



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE  
ESCUELA DE INGENIERÍA  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS  
ICS2122 – TALLER DE INVESTIGACIÓN OPERATIVA  
PROFESORES: GONZALO PÉREZ, MARCELO PÉREZ, ARIEL ESPINOZA, MATÍAS DE GEYTER,  
ELBIO AVANZINI, ALEJANDRO MAC CAWLEY  
AYUDANTE JEFE: FRANCISCO TAMARÍN  
SEGUNDO SEMESTRE 2023

## Informe Final

---

### Mantenimiento predictivo mediante datos categóricos de una línea de embotellado

---

#### Grupo 5

TOMÁS ARANCIBIA  
CONSUELO HONORATO  
FLORENCIA LIRA  
ANTONIA ORREGO  
RAFAELA RÁMIREZ  
MACARENA TAGLE  
GASPAR VILLARROEL

# Índice

<b>1 Resumen Ejecutivo</b>	<b>4</b>
<b>2 Introducción</b>	<b>4</b>
<b>3 Descripción del problema</b>	<b>5</b>
3.1 Antecedentes . . . . .	5
3.2 Definiciones . . . . .	6
3.3 Exploración de datos . . . . .	6
<b>4 Metas e Indicadores clave</b>	<b>7</b>
<b>5 Discusión Metodológica</b>	<b>8</b>
5.1 Método de Redes Neuronales Recurrentes (RNN) . . . . .	8
5.2 Comparación de metodologías . . . . .	8
<b>6 Análisis de datos</b>	<b>9</b>
6.1 Frecuencia . . . . .	9
6.2 Avisos y Alarms . . . . .	10
6.3 Producción . . . . .	12
6.4 Status . . . . .	13
<b>7 Metodología</b>	<b>13</b>
7.1 Estudio de los datos . . . . .	14
7.1.1 Definición de Falla . . . . .	14
7.1.2 Relaciones extraídas . . . . .	15
7.1.3 Caso base analítico . . . . .	15
7.1.4 Correlaciones . . . . .	16
7.1.5 Clusterización . . . . .	18
7.2 Modelo de predicción . . . . .	21
7.2.1 Predicción Mediante RNN . . . . .	21
7.2.2 Desarrollo de modelo base . . . . .	21
7.2.3 Resultado del modelo base . . . . .	22
<b>8 Desarrollo del modelo</b>	<b>23</b>
8.1 Análisis casos extremos en la predicción . . . . .	24
8.2 Predicción de probabilidad de falla mediante RNN . . . . .	25
8.2.1 Predicciones . . . . .	25
<b>9 Análisis de Sensibilidad</b>	<b>26</b>
9.1 Batch size y Epoch . . . . .	26
9.2 Test Size . . . . .	29
<b>10 Fast Fourier Transform</b>	<b>30</b>
<b>11 Evaluación de KPI'S</b>	<b>30</b>
<b>12 Propuesta de política</b>	<b>30</b>

<b>13 Conclusiones</b>	<b>31</b>
<b>14 Referencias Bibliográficas</b>	<b>32</b>
<b>15 Anexos</b>	<b>33</b>
15.1 Carta Gantt . . . . .	33
15.2 Anexo 1 . . . . .	33
15.3 Anexo 2 . . . . .	34
15.4 Anexo 3 . . . . .	35
15.5 Anexo 4 . . . . .	36
15.6 Anexo 5 Gráficos frecuencia status . . . . .	39
15.7 Anexo 6 Correlación de series de tiempo para las fallas . . . . .	47
15.7.1 Fallas y Series Cluster 1 . . . . .	47
15.7.2 Fallas y Series Cluster 2 . . . . .	47
15.7.3 Fallas y Series Cluster 3 . . . . .	49
15.8 Anexo 7 Listado Clusters . . . . .	53
15.9 Anexo 8 Gráficas Clusters . . . . .	56
15.10 Anexo 9 Matrices de Correlación por cluster . . . . .	61
15.11 Anexo 10 Mantenciones por Cluster . . . . .	63
15.12 Anexo 11 Predicción y pérdida para cada falla . . . . .	64
15.13 Anexo 12 Análisis FFT . . . . .	94

## 1 Resumen Ejecutivo

En el presente informe se desarrolla un análisis detallado sobre el trabajo de investigación realizado con la línea de embotellamiento de vinos de una empresa vitivinícola. El objetivo es actualizar las políticas de mantenimientos ya existentes, pasando de uno preventivo/correctivo hacia uno predictivo, el cual busca mejorar la eficiencia operativa y la gestión del mantenimiento de la máquina automatizada.

El proyecto está sustentado en datos de tipo categóricos, los cuales están presentes en cinco bases de datos distintas: avisos, alarmas, mantenciones, status y producción. Estas contienen información de distintas ocurrencias en un periodo de 20 meses, juntando entre ellas más de 5 millones de observaciones.

Gracias a la investigación de las bases de datos en conjunto, se pudo analizar el comportamiento de las ocurrencias de la maquinaria con respecto a las mantenciones realizadas, y se pudo observar que las mantenciones no se realizaban una vez que ésta fallaba, por lo que estas no solucionaban totalmente los errores emergentes. De aquí surge la importancia de resolver este problema, ya que fallas en la linea implican pérdidas de materia prima, tal como lo son las botellas y el vino.

A partir de esto, se establecen metas a cumplir e indicadores que permitan medir el desempeño, para poder testear la eficiencia de la metodología de predicción.

Para esto, se analizan y discuten distintas metodologías encontradas en la literatura, incluyendo las redes neuronales recurrentes, métodos de suavización exponencial, método Holt-Winter, método de Croston y SBA y Fast Fourier Transform. Se investigó sobre fortalezas y debilidades de cada metodología y se realizó un análisis comparativo para determinar que metodología se adapta mejor a la resolución de la problemática, poniendo énfasis en la complejidad y tamaño de los datos, parámetros a estimar y variables endógenas y exógenas. A partir de este análisis se optó por trabajar con las Redes neuronales Recurrentes (RNN) para la predicción de las series de tiempo.

Se llevó a cabo análisis de correlación y autocorrelación para comprender el comportamiento de las ocurrencias en conjunto y analizar efectos causales, como también, clusterización para agrupar eventos relacionados.

Posteriormente se pudo desarrollar un modelo de predicción de fallas en la línea de embotellado mediante las RNN para el caso base, describiendo los pasos de testeо, desarrollo y entrenamiento, llegando a grandes resultados de la precisión y pérdida durante el entrenamiento del modelo.

Se entrenó un modelo de predicción de comportamiento de fallas, a partir de cada uno de los clusters a los cuales estaban designadas. Con esto, se analizaron los casos extremos que revelaron grandes desafíos y limitaciones de la aplicación del modelo, observando el coeficiente de determinación y error absoluto medio para saber el nivel de éxito de la predicción, según esto se concluyó la importancia de ajustar el tamaño del lote y las épocas en función de la cantidad de datos de entrenamiento, lo cual se pudo ahondar a profundidad en el análisis de sensibilidad.

Luego se profundizó en el desarrollo del modelo de predicción y se decidió utilizar un modelo de dos etapas, en primer lugar de una predicción mediante RNN de los avisos y alarmas que condicionan el comportamiento de la falla, esto a través de los datos del clusters. Esto permite extender los datos en 5 días y así conocer el comportamiento de los avisos y alarmas en los 5 días futuros de interés. El resultado de predicción se almacena y se utiliza como input para una segunda RNN que tiene por objetivo determinar la probabilidad de falla en los próximos 5 días, esto lo hace a través del estudio de los patrones de los avisos y las alarmas, identificando las condiciones de estos que originan la falla, este dato es el esperado para entregar la política de mantenimiento a la empresa.

Finalmente, se propuso la implementación de política para mejorar la calidad e integridad en la medición de datos, con el fin de diferenciar claramente entre avisos y alarmas y poder definir de manera temprana una posible falla, catalogándolos de manera precisa, como también pasos futuros para el seguimiento del problema.

## 2 Introducción

En el entorno de los procesos productivos empresariales, la optimización y agilización de la producción son pilares fundamentales para satisfacer la demanda y alcanzar una máxima rentabilidad. Sin embargo, estos objetivos pueden verse amenazados por las incidencias en los recursos operacionales, específicamente en la maquinaria responsable de los procesos de la empresa, lo que impulsa la necesidad de una aproximación mas exacta a los momentos en que se requiere de un mantenimiento, y así disminuir los periodos de inoperancia.

