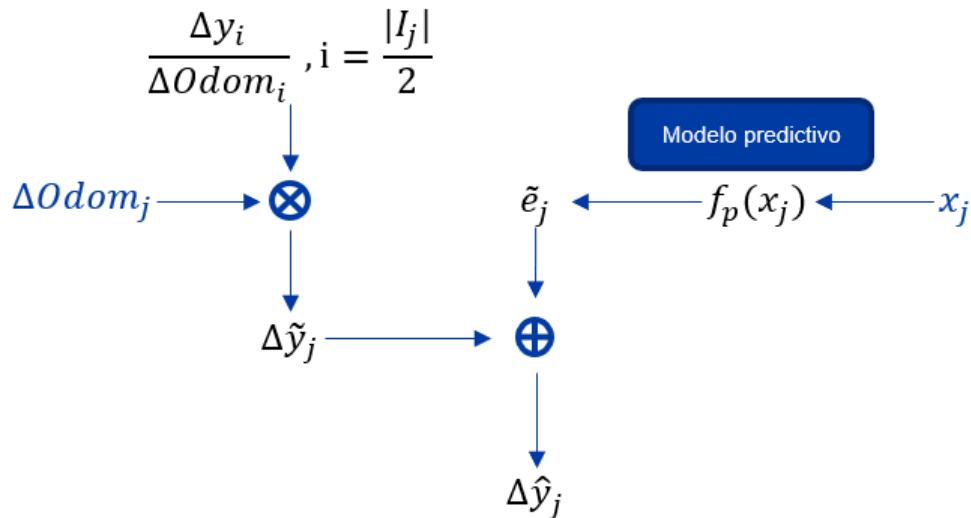
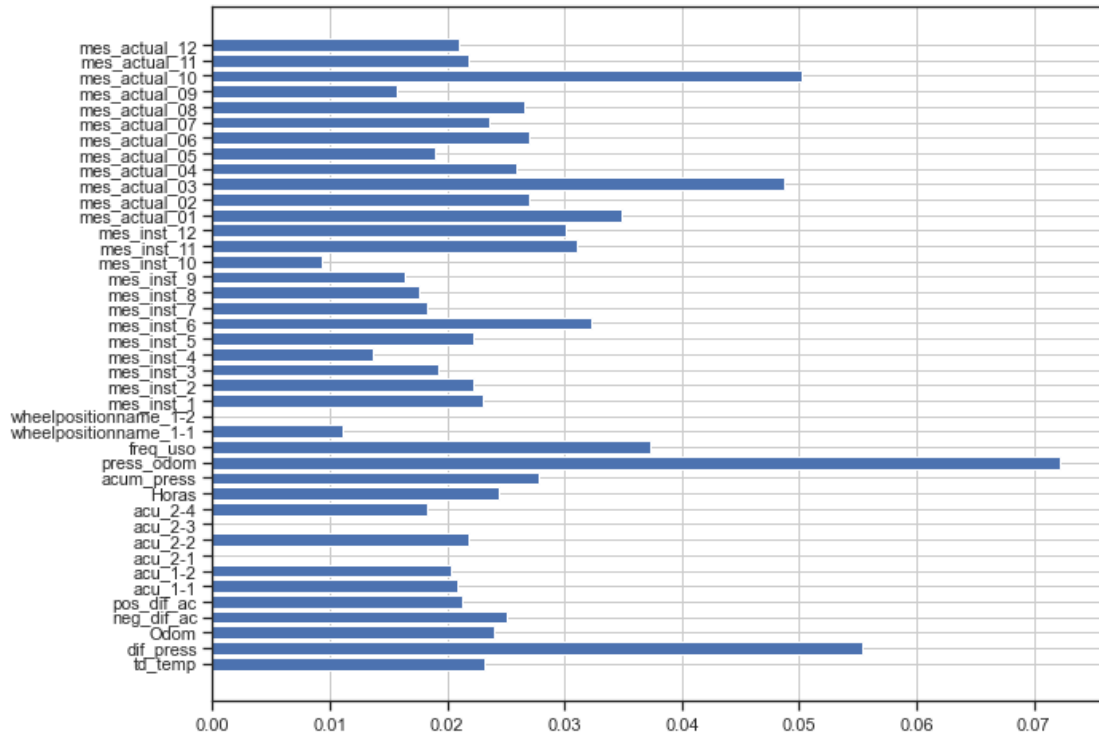


Figura 4.4: Esquema del algoritmo computacional de 2 etapas generado.

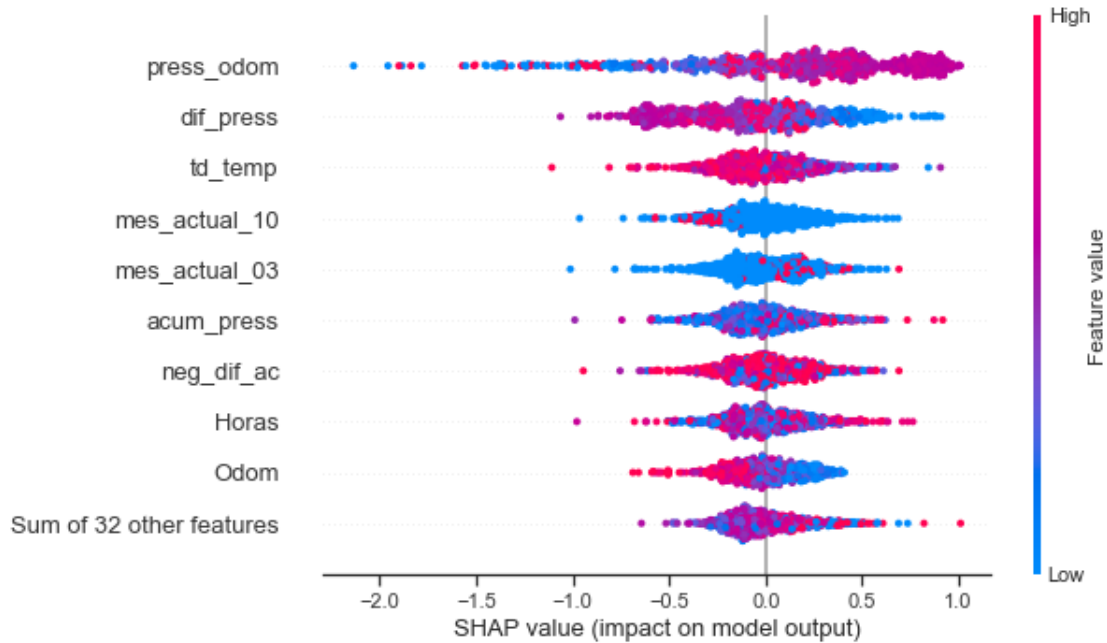
En la figura 4.4, se muestra un esquema del algoritmo computacional programado.  $I_j$  corresponde al conjunto de mediciones de todos los neumáticos para el mes en que se tomó la medida  $j$  (de los 2 años de datos disponibles), considerando que está ordenado de menor a mayor tasa de desgaste,  $\tilde{e}_j$  es la estimación del error respecto del desgaste real,  $f_p$  representa el modelo predictivo entrenado, con  $p$  el indicador de posición del neumático (delantero, trasero externo o trasero interno),  $x_j$  es el conjunto de variables operacionales de la medida  $\Delta y_j$  y  $\Delta \hat{y}_j$  es la predicción final del algoritmo.