Los sistemas de embotellado, con su maquinaria especializada y ritmo constante, están sometidos a un desgaste que puede llevar a fallas costosas y tiempos de inactividad no planificados, que pueden significar

la pérdida de materia prima y el no cumplimiento de la demanda. Si se cuenta con un sistema de mantenimiento correctivo o preventivo, como es el caso de la empresa en cuestión, no se está abordando de manera óptima las fallas imprevistas y los recursos, lo que puede resultar en tiempos de inactividad, gastos innecesarios y falta de capacidad para prever y evitar problemas críticos de manera anticipada. En la actualidad, numerosas organizaciones están optando por estrategias de mantenimiento predictivo por una serie de razones fundamentadas. El enfoque de mantenimiento predictivo se basa en la utilización de datos históricos y técnicas de análisis para prever posibles fallos en la máquinaria antes de que ocurran. Además de prolongar la vida útil de los equipos, es una herramienta esencial para prevenir las interrupciones no planificadas, reducir costos operativos y garantizar un óptimo funcionamiento que conlleve a una mayor seguridad en los procesos productivos.

Ahora se cuenta con el desafío de desarrollar una metodología de mantenimiento predictivo para la línea de embotellamiento de una empresa vitivinícola. Para lograrlo, se cuenta con bases de datos que contienen un gran volumen de información registrada en los últimos diez meses. Esta tarea implica la aplicación de algoritmos y modelos avanzados de análisis de datos para identificar patrones, tendencias y señales tempranas de problemas o anomalías. A medida que se avanza en la investigación, se cree que el enfoque en el mantenimiento predictivo ofrecerá beneficios significativos para la empresa y contribuirá al éxito continuo de la línea de embotellamiento en este mercado altamente competitivo. En las secciones siguientes de este informe, se detallan las etapas que se llevaron a cabo, con el objetivo de proporcionar una visión completa del trabajo realizado y sus implicaciones para el proceso de embotellamiento.

### 3 Descripción del problema

#### 3.1 Antecedentes

Una línea de embotellamiento de vinos consta de 4 etapas; enjuague, lavado y secado, llenado y tapas, siguiendo un proceso de Monoblock en una máquina automatizada, con una serie de estaciones con eventos y operaciones de manera secuencial. Cada evento presente en un proceso de embotellamiento en empresas vitivinícolas, es de suma importancia para lograr un procedimiento eficiente y eficaz. Es por esto que es necesario analizar cada una de estas y cómo afectan a la producción cuando se generan distintos tipos de fallas.

Actualmente la empresa solamente lleva a cabo mantenimientos preventivos, los cuales se realizan una vez que la máquina alcanza una cierta cantidad de horas en uso o cuando se requiere de una mantención general. Este tipo de mantenimientos no permiten aprovechar el 100% de la capacidad de la máquina al detener su funcionamiento sin tener completa claridad de las fallas. Sin embargo, una vez que la máquina presenta fallas que requieren de atención inmediata para seguir con su funcionamiento, esta es intervenida, para lo cual se detiene totalmente y se soluciona la falencia.

La detección y gestión de defectos es crucial en el ámbito del correcto funcionamiento de la línea productiva, debido a las enormes pérdidas económicas y de recursos que pueden generar, incluso si algunos de esos defectos resultan ser falsas alarmas. Existe un claro trade-off entre las tres metodologías de mantenimiento mencionadas con anterioridad. Por un lado, el mantenimiento correctivo implica abordar las fallas una vez que se detectan, lo que puede ser costoso y peligroso si la detección es tardía. Por otro lado, el mantenimiento preventivo se basa en programar revisiones periódicas de componentes en función de su edad, lo que puede ser efectivo en casos donde la probabilidad de falla está relacionada con la edad del equipo, pero puede llevar a un mantenimiento repetitivo e innecesario en otros casos.

Para superar estas limitaciones y anticipar defectos mientras se evitan mantenimientos innecesarios, se introduce el mantenimiento predictivo. Esta metodología se basa en datos en tiempo real y datos históricos para estudiar y comprender las causas de los defectos. El objetivo es identificar las condiciones que generan defectos y, si los resultados son prometedores, se busca encontrar el equilibrio entre la detección temprana de defectos y la minimización de mantenimientos innecesarios en la máquinaria (Lopez, 2021).

La empresa vitivinícola en cuestión cuenta con una gran recopilación de datos ordenados, tanto categóricos como continuos, que ha utilizado para optimizar su operación en la línea de embotellamiento. Para esto, es necesario también hacer cuenta de los mantenimientos que se deben realizar, y minimizar su ocurrencia, ya que estos significan detener la producción. Al observar los datos entregados, los mantenimientos que se realizan no cuentan con ningún tipo de periodicidad definida, por lo que es necesario definir un mantenimiento predictivo. El cambio de política propuesto busca priorizar este tipo de mantenimiento por sobre las otras como estrategia principal para la gestión de la línea de embotellado basándose en una premisa proactiva para anticiparse a posibles fallas con el fin de reducir los tiempos de inactividad no planificada.

### 3.2 Definiciones

En esta sección, se presentan definiciones fundamentales para el buen entendimiento del estudio, las cuales se hará referencia a ellas durante todo el informe.

- Intervenciones: Esta acción hace alusión a aquella actividad destinada a corregir o prevenir. Son aplicadas a la maquinaria para garantizar un funcionamiento óptimo no necesariamente debido a una falla. Implican la detención de la máquina durante una breve cantidad de tiempo (Status: STOP).
- Mantenimiento correctivo: Actividad que implica una acción reactiva frente al status del equipo después de que ha presentado un problema, se realizan reparaciones o ajustes para restablecer el correcto funcionamiento.
- Mantenimiento preventivo: Actividad planificada realizada de manera regular (horas de funcionamiento de la máquina) o irregular para corregir comportamientos hasta el minuto y/o reducir el riesgo de avería en la maquinaria con el fin de evitar problemas potenciales.
- Mantenimiento predictivo: Actividad que conlleva un monitorización continua y análisis de datos en tiempo real para identificar patrones y tendencias anómalas que indiquen la necesidad de un mantenimiento, permitiendo una intervención proactiva.
- Epoch: Representa una pasada completa a través de todo el conjunto de datos de entrenamiento durante el proceso de ajuste de los parámetros del modelo. Cuando el número de épocas es más pequeño, el modelo puede subajustar, lo que significa que no aprenderá completamente los patrones en los datos. Si el número de épocas es más grande, el modelo tiene más oportunidades de aprender, pero también existe el riesgo de sobreajustar, donde el modelo se adapta en exceso a los datos de entrenamiento y no generaliza bien a nuevos datos. Generalmente, se necesita un equilibrio, y es común utilizar técnicas como la validación cruzada o detener el entrenamiento temprano si la pérdida en el conjunto de validación deja de mejorar.
- Batch size: Cantidad de ejemplos de entrenamiento que se utilizan en cada iteración o paso de entrenamiento del modelo. En otras palabras, es el número de muestras de datos que se procesan en paralelo antes de actualizar los parámetros del modelo. En un proceso de entrenamiento, los datos de entrenamiento se dividen en varios lotes, y el modelo se ajusta después de procesar cada uno de estos lotes. En la práctica, es común experimentar con diferentes tamaños para determinar cuál funciona mejor para un problema específico. Los valores típicos suelen estar en el rango de 32, 64, 128 o 256, pero pueden variar según el caso.

### 3.3 Exploración de datos

Para el análisis de datos se dispuso de 5 bases de datos. Algunos elementos comunes que poseen estas bases de datos es el registro temporal el cual al hacerlas coincidir definen el marco temporal de estudio desde el 06/09/2022 hasta el 14/06/2023, es importante mencionar que se está trabajando con una máquina en específico de la marca “Bertolaso”, por lo que los datos son codificados por ella y se encuentran relacionados con esta. Los únicos datos que no son obtenidos directamente de la máquina son las fechas de las mantenciones, las cuales fueron proporcionadas directamente por la empresa.