## 4.5. Inclusión de Data de Presión y Temperatura

A lo largo del trabajo realizado, se distinguió una distribución homogénea de importancia de las variables operacionales usadas en el resultado final. Pero, el indicador de *feature importance* y los valores *SHAP*, muestran que ninguna de estas variables, por separado, contienen información determinante del proceso. En los gráficos de la figura 4.5 se observan los resultados obtenidos para la importancia de las variables en el modelo entrenado.



(a) Feature importances



(b) SHAP Values

Figura 4.5: Gráfico de *feature importance* y *SHAP Values* para el algoritmo de 2 etapas generado, en posiciones delanteras, usando XGBoost para entrenar el modelo predictivo.

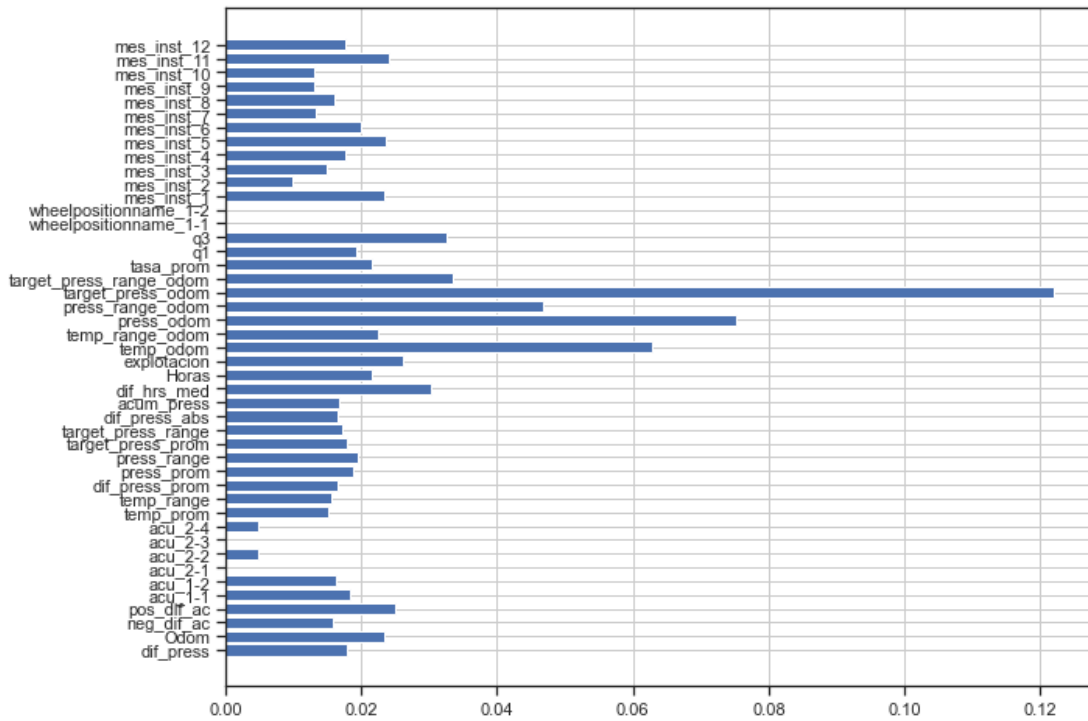
Hasta este instante, se han utilizado variables de presión y temperatura instantáneas del momento de la medición del neumático, que son las más fácilmente transferibles al cruzar los *datasets*, debido a la coincidencia temporal. Pero, es claro que estas mediciones de un instante particular no son representativas del intervalo de tiempo transcurrido desde la última

inspección de los neumáticos, ya que tanto las presiones como las temperaturas varían a lo largo de los días. Bajo esta premisa, las variables que son claras candidatas para añadir al entrenamiento son aquellas que tomen en cuenta estas variaciones de temperatura y presión en un intervalo de tiempo.

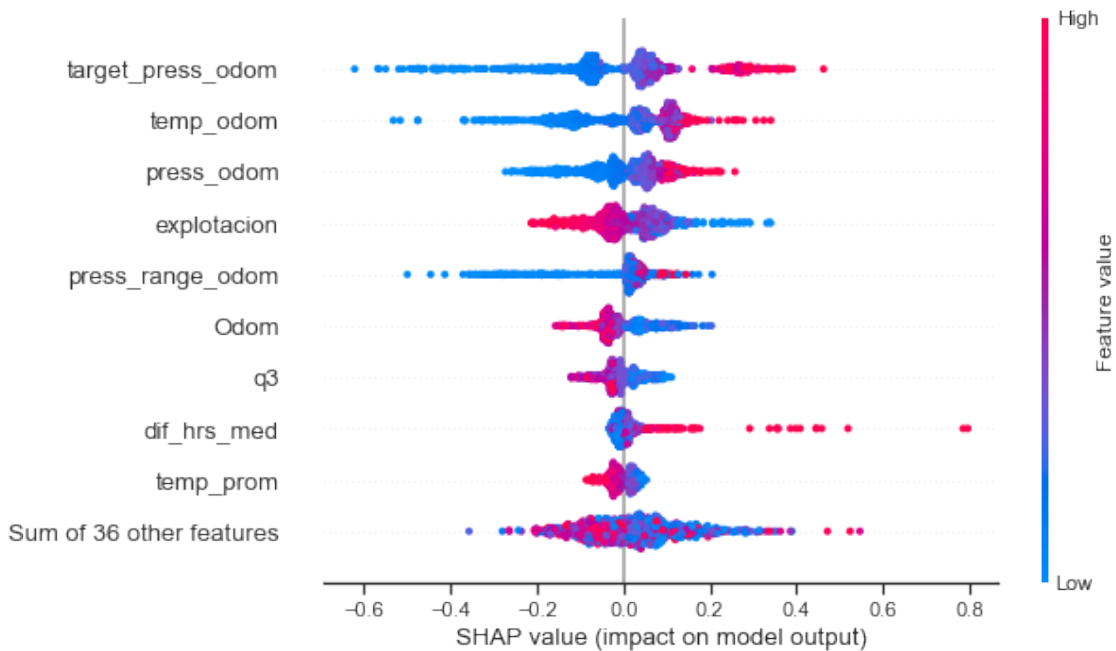
Se crean, entonces, las variables:

- *temp\_prom*: promedio de temperatura del neumático desde la última inspección.
- *temp\_range*: rango de temperatura total alcanzado desde la última inspección (temperatura máxima menos temperatura mínima).
- *press\_prom*: promedio de presión del neumático desde la última inspección.
- *press\_range*: rango de presión alcanzado desde la última inspección (igualmente, presión máxima menos presión mínima).
- *target\_press\_prom*: promedio de presión objetivo desde la última inspección.
- *target\_press\_range*: rango de presión objetivo alcanzado desde la última inspección (nuevamente, presión objetivo máxima menos la mínima).