A continuación se detallan cada una de las bases de datos:

**Alarms:** Esta base de datos corresponde a información entregada por la máquina que muestra la ocurrencia de estos eventos denominados alarmas en varios instantes de tiempo, este tipo de evento se caracteriza por mostrar un cambio en el comportamiento de la máquina que debe ser considerado según su gravedad, ya que se registran desde ocurrencias de rutina hasta fallas que significan su mal funcionamiento. Esta posee una longitud de 14,952 observaciones, de las cuales se logró extraer un total de 81 alarmas distintas, las cuales se encuentran detalladas en el Anexo (1). En particular esta base posee una columna denominada causa la cual clasifica las alarmas en 6 tipos : “No Asignado”, “Generic Cause”, “Internal Cause”, “External Cause”, “Operator Action”, “Outfeed Accumulation”

**Avisos:** Esta base de datos sigue el mismo comportamiento que la base de alarmas, es un registro en el tiempo de la ocurrencia de eventos que se catalogan como avisos. Su longitud es de 199,287 observaciones, debido a lo que se distinguieron 49 tipos de avisos distintos que se encuentran detallados en el Anexo (2), debido a las descripciones y la frecuencia con que ocurren demuestran que la gravedad de los avisos es menor a de

las alarmas.

**Mantenciones:** Contiene un registro de las mantenciones realizadas por la empresa, con la especificación de la duración y fecha realizada. Se cuentan con 13 instancias de mantención, de las cuales solo 7 son efectivas para el análisis ya que pertenecen al marco temporal. De las fechas se extrae que estas no se realizan de forma periodica, ocurriendo en intervalos de tiempos variables entre una y otra.

**Producción:** Con un total de 3,676,125 observaciones, esta base de datos muestra los ritmos de producción registrados por la máquina a lo largo del tiempo.

Tanto la base de datos de alarmas y avisos seguían un patrón de registro escrito horizontalmente con las ocurrencias de eventos en un intervalo de tiempo específico. Con el propósito de individualizar las ocurrencias se genero una nueva base de datos con un registro vertical, incluyendo las siguientes variables:

- **UNIX\_T:** El “tiempo Unix” o “marca de tiempo Unix” se refiere a una forma particular de representar el tiempo en sistemas operativos tipo Unix, como Linux. En esta representación, el tiempo se mide en segundos transcurridos desde un punto de referencia específico conocido como el “epoch” o “época” de Unix. El “epoch” de Unix es el 1 de enero de 1970 a las 00:00:00 UTC (Tiempo Universal Coordinado).
- **Tipo:** Esta variable fue creada al juntar las dos bases de datos para indicar si la ocurrencia de la fila corresponde a una Alarma o Aviso.
- **Descripción:** Indica los tipos de avisos o alarmas que existen.
- **CAUSA:** Es una variable de los datos de alarmas que indica el origen de la alarma, de estas pueden existir 6 tipos mencionados anteriormente.
- **FECHAHORA:** Entrega el año, mes, día, hora, minuto y segundo exacto en que se guardo la información.

La disposición vertical nos permitió juntar en una sola base de datos la información tanto de avisos, alarmas, mantenciones y producción, la cual se ordeno en base al tiempo para tener un visión temporal.

**Status:** Esta base de datos, contiene un registro del funcionamiento de la máquina, es decir, su contenido de 71,503 observaciones muestra los distintos status operacionales de la empresa. En particular esta base para demostrar el funcionamiento de la máquina, cuenta con 21 descripciones del cambio en el funcionamiento de la máquina los cuales se encuentran definidos en el Anexo (3). Además, la base cuanta con una columna llamada MACRO, en la que se registra de que forma cambia el status de la máquina, entra las posibles macros encontramos: “STOP”, “RUN”, y “PAUSA”. La cuales hacen referencia a, que la máquina no se encuentra funcionando, la máquina esta en funcionamiento, y que la producción de la máquina esta en una pequeña pausa respectivamente.

## 4 Metas e Indicadores clave

A partir de los datos entregados, es necesario identificar aquellos avisos y alarmas que generan un mayor impacto en la productividad de la máquina, es decir, aquellas situaciones en las que la productividad de esta baja considerablemente. Para esto, se identificarán correlaciones entre los distintos tipos de alerta y avisos con las fechas de mantenciones efectivas que se tienen. Esto tiene el propósito de generar una política de mantención, con la cual se puedan predecir los comportamientos futuros y evitar una baja en la producción debido a problemas operacionales.

Para poder ir testeando el desempeño del modelo, definimos una serie de indicadores (KPI's) con el fin de detectar anomalías, continuidades y/o alertas tempranas. Estos nos permitirán ir comparando los resultados que entreguen las distintas metodologías y así poder escoger aquella que se ajuste de mejor manera a nuestros datos.

- (a) Tasa de avisos y alarmas diarias.
- (b) Tiempo de actividad en producción (estacionalidades).

(c) Diferencia de Capacidad Real vs. Capacidad Propuesta.

A partir de estos KPIs se podrá comprobar la efectividad operacional de la solución entregada, se espera que la tasa de avisos y alarmas disminuya en el largo plazo, y los otros indicadores aumenten. A la medida que se desarrolle la predicción existirán otros tipos de KPIs que ganan protagonismo como son el porcentaje de acierto, porcentaje de error y diferencia entre el valor predicho y el real, estos valores miden la efectividad de la predicción y permiten conocer el estado actual del modelo frente a la realidad, permitiendo a través de esta evaluación de rendimiento tomar decisiones para mejorar los comportamiento actuales o iterar en la metodología.

## 5 Discusión Metodológica

En la sección de Discusión Metodológica, se ha llevado a cabo un análisis exhaustivo de la literatura, centrándose en metodologías aplicadas para abordar problemas con las mismas características al presente. A continuación se define brevemente el método de RNN el cuál se implementará en este caso. Para obtener mayor información sobre esta metodología y las de suavización exponencial y Fast Fourier Transform se puede revisar el Anexo (4). Se han comparado las fortalezas y debilidades de estas metodologías y se ha destacado la necesidad de identificar objetivos y enfoques específicos para adaptar la metodología al contexto de nuestro problema.

### 5.1 Método de Redes Neuronales Recurrentes (RNN)

Una red neuronal recurrente (RNN) es un tipo de modelo de redes neuronales diseñado para trabajar con datos secuenciales o series temporales en sistemas dinámicos no lineales. Estas redes utilizan datos de entrenamiento junto con técnicas algorítmicas de gradiente descendente para aprender patrones, ya que utilizan información previa para generar respuestas. En otras palabras, las RNN tienen elementos de entrada y salida que están conectados a través de conexiones recurrentes. (Bonet, I., Et. Al., 2007)

Cuando se aplica este enfoque a un problema como el de una línea de embotellado, se vuelve especialmente relevante debido a su capacidad para lidiar con la no linealidad inherente al proceso de producción, su adaptabilidad para ajustarse a cambios en la operación y su capacidad para tolerar fallas sin interrupciones significativas. Esto se justifica aún más cuando se dispone de datos históricos secuenciales, como registros de avisos y alarmas a lo largo del tiempo, y un conjunto de datos relativamente grande recopilado durante un período de 20 meses permitiendo así estudiar profundamente la dimensionalidad del tiempo.

Como antecedente que respalda y otorga valor a la utilización de esta metodología en el mantenimiento predictivo, se han encontrado estudios y tópicos que abordan la misma problemática en otros ámbitos, como por ejemplo, en la industria de la aviación. A partir del análisis lograron llegar a la siguiente conclusión, “Nuestros mejores modelos han sido capaces de detectar entre un 45-55% de los casos anteriormente mencionados correctamente. Esta cifra debe contrastarse con el hecho de que no podemos asegurar que todos los fallos dentro de nuestro conjunto de datos tengan un origen relacionado con los factores climatológicos estudiados. Por tanto, se ha cumplido el objetivo: demostrar el potencial de predicción de estos datos aplicando Deep Learning.” (López, J.M., 2021). Esto valida justamente que para el estudio de mantenimientos predictivos logra ser una herramienta de gran utilidad para abordarlo por Deep Learning mediante RNN.

### 5.2 Comparación de metodologías

Dado que el objetivo es crear un modelo predictivo de fallas en una línea embotelladora, la elección de la metodología depende de la complejidad de los datos y de la precisión requerida.

Si los datos de fallas son altamente secuenciales y presentan patrones temporales complejos, las Redes Neuronales Recurrentes (RNN) son la elección más adecuada debido a su capacidad para capturar patrones no lineales y adaptarse a cambios en la operación.

Si los datos son relativamente estables y simples, y se requieren alertas tempranas, los Métodos de Suavización Exponencial (SES o Holt-Winters) pueden ser útiles.

El Fast Fourier Transform (FFT) es valioso para analizar patrones de frecuencia, pero no es un método predictivo en sí mismo. Puede utilizarse en combinación con otras metodologías, como las RNN, para comprender mejor los patrones cíclicos en los datos.

De esta forma, las Redes Neuronales Recurrentes (RNN) son la elección más prometedora para este problema debido a su capacidad para lidiar con la no linealidad inherente, adaptarse a cambios y manejar datos secuenciales. Sin embargo, es importante evaluar el costo computacional y considerar las limitaciones mencionadas.

## 6 Análisis de datos

En el proceso de análisis de datos, además de realizar una observación general de las bases de datos, se consideró esencial la representación gráfica para visualizar la relación entre diversos parámetros seleccionados. Esto permitió obtener una comprensión más clara y dinámica de cómo se comportan los datos a lo largo del tiempo, y facilitó la identificación de patrones y tendencias.

## 6.1 Frecuencia

Se confeccionaron gráficos de frecuencia que representaran las ocurrencias totales de las distintas alarmas y avisos. Estos otorgan una visión general de los datos en todo el período de tiempo de análisis. Este enfoque permitió identificar cuáles de estos eventos eran más recurrentes a lo largo de la serie temporal. Dado que el objetivo principal implica la creación de un modelo predictivo, es necesario contar con una amplia cantidad de datos para comprender el comportamiento de la máquina en cuestión. Por esto, se inició el análisis con un enfoque en los eventos más recurrentes en nuestra base de datos, ya que son estos los que aportan una mayor cantidad de información.

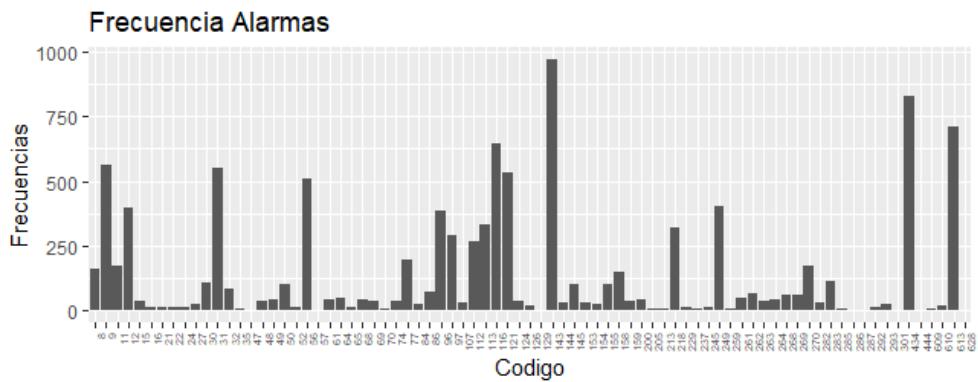


Figure 1: Frecuencias de cada alarma.

Del gráfico se distingue claramente que existen 3 tipos de alarmas las cuales poseen una mayor frecuencia estas son las siguientes: “Falta cera en botella”, “Avería sonda presión sondra” y “Volante no en reposo”, con las frecuencias correspondientes a 971, 829 y 713 ocurrencias respectivamente. Además, se observa que existe una gran cantidad de alarmas que su ocurrencia esta muy por debajo de las 250 ocurrencias, aún más específico 18 tipos de alarma no poseen más de 10 ocurrencias en la totalidad del periodo de análisis.

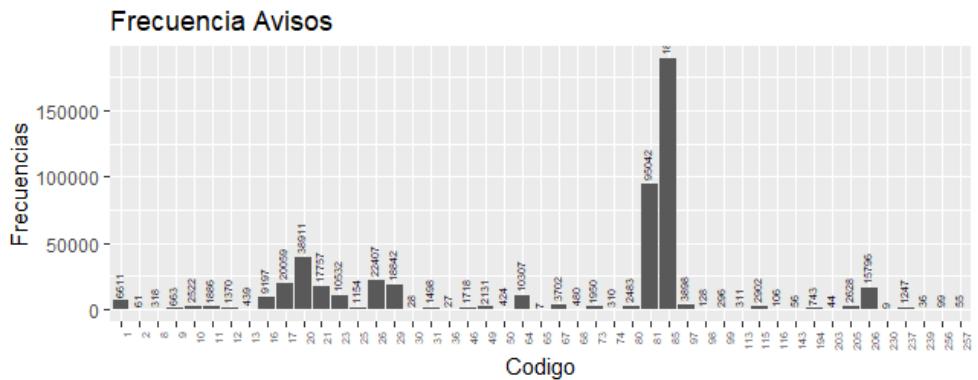


Figure 2: Frecuencias de cada aviso.

A primera vista se puede ver como las frecuencias de los avisos se disparan en comparación a la ocurrencia de las alarmas alcanzando valores por sobre los 1000 en más del 50% de los tipos de alarmas, específicamente en 28 tipos de 49. Los 3 tipos mas recurrentes de avisos son Control automático ventilador campana, Modulación entrada producto equivocada y Señal bloqueo botellas activada en linea con una frecuencia de 189181, 95042 y 38911 respectivamente.

Como se observó de los gráficos anteriores, la frecuencia de los avisos es superior a la de las alarmas en gran medida. Con el objetivo de visualizar la relación entre la cantidad de alarmas y avisos totales en el tiempo se realizó el siguiente gráfico:

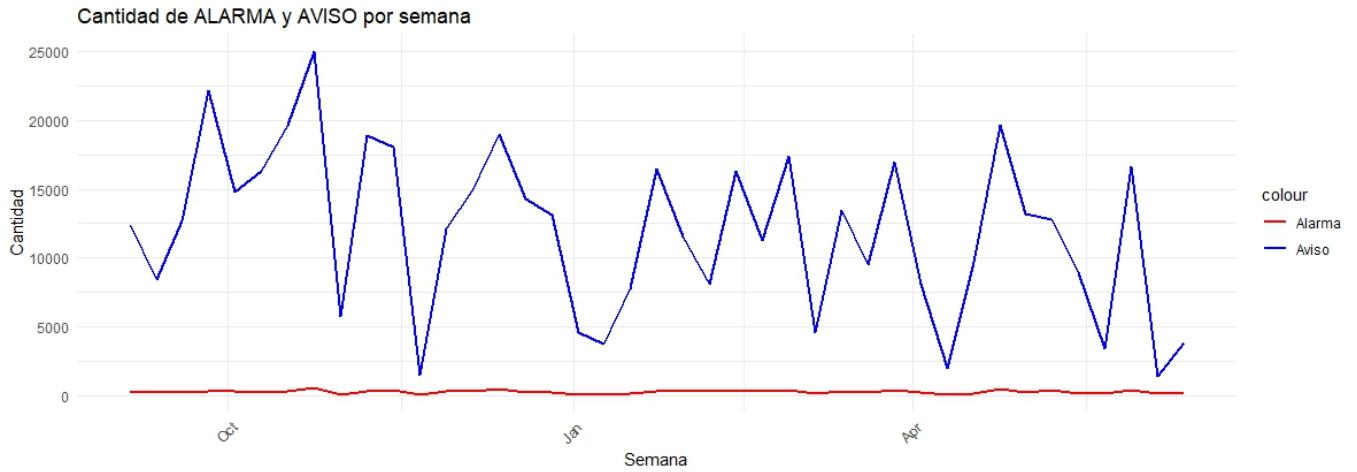


Figure 3: Cantidad de avisos y alarmas por semana.

Este gráfico demuestra que en todo el periodo de tiempo la cantidad de avisos es superior a la de alarmas. Además, por el nivel de detalle alcanzado para los avisos se ven grandes fluctuaciones en períodos del tiempo identificados que es interesante analizar.

Del análisis de frecuencia quedan instancias de análisis abiertas, tal como ver la relación de las alzas y descensos de cantidad de avisos en un instante de tiempo con la realización de una mantención. Por otro lado, si bien es importante ver cuales son los eventos más frecuentes se deben estudiar a su vez, aquellos avisos y alarmas con baja recurrencia debido a que pueden exitir situaciones de alta gravedad que ocurrán una menor cantidad de veces y que alteren inmediatamente el proceso productivo.

## 6.2 Avisos y Alarmas

Teniendo en cuenta cuales son los eventos con mayores recurrencias, tanto para las alarmas como para los avisos, se verificó el comportamiento de los 10 eventos con mayor frecuencia en ambas categorías a lo largo de la serie de tiempo, para identificar como estos se relacionan con las fechas de las mantenciones, y evaluar si aquel evento afecta de alguna manera la producción en la linea de embotellamiento.