Luego de obtener resultados con estas nuevas variables, se determinó que no representan una mejora observable en el desempeño de los algoritmos de predicción, lo que resulta extraño, ya que las variables de presión y temperatura han sido, históricamente, las más consultadas para medir el desempeño de neumáticos de CAEX. Dada esta incongruencia en los resultados, se realizaron pruebas añadiendo más información a las variables creadas, modulándolas con el kilometraje recorrido desde la última inspección. Por ejemplo, la teoría indica que a mayor temperatura, mayor es la tasa de desgaste de los neumáticos y, por otro lado, los algoritmos restan desgaste a la predicción a mayor kilometraje (esto se observa en los resultados de *SHAP Values* de la figura 4.5), por lo que se crea una variable que divide la temperatura por el kilometraje recorrido. Esta variable aumentaría a mayor temperatura y menor kilometraje. Por otro lado, se observa que a menor presión los hombros de los neumáticos se ven más afectados, debido a una menor prominencia de la banda de rodadura, por lo que se crea una variable que aumenta a menor presión y menor kilometraje. Siguiendo esta lógica, se crean nuevas variables con el sufijo “\_odom”. En la figura 4.6, se presentan los resultados de *feature importance* y *SHAP Values*, usando el algoritmo *XGBoost*, con estas nuevas variables.



(a) Feature importances



(b) SHAP Values

Figura 4.6: Gráfico de *feature importances* y *SHAP Values* para el algoritmo de 2 etapas generado, usando *XGBoost* como modelo predictivo en posiciones delanteras, con nuevas variables de presión y temperatura.

Es claro, al observar los gráficos, que estas nuevas variables creadas tienen mayor información del proceso que las usadas anteriormente y representan una mejora en el modelo. En efecto, los valores de MSE para el algoritmo entrenado con las variables iniciales y con las nuevas variables, usando *XGBoost* para el modelo predictivo, se observan en la tabla 4.2.

Tabla 4.2: Valor de MSE de modelos predictivos según posición en el camión, al ser entrenados con las variables iniciales y al agregar data obtenida de presión y temperatura.

| <b>Posición</b> | <b>Modelo inicial</b> | <b>Modelo con nuevas variables</b> |
|-----------------|-----------------------|------------------------------------|
| Delantero       | 0.72                  | 0.57                               |
| Trasero externo | 4.09                  | 2.62                               |
| Trasero interno | 1.39                  | 1.10                               |
| General         | 2.09                  | 1.44                               |



# Capítulo 5

## Resultados

Al realizar predicciones con el modelo de 2 etapas (4.4), que separa el *dataset* según posición en el camión, e incluyendo las nuevas variables creadas, se obtienen los resultados de MSE presentes en la tabla 5.1.

Tabla 5.1: Valor de MSE de modelos predictivos según posición en el camión y tipo de algoritmo usado, para la arquitectura de 2 etapas.

| Posición        | XGBoost | Adaboost | Regresión Lineal |
|-----------------|---------|----------|------------------|
| Delantero       | 0.579   | 0.594    | 0.590            |
| Trasero externo | 2.621   | 2.797    | 2.861            |
| Trasero interno | 1.108   | 1.179    | 1.123            |
| General         | 1.448   | 1.536    | 1.538            |

Por otro lado, es posible calcular el valor de MSE para el seguimiento del desgaste real, es decir, comparar las variables acumuladas  $\hat{y}_i$  (predicción del desgaste acumulado desde la instalación) e  $y_i$  (desgaste real acumulado desde la instalación). Para los distintos modelos entrenados, este valor se presenta en la tabla 5.2.

Tabla 5.2: Valor de MSE de todos los modelos predictivos entrenados, midiendo el error del seguimiento de desgaste real acumulado  $y_i$ .

| Modelo                    | MSE    |
|---------------------------|--------|
| 2 etapas XGBoost          | 15.350 |
| 2 etapas Adaboost         | 17.076 |
| 2 etapas Regresión Lineal | 16.956 |
| XGBoost                   | 64.004 |
| Adaboost                  | 16.316 |
| Regresión Lineal          | 18.751 |

Viendo estos resultados, es claro que el modelo de 2 etapas creado, usando *XGBoost* como modelo predictivo, presenta el mejor rendimiento en seguimiento del desgaste real y es el

modelo más completo en términos de la información que recoge del proceso, al considerar tanto la fenomenología subyacente de la mina, como los datos operacionales individuales de los neumáticos de la flota estudiada.

Usando entonces *XGBoost* como modelo predictivo, se obtienen los siguientes resultados para el valor de  $\hat{y}_i$  (4.3) y de  $y_i$ , en 9 de los neumáticos más problemáticos (de un total de 222), se obtienen los resultados presentes en la figura 5.1.