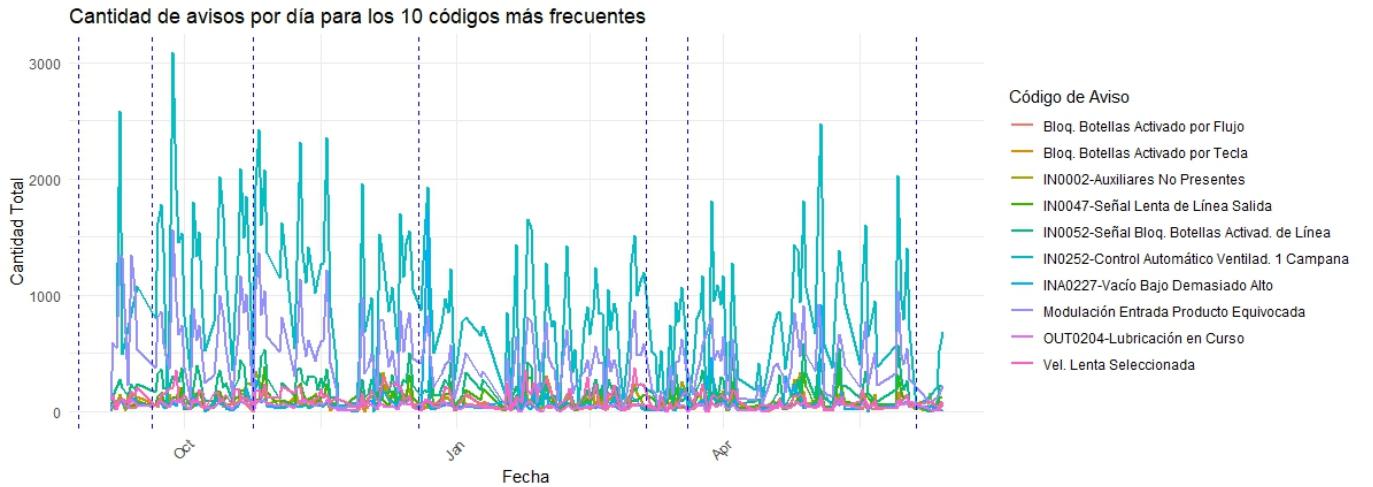


Figure 4: cantidad de avisos por día para los 10 mas frecuentes y mantenciones.

Se puede observar en los gráficos anteriores diversos comportamientos. Por una parte hay eventos que no presentan ningún tipo de correlación con la cercanía al momento de una mantención (lineas punteadas verticales) como son los avisos “Cambio de velocidad en la máquina” o de “Bloqueo automático”, esto puede ocurrir debido a que son avisos que no perjudican el funcionamiento de la máquina. Por otro lado se observa que hay avisos cuya frecuencia tiene un grado de correlación con la cercanía de las mantenciones, por ejemplo el aviso “Control automático ventilador”, vemos que su frecuencia va al alza hasta la mantención y posteriormente desciende, este fenómeno da indicios de que ese aviso es un factor de alerta de una falla.

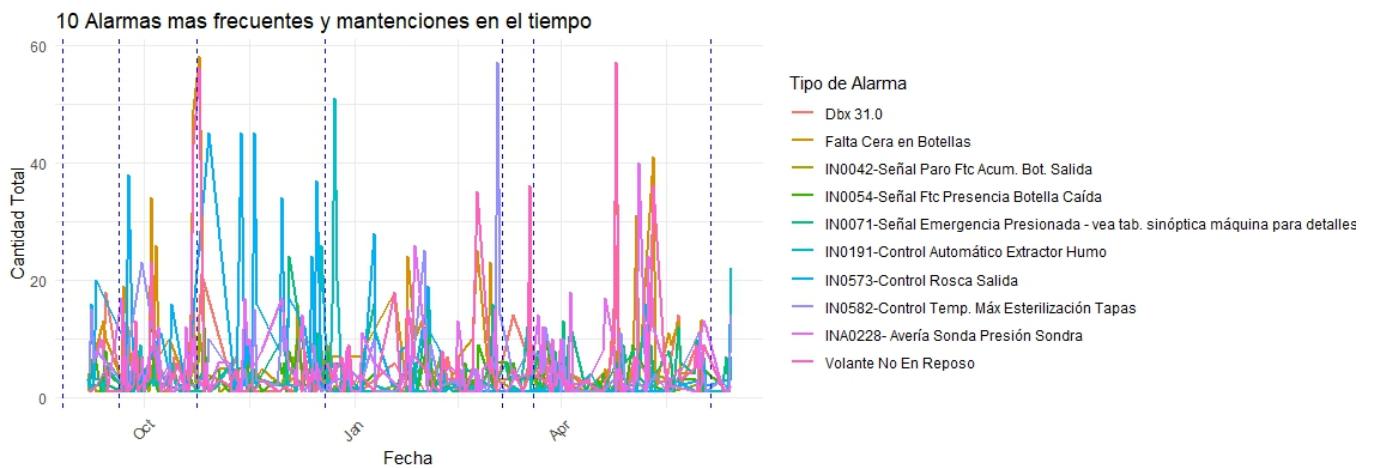


Figure 5: cantidad de alarmas por día para las 10 mas frecuentes y mantenciones.

En el gráfico anterior, se logran identificar diferentes comportamientos con respecto a las fechas de las mantenciones. Hay ciertas alarmas que su frecuencia si presenta una correlación con las mantenciones (aumenta previo a la fecha y posteriormente disminuye). Por otro lado, hay grupos de alarmas cuya frecuencia es independiente de la presencia de una mantención. El comportamiento de una alarma específica, previo a una mantención, es de gran utilidad porque permite verificar si se realizó mantenimiento preventivo o una intervención correctiva, lo que permite clasificar el tipo de alarma como un posible factor que puede producir una falla, en caso de que haya ocurrido una intervención.

Junto con lo anterior, se graficó la cantidad de alarmas totales por día, para corroborar si el comportamiento particular de las alarmas anteriores con las mantenciones se asemeja al comportamiento general de las alarmas y estas fechas particulares, las cuales están demarcadas por lineas verticales en los gráficos.

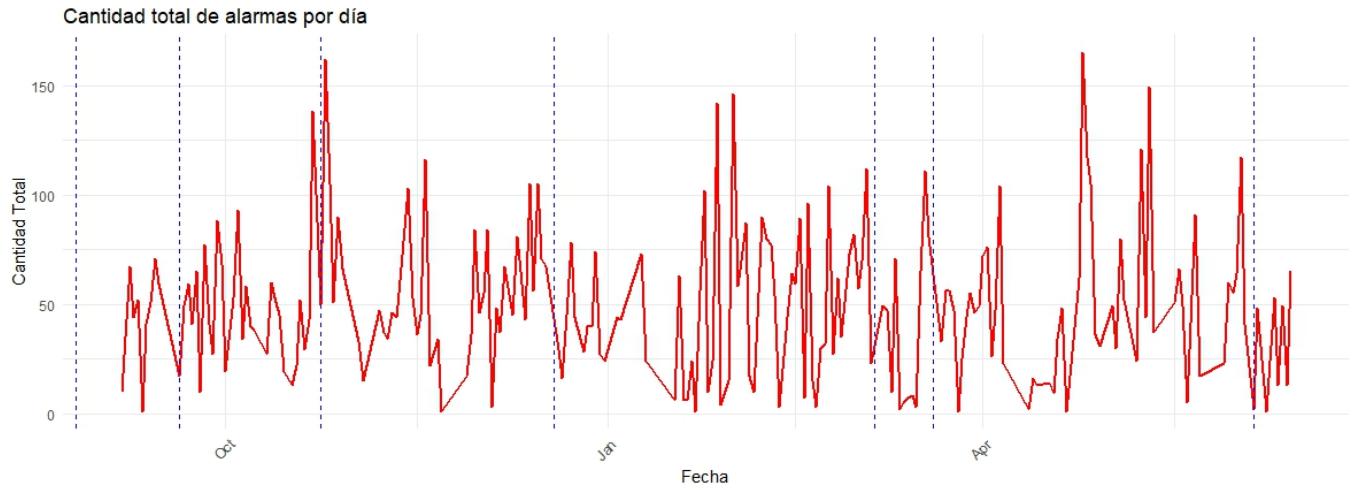


Figure 6: Cantidad de alarmas por día y mantenciones realizadas.