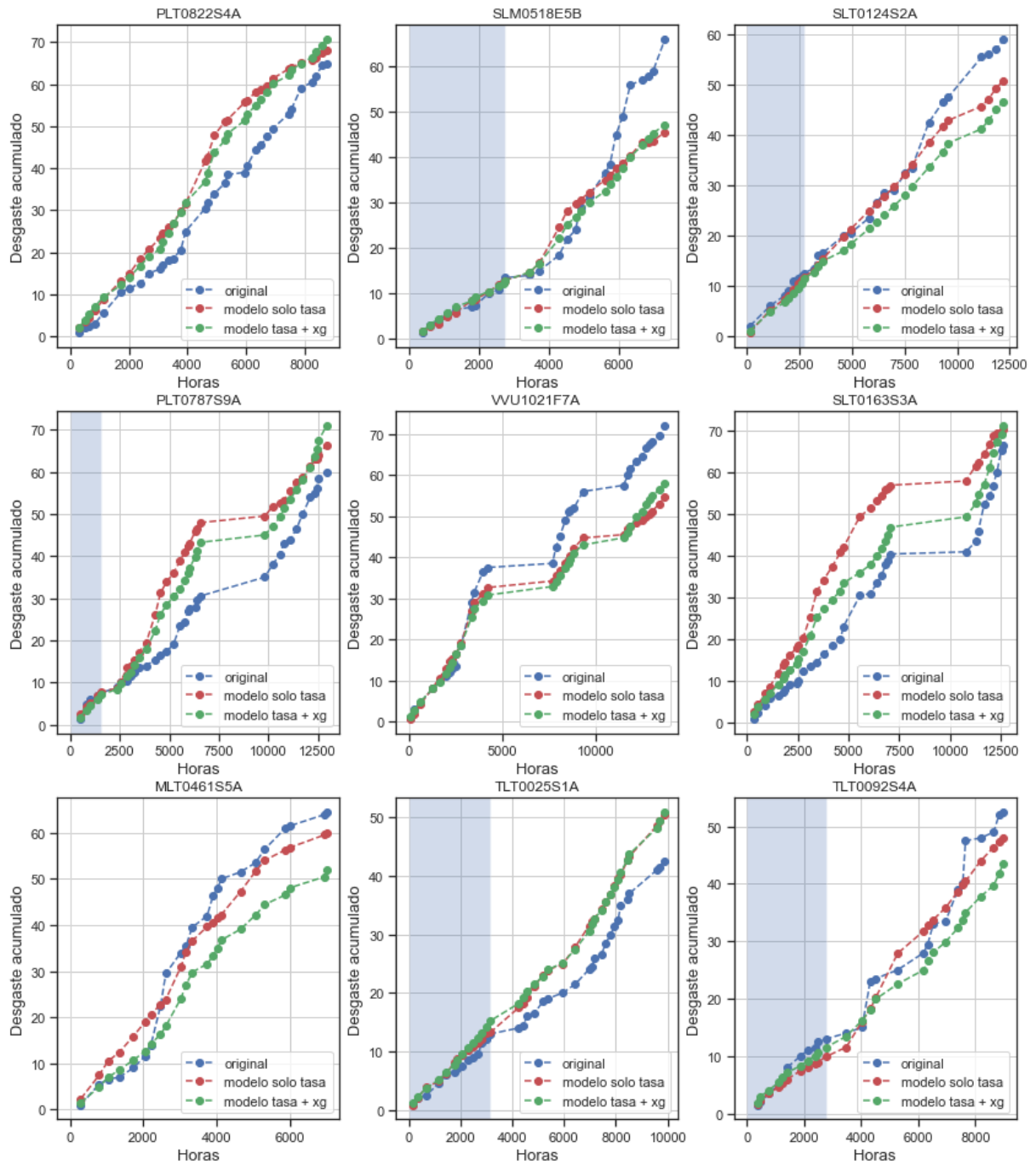


Figura 5.1: Curvas de desgaste predicho (usando solo tasa mensual y, por otra parte, agregando ajuste de XGBoost) y real, para los 9 neumáticos más problemáticos de la flota (mayor error de seguimiento de desgaste real). El área azul representa el periodo que el neumático estuvo en posiciones delanteras.

En la figura 5.2 se muestra el seguimiento del desgaste real para 9 neumáticos de la flota, esta vez escogidos al azar. Se puede observar que la mayoría de las veces el algoritmo sigue la referencia con gran precisión.

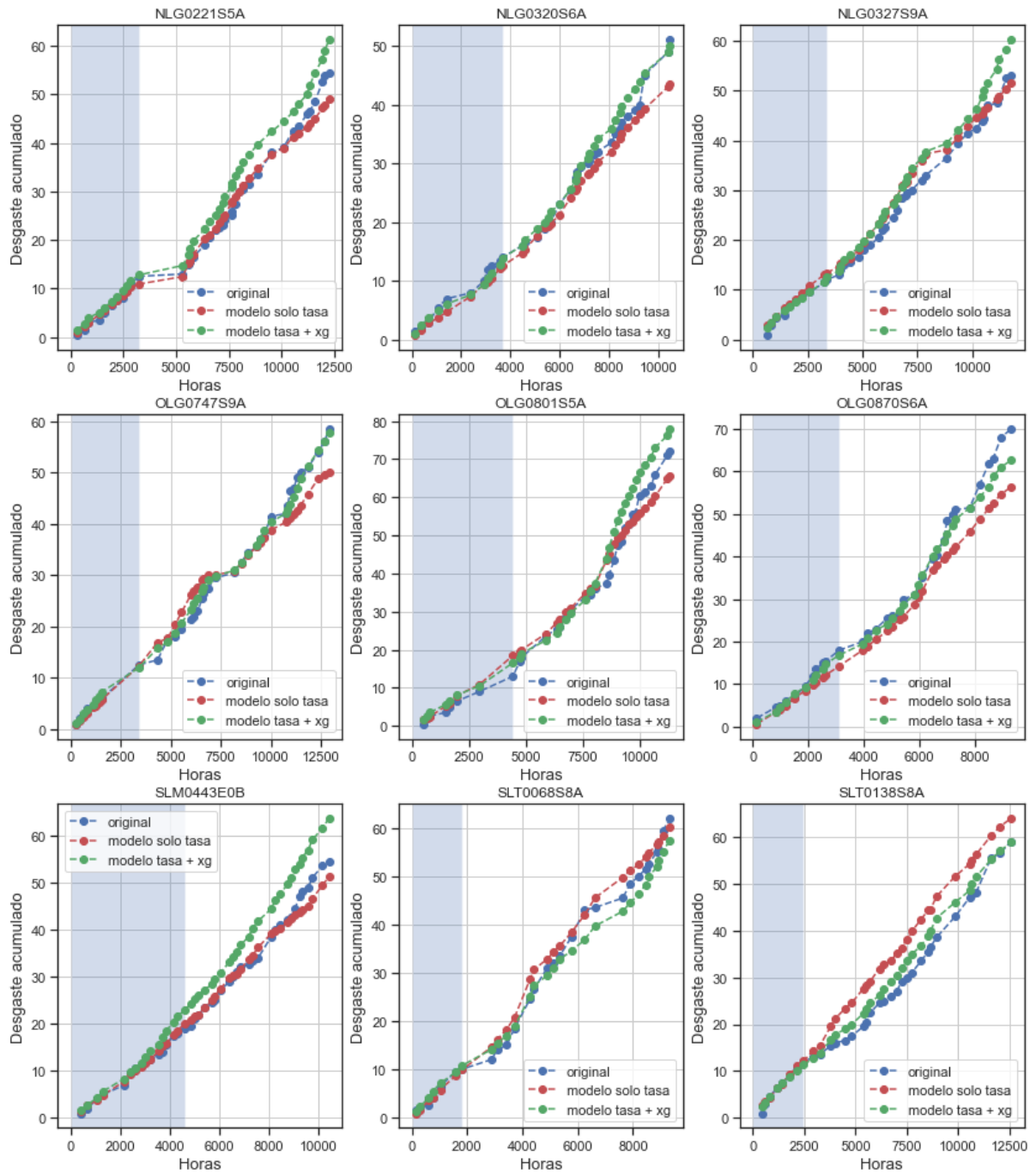


Figura 5.2: Curvas de desgaste predicho (usando solo tasa mensual y, por otra parte, agregando ajuste de XGBoost) y real, para 9 neumáticos escogidos al azar de la flota. El área azul representa el periodo que el neumático estuvo en posiciones delanteras.