Si bien en algunas mantenciones se logra identificar el comportamiento esperado que refleja un alza de las frecuencias justo antes de la fecha y un descenso posterior a esta, lo cual permite inferir que la mantención solucionó las fallas; también se observa la existencia de un desfase entre el alza y la baja con las mantenciones, por lo que se ha optado por trabajar con el comportamiento a nivel individual para no perder información relevante.

### 6.3 Producción

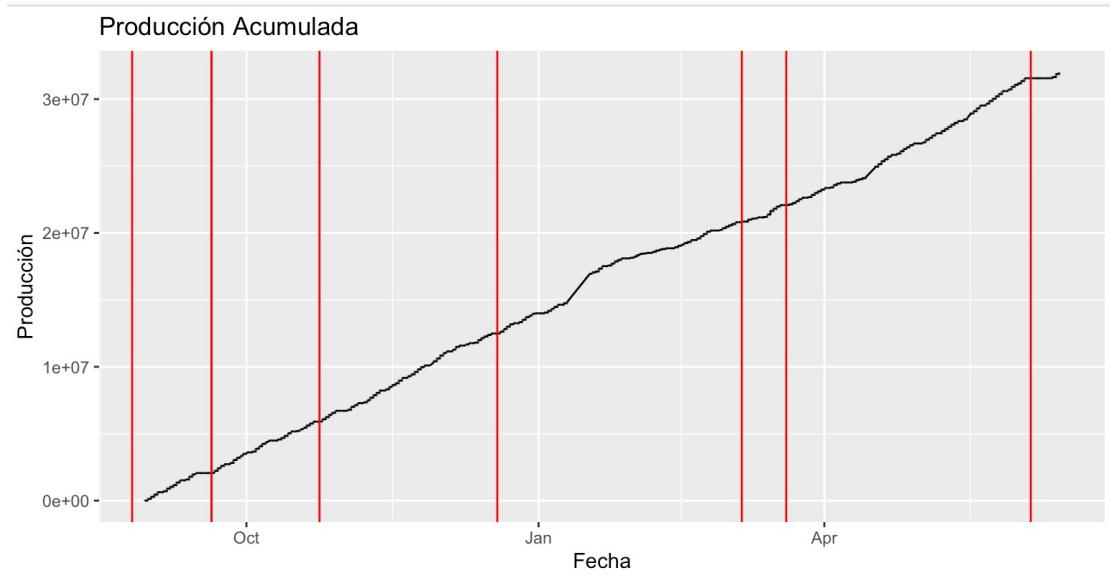


Figure 7: Producción de botellas de vino acumuladas y mantenciones.

El gráfico muestra la cantidad de botellas acumuladas en todo el periodo de tiempo. En su generalidad el crecimiento es estable, pero se logran identificar pendientes que dejan en evidencia la presencia de inestabilidades interesantes para el análisis.

Se puede ver que en la parte central de la gráfica, entre la cuarta y quinta mantención, hay una zona con una pendiente muy positiva, lo que significa un gran volumen de producción en una menor cantidad de tiempo. Dada la gran cantidad de valores, se decidió trabajar con la información diaria para poder simplificar el análisis.

## 6.4 Status

En el análisis de los status surgieron una serie de hipótesis. En primer lugar, que la frecuencia de STOPS o PAUSAS disminuya posterior a una mantención. Para confirmar o refutar la hipótesis, se tomó en cuenta solo las paradas que fueron producidas por algún mal funcionamiento fuera de lo operacional, para esto se calificaron las descripciones de status de la máquina, en operaciones rutinarias y no rutinarias tal y como se puede ver en el Anexo (3). Además, esta base de datos se utilizó para encontrar las posibles fallas en las bases de datos de alarmas y avisos, identificando aquellas ocurrencias que producen detención de la máquina. Luego de la clasificación mencionada, se graficó la relación entre la frecuencia de estas paradas con respecto al tiempo, para comprobar la hipótesis. Tomando como ejemplo el siguiente gráfico:

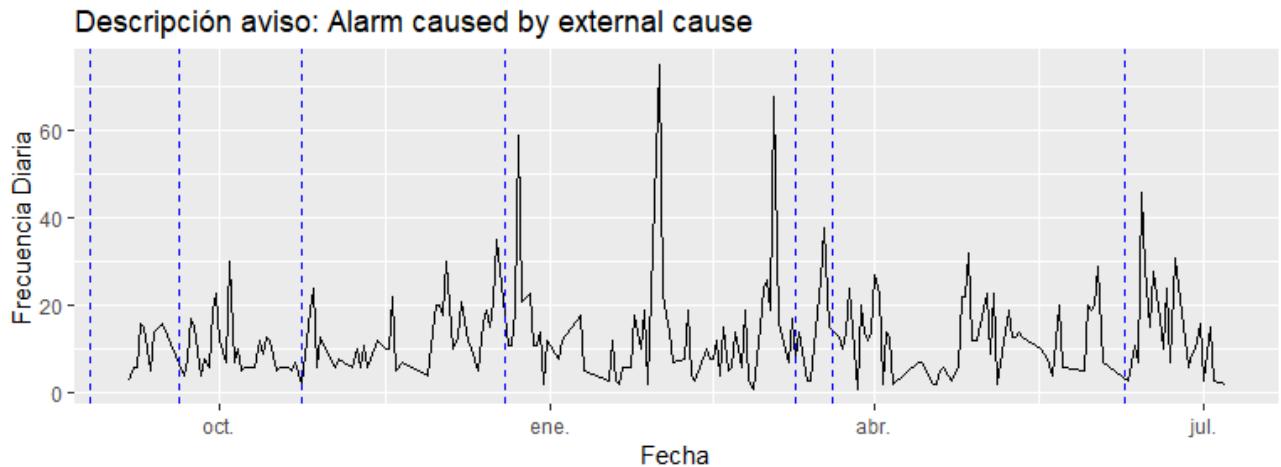


Figure 8: Frecuencia STOP por causa externa

Se puede evidenciar que para este tipo de STOP, ocurre que luego de una mantención, la frecuencia de las paradas de ésta por causas externas tiende a disminuir, cabe mencionar que este comportamiento no es común para todas las paradas que se definieron como “NO” rutinarias, como se muestra en los gráficos del Anexo (5), es por esto que la hipótesis planteada no puede ser validada.

## 7 Metodología

La recopilación de información junto al análisis de datos que se ha desarrollado, nos permite definir una línea de acción o metodología para la búsqueda de una solución a la problemática. El diseño y desarrollo para la creación de un modelo predictivo para la linea de embotellado se dividirá en dos etapas fundamentales: estudio de los datos y generación de un modelo de predicción.

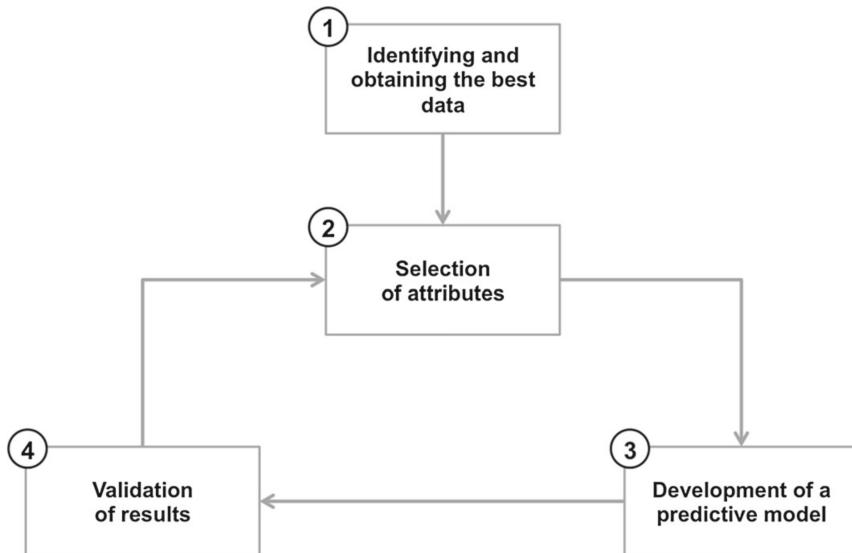


Figure 9: Etapas del proceso de construcción de un modelo predictivo. Fuente: ResearchGate

## 7.1 Estudio de los datos

El estudio se basa en el entendimiento y manejo de la información proveniente de las distintas bases de datos. El objetivo principal que tiene esta etapa es entender las relaciones existentes en la data para posteriormente trabajarlas mediante algoritmos o métodos que nos permitan obtener proyecciones a futuro a partir de los datos históricos.