# Capítulo 6

## Conclusiones

En este trabajo de título se analizó una flota de neumáticos de camiones de extracción (CAEX), para desarrollar modelos predictivos capaces de detectar el desgaste ocurrido en la goma de los hombros del neumático desde la última inspección. Para ello, se crearon múltiples variables con información parcial del proceso, basándose en la fenomenología del mismo y en los comportamientos mensuales de la mina en estudio. Se decidió separar el algoritmo predictivo en 3 modelos según la posición (delantera, trasera externa y trasera interna) y 2 etapas: una primera etapa que considera la mediana de la tasa de desgaste mensual por kilómetro y una segunda etapa que involucra el uso de modelos predictivos, para estimar el error de la variable obtenida en la primera etapa respecto del desgaste real. Gracias a esta arquitectura, se obtiene un modelo con un MSE de 1.44, en el total de predicciones  $\Delta\hat{y}_j$  y 15.35 en el seguimiento de la curva de desgaste real  $y_i$  de la flota. Se determina, además, que el algoritmo con mejor rendimiento para este caso ha sido *XGBoost*, siendo el más complejo y completo de los usados en este trabajo.

Dado que en este caso se dispone de datos de neumáticos de 2 años, idealmente se realizaría este estudio considerando más épocas, de manera de reducir el error ocasionado por situaciones particulares ocurridas en un año. Esto puede ser analizado en un trabajo futuro.

El modelo generado no se basa en la predicción directa de la variable objetivo a partir de las variables de operación, demostrando que obtener un buen rendimiento en algoritmos predictivos no es tan directo como ingresar los datos en entrenamiento y obtener un buen resultado, sino que requiere de un estudio del proceso y sus características fundamentales.

Los resultados obtenidos, a su vez, demuestran que es posible predecir el desgaste ocurrido en el neumático con gran precisión, usando variables de operación y medianas mensuales de tasas de desgaste por kilómetro, lo que entrega información suficiente de las características de la mina. Así, se logra obtener un modelo fácilmente escalable y aplicable a cualquier operación minera con data suficiente de sus neumáticos (en particular, aquellas con sensores de presión, temperatura y GPS, que ya recogen gran parte de la data necesaria), ofreciendo así una solución de fácil implementación y que no requiere de mayor estudio de caminos de la mina u otro tipo de inversión.

# Bibliografía

- [1] CODELCO, “Rt refuerza mantenimiento de neumáticos de camiones de extracción para aumentar su productividad,” 2016, [https://www.codelco.com/rt-refuerza-mantenimiento-de-neumaticos-de-camiones-de-extraccion-para/prontus\\_codelco/2016-05-04/182325.html](https://www.codelco.com/rt-refuerza-mantenimiento-de-neumaticos-de-camiones-de-extraccion-para/prontus_codelco/2016-05-04/182325.html) (visitado el 5/10/2021).
- [2] Safety, M., “Maintenance guide for earthmover tyres: Tyre inspection,” 2021, <https://www.miningsafety.co.za/maintenance-guide-for-earthmover-tyres-tyre-inspection/> (visitado el 8/10/2021).
- [3] Barrientos, V., “Disminución de los costos por hora en neumáticos de camiones mineros,” 2019, <https://www.mch.cl/reportajes/disminucion-de-los-costos-por-hora-en-neumaticos-de-camiones-mineros/> (visitado el 20/09/2021).
- [4] Wenco, “Think tire temperature matters most? think again.,” 2021, <https://www.wenco.com/post/think-tire-temperature-matters-think> (visitado el 7/10/2021).
- [5] CODELCO, “Monitoreo automatizado de condiciones de neumáticos caex,” 2021, <https://expandemineria.cl/desafio/monitoreo-automatizado-de-condiciones-de-neumaticos-caex/> (visitado el 1/10/2021).
- [6] Nyaaba, W., Scholars’ Mine Thermomechanical fatigue life investigation of an ultra-large mining dump truck tire Wedam Nyaaba Recommended Citation. PhD thesis, 2017.
- [7] Euromaster, “Descubre todas las partes de un neumático,” 2021, <https://www.euromaster-neumaticos.es/blog/partes-de-un-neumatico> (visitado el 27/10/2021).
- [8] Thoeni, K., Servin, M., W.Sloan, S., y Giacomini, A., “Designing waste rock barriers by advanced numerical modelling,” 2019, <https://www.sciencedirect.com/science/article/pii/S1674775518303895> (visitado el 27/10/2021).
- [9] Moore, P., “Michelin introduces its mems 4 tpms adding real-time gps and accelerometer data into a complete platform,” 2017, <https://im-mining.com/2017/10/13/michelin-introduces-mems-4-tpms-adding-real-time-gps-accelerometer-data-complete-platform/> (visitado el 7/10/2021).
- [10] Michelin, “Michelin® mems® 4,” 2021, <https://business.michelinman.com/mines-quarries/services-mining-quarries/michelin-mems-4> (visitado el 7/10/2021).
- [11] Quadrianto, N. y Buntine, W. L., cap. Linear Regression, pp. 747–750. Boston, MA: Springer US, 2nd ed., 2017, doi:10.1007/978-1-4899-7687-1\_481.
- [12] James, G., Witten, D., Hastie, T., y Tibshirani, R., cap. Tree-Based Methods, pp. 327–365. New York, NY: Springer US, 2nd ed., 2021, doi:10.1007/978-1-0716-1418-1\_8.
- [13] Schapire, R., Explaining AdaBoost, pp. 37–52. 2013, doi:10.1007/978-3-642-41136-6\_5.

- [14] Friedman, J., “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, pp. 367–378, 2002, doi:10.1016/S0167-9473(01)00065-2.
- [15] Chen, T. y Guestrin, C., “Xgboost: A scalable tree boosting system,” pp. 785–794, 2016, doi:10.1145/2939672.2939785.
- [16] Sammut, C. y Webb, G. I., eds., *Mean Squared Error*, pp. 808–808. Boston, MA: Springer US, 2017, doi:10.1007/978-1-4899-7687-1\_528.
- [17] Scikit-Learn, “Parameter estimation using grid search with cross-validation,” 2021, [https://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_grid\\_search\\_digits.html](https://scikit-learn.org/stable/auto_examples/model_selection/plot_grid_search_digits.html) (visitado el 21/10/2021).
- [18] Mastery, M. L., “How to calculate feature importance with python,” 2021, <https://machinelearningmastery.com/calculate-feature-importance-with-python/> (visitado el 21/10/2021).
- [19] Lundberg, S. y Lee, S.-I., “A unified approach to interpreting model predictions,” 2017.
- [20] Science, T. D., “Explain your model with the shap values,” 2019, <https://towardsdatascience.com/explain-your-model-with-the-shap-values-bc36aac4de3d> (visitado el 29/10/2021).
- [21] Anzabi, R., Nobes, D., y Lipsett, M., “Haul truck tire dynamics due to tire condition,” *Journal of Physics: Conference Series*, vol. 364, 2012, doi:10.1088/1742-6596/364/1/012005.
- [22] Rodovalho, E., Quaglio, O., Felsch, W., Pascual, R., Tomi, G., y Tenório, J., “Reducing ghg emissions through efficient tire consumption in open pit mines,” *Journal of Cleaner Production*, vol. 255, p. 120185, 2020, doi:10.1016/j.jclepro.2020.120185.
- [23] Kowshik, C. S., Naik, N., Bhat, R., y Silvam, N., “Predicting the effect of operating parameters on the radial tire treadwear using response surface method,” *Journal of Engineering Science and Technology*, vol. 14, pp. 2824–2835, 2019.
- [24] Morales Navarro, J., “Aumento de vida útil de los neumáticos para la flota de transporte de minera centinelas,” 2020, <http://repositorio.uchile.cl/handle/2250/177975>.
- [25] Barrientos, V., “Disminución de los costos por hora en neumáticos de camiones mineros,” 2019, <https://www.mch.cl/reportajes/disminucion-de-los-costos-por-hora-en-neumaticos-de-camiones-mineros/>.

# Anexo A

## Tasa de Desgaste en Base a Horas Entre Inspecciones

Durante este trabajo de título se dispuso de dos principales variables para medir el avance del uso de neumáticos en CAEX: odometría y horas entre mediciones, siendo ambas variables útiles para graficar y medir la tasa a la cual se desgastan los neumáticos. Sin embargo, es claro que la variable de odometría resulta más representativa que las horas entre mediciones, dado que las horas entre mediciones no entregan información respecto del uso real del camión, razón por la cual durante este trabajo se usó la odometría para calcular la tasa de desgaste de los neumáticos. A modo de experimento, de todas formas, se invirtieron estas variables, obteniendo resultados para MSE, gráficos de seguimiento de desgaste real, e importancia de las variables. Las variables usadas para entrenar los algoritmos predictivos se mantuvieron idénticas, salvo por la inclusión de la variable *dif\_odom* previamente usada para calcular la tasa de desgaste y el retiro de la variable *dif\_hrs\_med*.

De esta forma, la nueva tasa de desgaste usada en este experimento se calcula como:

$$tasa\_desgaste = \frac{\Delta y_j}{\Delta Horas} \quad (A.1)$$

### A.1. MSE

En términos de MSE, los resultados son de esperar, dada la menor información aportada por la variable de horas entre mediciones. Estos resultados se muestran en la tabla A.1.



Tabla A.1: Valor de MSE de modelos predictivos según posición en el camión y tipo de algoritmo usado para la arquitectura de 2 etapas, usando la variable  $\Delta Horas$  para el cálculo de la tasa de desgaste.

| <b>Posición</b> | <b>XGBoost</b> | <b>Adaboost</b> | <b>Regresión Lineal</b> |
|-----------------|----------------|-----------------|-------------------------|
| Delantero       | 0.653          | 0.727           | 0.662                   |
| Trasero externo | 4.149          | 5.058           | 7.510                   |
| Trasero interno | 6.359          | 6.778           | 7.886                   |
| General         | 3.703          | 4.174           | 5.350                   |
| Seguimiento     | 39.437         | 46.149          | 64.776                  |

## A.2. Seguimiento de Desgaste Real

De igual forma al método usado en la sección de Resultados de este documento, en primer lugar se grafican los 9 neumáticos más problemáticos usando esta configuración (mayor error de seguimiento de desgaste real), lo que se presenta en la figura A.1.

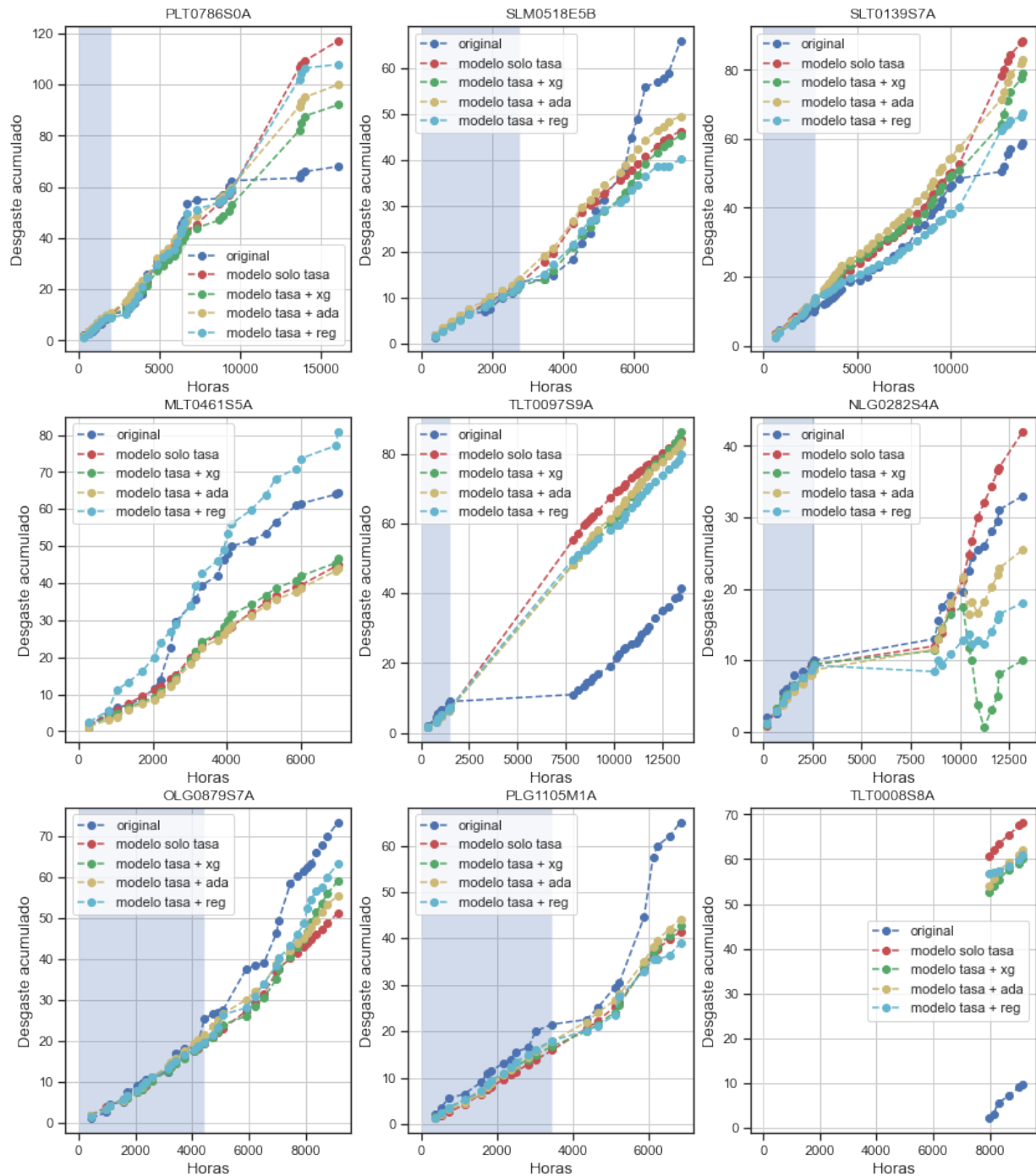


Figura A.1: Curvas de desgaste predicho (usando por un lado sólo tasa mensual y por otro agregando los ajustes de algoritmos predictivos) y real, para los 9 neumáticos más problemáticos de la flota. El área azul representa el periodo que el neumático estuvo en posiciones delanteras.