### 7.1.1 Definición de Falla

Para poder definir cuales son aquellas situaciones graves para la máquina es necesario definir claramente qué es una falla.

Como ya se mencionó anteriormente, la máquina funciona como monoblock, es decir, en ella se llevan a cabo todas las etapas por las que pasa la botella de vino, desde que está vacía hasta que está lista para ser empaquetada y distribuida. De esta forma, una falla es todo aquel evento que genera pérdidas en la producción, que detiene este proceso y requieren de supervisión o atención de algún funcionario de la fabrica.

A partir de la base de datos de Status, se ha logrado identificar aquellos avisos y alarmas que corresponden a fallas según la definición propuesta, es decir, aquellas que generan STOP en la máquina. Las fallas obtenidas se encuentran listadas en la siguiente tabla. La frecuencia o cantidad de ocurrencias de estos eventos no determina su gravedad, si no mas bien, el tiempo de detención que estas fallas generan en la maquina. Estas han sido filtradas según distintos criterios para obtener las potenciales fallas más importantes de tener en cuenta. La tabla está dividida en dos partes para que su entendimiento sea más fácil, donde el tiempo está representado en minutos.

Falla	T. prom. detención	Frecuencia	T. total detención
OUT0210-Válvula Entrada {...}C.I.P.	12,6146	178	1425,45
Error Acoplam. Enjuagadora	14,18951	55	766,2333
Falta Comunicación PLC Depósito	6,94245	120	812,2667
IN0054-Señal Ftc Presencia Botella Caída	10,76413	296	2970,9
INA0228- Avería Sonda Presión Sondra	3,2831940	493	1.566,083
INA0227-Vacio Bajo Demasiado Alto	4,0166	1993	1000,133
IN0031-Grasa Bloqueada	5,000463	192	180,0167
INA0200-Nivel Máximo Depósito	14,95	46	74,75
Puertas Desbloq.	5,864583	201	93,83333
IN0086-Control Sobretemperatura Armario	8,363636	43	184
INA0227-Vacio Bajo Demasiado Bajo	3,681081	349	136,2
IN0577-Control Botella Colgada	2,75	18	49,5
Error Acoplamiento Tapadora 2	10	2	19,35
IN0064-Señal Mesa Acum. Bot. Llena	2,5	16	17,5
Botella Rota TV	3,166667	1	3,166667
OUT0207-Presión Cilindros Demasiado Baja	0,233333	1	0,233333
Máquina Desfasada	0	6	0
INA0000-Falta Ref. Vel. de Inversor Motor ppal	0	0	0

Table 1: Datos de Tiempo de Falla (Parte 1)

Falla	Min T.	Max T.	Frec. siendo falla
OUT0210-Válvula Entrada {...} C.I.P.	0,1666667	185,5167	113
Error Acoplam. Enjuagadora	0,1666667	168,8667	54
Falta Comunicación PLC Depósito	0,0333333	145,35	117
IN0054-Señal Ftc Presencia Botella Caída	0,16666667	129,6167	276
INA0228- Avería Sonda Presión Sondra	0,1	121,75	477
INA0227-Vacio Bajo Demasiado Alto	0,0666667	107,55	249
IN0031-Grasa Bloqueada	0,15	72,03333	36
INA0200-Nivel Máximo Depósito	0,333333	70,1	5
Puertas Desbloq.	0,133333	68,3333	16
IN0086-Control Sobretemperatura Armario	0,1666667	40,66667	22
INA0227-Vacio Bajo Demasiado Bajo	0,0666667	24,68333	37
IN0577-Control Botella Colgada	0,1666667	18,83333	18
Error Acoplamiento Tapadora 2	5,35	14	2
IN0064-Señal Mesa Acum. Bot. Llena	0,1666667	10,68333	7
Botella Rota TV	3,166667	3,166667	1
OUT0207-Presión Cilindros Demasiado Baja	0,233333	0,233333	6
Máquina Desfasada	0	0	0
INA0000-Falta Ref. Vel. de Inversor Motor ppal	0	0	0

Table 2: Datos de Tiempo de Falla (Parte 2)

### 7.1.2 Relaciones extraídas

En la sección 6 de “Análisis de datos”, se estudió gráficamente la relación existente entre avisos, alarmas, status y producción con las mantenciones. Lo anterior, en conjunto con la definición de falla facilitará la búsqueda de aquellos avisos o alarmas que generan las detenciones.

Con esto se busca definir la probabilidad de ocurrencia de una falla para evitar que estas ocurran sin previo aviso y que detengan la producción por un plazo más extenso de tiempo. Todo esto con el propósito de generar mantenimientos predictivos.

### 7.1.3 Caso base analítico

A partir del análisis gráfico de los datos y la lectura de las descripciones de cada una de las alarmas, avisos y de los Status de la maquina, se pudo identificar dos eventos que permiten entender claramente la

identificación de fallas dentro de los avisos y alarmas. Esto ocurre con los avisos “Máquina ligeramente desfasada” y “Máquina desfasada”.

Se logró mostrar que la importancia de estos avisos radica en que la máquina cuenta con una cinta principal que transporta las botellas y las traslada desde una etapa a la otra. Una descoordinación de la cinta con otras partes, como lo son los inyectores de vino o la zona de puesta de tapas, puede generar pérdidas de vino o rupturas de botellas. De esta manera se identificó que el aviso “Máquina ligeramente desfasada” anticipa la falla “Máquina desfasada”.

Se logró evidenciar que la frecuencia de estos avisos a lo largo de todo el rango temporal es bastante baja, teniendo un total de 7 ocurrencias para el caso de “Máquina ligeramente desfasada” y 6 para “Máquina desfasada”.

#### 7.1.4 Correlaciones

En base al objetivo central del proyecto, el cual es predecir la ocurrencia de fallas en la linea de embotellamiento, se requiere conocer la relación entre las variables involucradas. Esto permite estudiar el comportamiento de los distintos eventos que ocurren en forma conjunta, dando la oportunidad de analizar causalidad entre unos y otros. Para esto se realizaron 2 tipos de análisis, los cuales se resumen en estudio de autocorrelaciones y estudio de correlaciones entre variables.

##### (a) Autocorrelación

Por la conformación de los datos y el funcionamiento de la máquina se sabe que en un día normal se emite una gran cantidad de avisos y alarmas de los cuales debido a la gravedad generan distintos comportamientos como pausar la máquina (PAUSA), poner en marcha la máquina (RUN) y detener la máquina por un periodo de tiempo más largo (STOP). Estos 3 status mencionados son indicadores de problemas en el funcionamiento, es decir, de estar frente a una falla. Por lo cual es de interés analizar su comportamiento en el tiempo.

Básicamente la definición de autocorrelación “trataría de explicar la relación que existe en la memoria de la serie observada a través del tiempo” (Gutiérrez, 2021), en este caso se hace uso de la función `acf` de RStudio. Esta entrega un gráfico, en donde en el eje x se tiene el LAG que en base a los datos entregados, que son series temporales, consiste en el desplazamiento de la serie temporal en días. Además, se encuentra en el eje y el coeficiente de ACF o de autocorrelación. Para cada uno de los status mencionados se obtuvieron los siguientes gráficos:

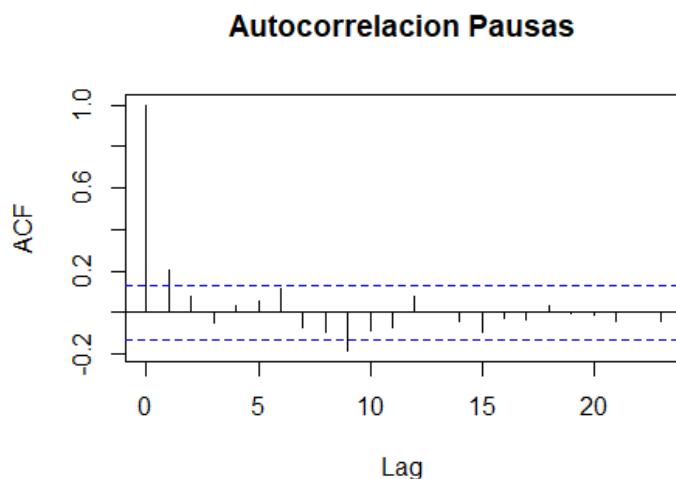


Figure 10: Autocorrelación para el estado de la máquina PAUSA

Este gráfico muestra como la ocurrencia de pausas en un periodo de tiempo afecta las pausas en otro periodo de tiempo. Se observa que entre los 0 y 5 lags los coeficientes de correlación se mantienen dentro del rango de precisión definida por las lineas horizontales, más específicamente en el lag 3 indicando que el comportamiento de las pausas se repite en promedio cada 3 días.