Se distinguen anomalías significativas en parte de los neumáticos, que pueden verse asociadas con camiones detenidos (horas avanzan pero el camión se mantiene sin uso), además de otros fenómenos que impulsan a una predicción de desgaste negativo, situación que puede deberse a la dispersión de la data en términos de horas. Esta dispersión se puede observar claramente en el histograma presente en la figura A.2.

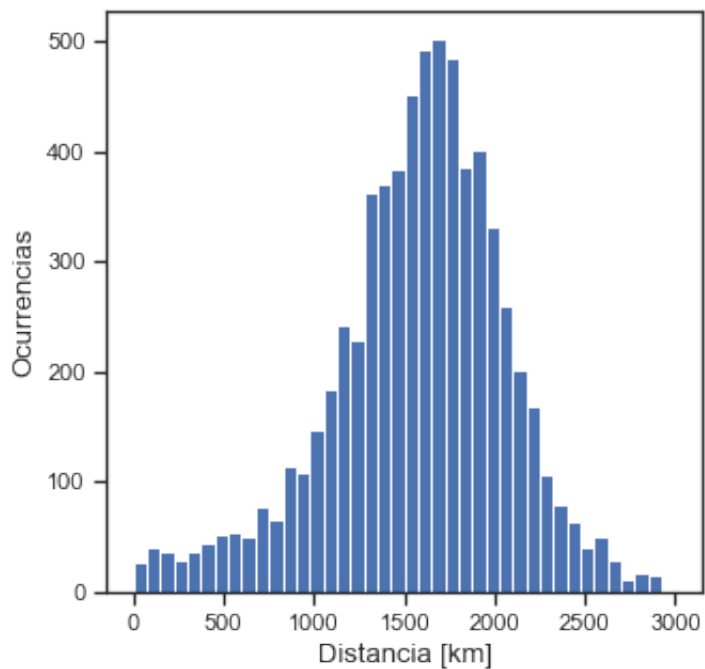


Figura A.2: Histograma de kilometraje alcanzable para un paso fijo de 250 horas entre mediciones, usando todas las tasas de desgaste disponibles en el *dataset* ( $250 \frac{\Delta Odom}{\Delta Horas}$  para todos los neumáticos).

Por otro lado, escogiendo neumáticos al azar, se obtienen los resultados presentes en la figura A.3.

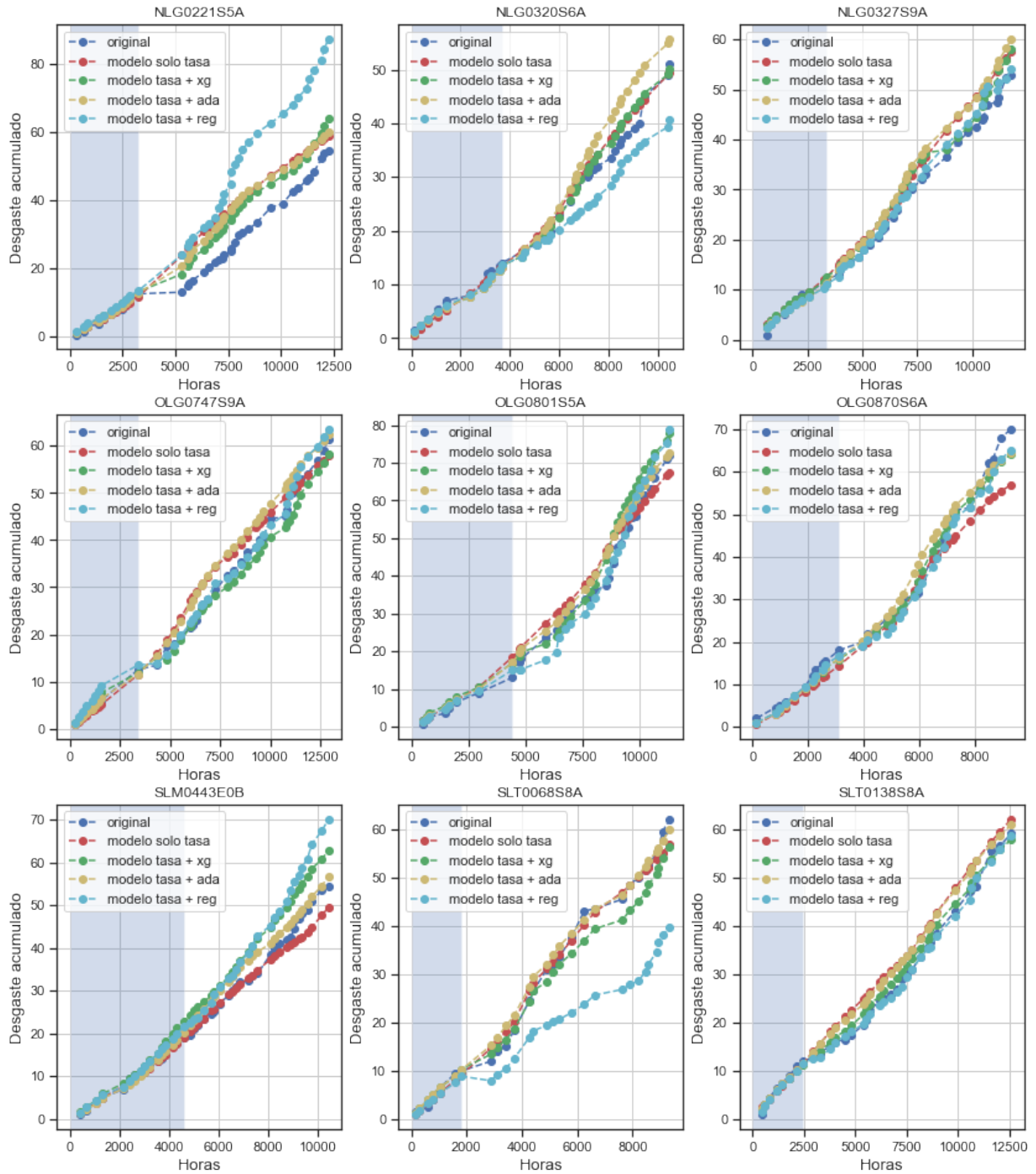


Figura A.3: Curvas de desgaste predicho (usando por un lado sólo tasa mensual y por otro agregando los ajustes de algoritmos predictivos) y real, para 9 neumáticos de la flota escogidos al azar. El área azul representa el periodo que el neumático estuvo en posiciones delanteras.

Como se observa en la figura A.3, el seguimiento de desgaste real no es particularmente deplorable, al tomar como referencia el algoritmo *XGBoost* usado anteriormente. Sin embargo, es claro que este modelo presenta un mayor error que el usado finalmente en la sección *Resultados* de este documento y presenta anomalías difíciles de eliminar.

### A.3. Importancia de Variables

En este caso no se observan cambios significativos y se mantienen las relaciones previamente establecidas, donde las variables más importantes son aquellas que combinan información de presión y temperatura con la variable de odometría.

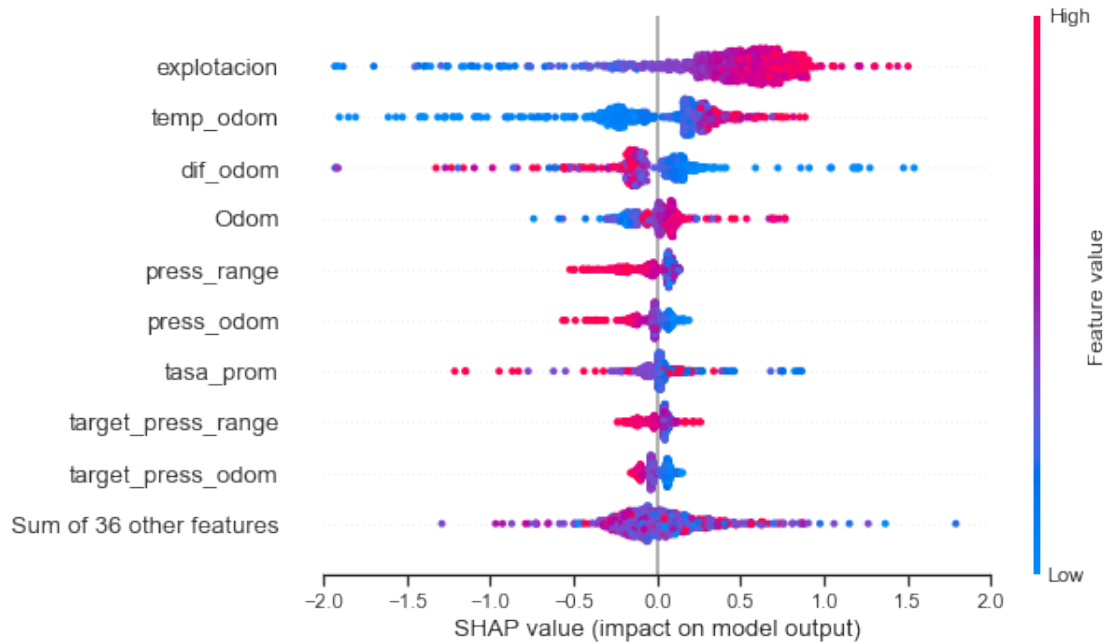


Figura A.4: *SHAP Values* para el experimento de tasa de desgaste en base a horas en neumáticos traseros externos.

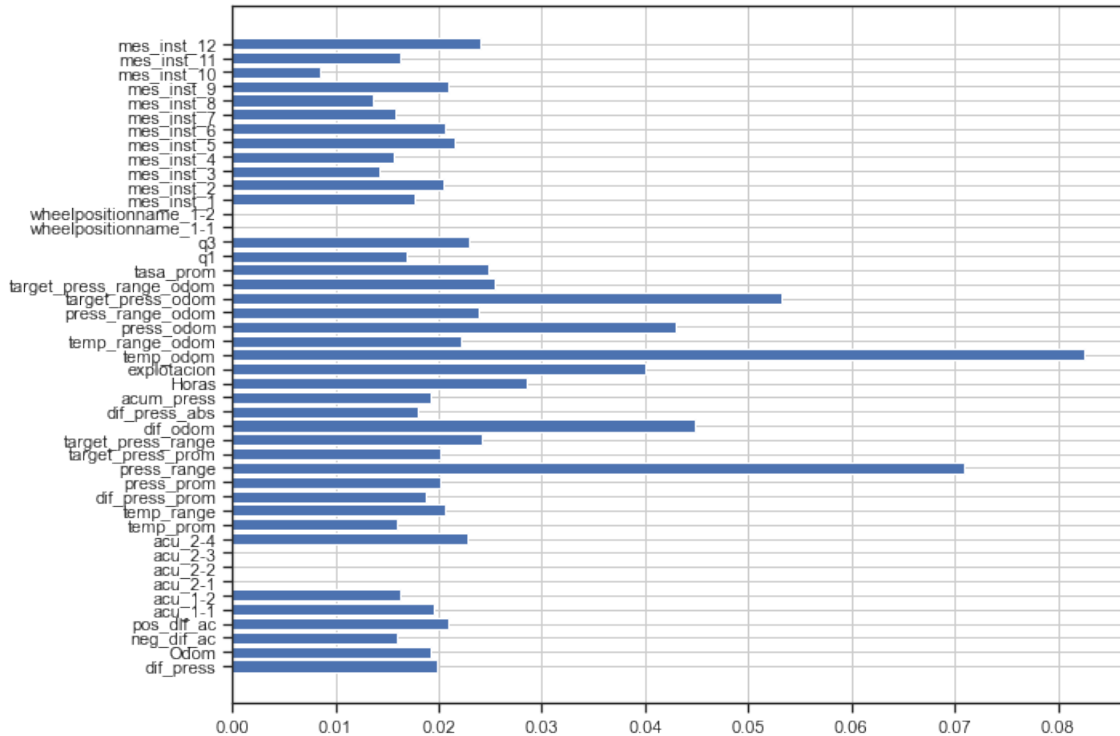


Figura A.5: *Feature importances* para el experimento de tasa de desgaste en base a horas en neumáticos delanteros.

Como se observa en las figuras A.4 y A.5, la variable *explotacion* presenta mayor importancia en los modelos, lo que es de esperar, ya que presenta información de kilometraje dividido por horas entre mediciones, aportando información relevante para los modelos que se basan en horas entre inspecciones. Que esta variable adquiera tal importancia es otro signo claro de que los modelos pueden maximizar su exactitud mediante la variable de kilometraje y que usar esta variable como base para las tasas de desgaste resulta en un mejor modelo. Fuera de esta particularidad, se mantiene la importancia de las variables que incluyen información de odometría en su cálculo.

## A.4. Conclusiones

Los resultados aquí obtenidos reafirman que la metodología usada en el documento principal es un mejor acercamiento a la resolución del problema, al usar la variable de odometría para calcular la tasa de desgaste de los neumáticos. Usando un acercamiento de tasa de desgaste basada en horas entre inspecciones resulta difícil para los algoritmos predictivos determinar el ajuste correcto, debido a la gran dispersión de la data y la posibilidad de que la variable avance con el camión detenido. Debido a esto, en el experimento realizado existe una gran desconexión entre el proceso real (desgaste del neumático debido al uso) y el algoritmo que busca modelarlo, posicionándolo como un útil marco de referencia para entender el desempeño del algoritmo entrenado en el documento principal, pero no como una propuesta de solución al problema.