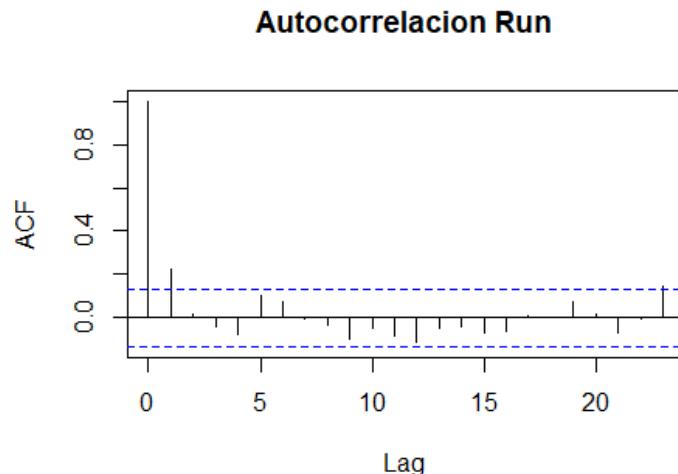


Figure 11: Autocorrelación para el estado de la máquina RUN

Para las puestas en marcha es esperable un comportamiento similar al de las pausas, debido a que cada vez que ocurre un evento de pausa se intenta volver lo más rápido posible a la producción normal. El gráfico apoya esto anterior, mostrando que en el lag 3 se estabiliza la función dentro de los límites de precisión, es decir, el patrón de Run se repite cada 3 días.

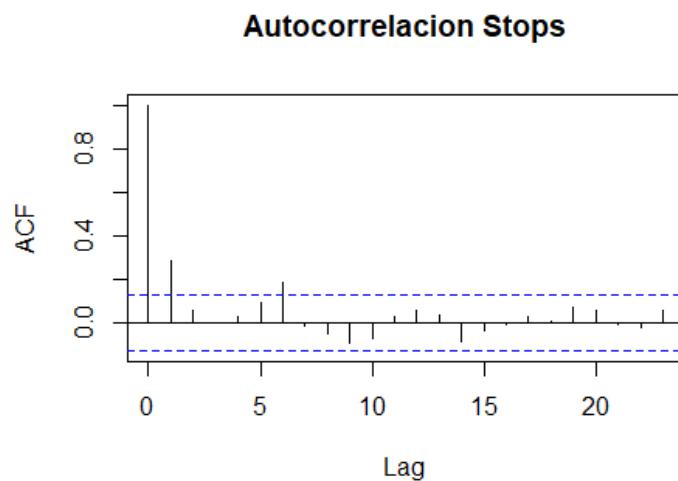


Figure 12: Autocorrelación para el estado de la máquina STOP

En el gráfico se observa una estabilidad de la serie de tiempo distinta a la presentada en los gráficos anteriores. En este caso ocurre entre los lags 5 y 10, más específicamente en el lag 7, es decir, cada 7 días se repite el comportamiento de los Stops.

Los 3 gráficos poseen elementos en común: son estabilidades menores a los 10 días y con una alta correlación entre los primeros 2 días, lo cual es natural debido a la ocurrencia de un evento. A pesar de esto, no se logran identificar ciclos claramente definidos que alienten a buscar una predicción a partir de la periodicidad de estos fenómenos, debido a que existen respuestas operativas que justifican los períodos identificados como es el trabajo de turnos semanales para los Stops de lag 7.

#### (b) Matriz de correlación

En esta parte del estudio se implementaron 3 matrices distintas de correlaciones: la primera relaciona los avisos, la segunda las alarmas y la tercera los avisos con las alarmas. Este método fue seleccionado para conocer las relaciones entre variables, debido a la gran cantidad de involucradas. En específico se deseaba conocer qué tipos de avisos y alarmas mostraban dependencia, ya que esto aporta información clave para la anticipación y predicción de una falla.

De las 3 matrices se calcularon medidas estadísticas básicas, que se muestran a continuación en la siguiente tabla.

	Avisos - Avisos	Avisos - Alarmas	Alarmas - Alarmas
Media	0.135	0.05	0.06
Mediana	0.059	0.011	-0.0047
SD	0.22	0.126	0.1

Table 3: Medidas estadísticas de las matrices de correlación

Los resultados mostrados en la tabla permiten observar la composición general de cada una de las matrices, si bien se esperaban valores superiores a 0.5 para asegurar una correlación fuerte los valores obtenidos en su mayoría se encuentran entre 0,0 y 0,2. Esto gatilló que se analizara de forma particular cada una de las relaciones en busca de resultados significativos.

Con esto se pudo visualizar de manera gráfica el significado de una correlación en series de tiempo, como se puede observar a continuación con la falla “OUT0210-Válvula Entrada {...} C.I.P.” y sus series de tiempo con mayor correlación que corresponden al aviso “Señal Falta Solicitud lista de CIP” y la alarma “Mantenimiento Programado”. Se pueden observar las distintas correlaciones para las fallas de forma gráfica en el Anexo (6).

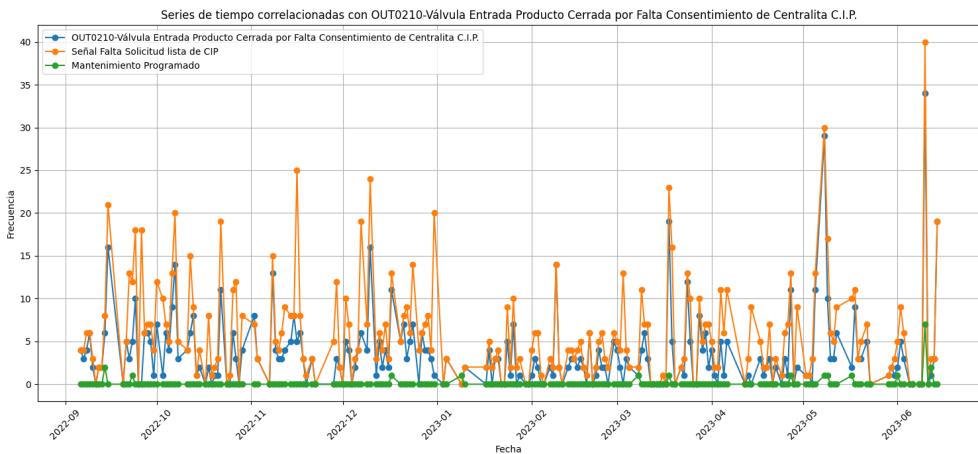


Figure 13: Correlaciones entre series de tiempo: OUT0210-Válvula Entrada {...} C.I.P.

A pesar de la obtención de resultados útiles, estos se obtuvieron en su mayoría en base a un trabajo manual el cual es difícil debido al volumen de datos. Esto motivó la búsqueda de otras herramientas como son los clusters que se detallan más adelante, los cuales permiten hacer la agrupación de estos eventos en forma computacional y no manual, de igual forma tener los valores en las matrices permite solucionar dudas puntuales sobre los datos.

#### 7.1.5 Clusterización

Con el fin de extraer mayor información a partir del análisis de datos y las correlaciones, se decidió realizar una clusterización de los datos con el propósito de conseguir subconjuntos de la data y reconocer patrones similares entre los objetos, para así, obtener más precisión en la definición de fallas que se mencionó anteriormente. Para este proceso, se aplicó un algoritmo particional, en el cual se eligió una cantidad de clusters relativo a lo que se requería en este estudio, en este caso 4, una vez corrido el algoritmo, se van reasignando iterativamente los objetos entre los grupos hacia la convergencia, asignando todos los grupos a la vez. Además, el método heurístico utilizado dentro del algoritmo es el de “k-means”.

Para el caso de la línea de embotellado, el método de k-means calcula el centroide o valor promedio de cada aviso/alarma, para luego definir un umbral de distancia a los centroides que indique cuando el punto está demasiado lejos del centroide y así, poder agrupar cada variable en grupos con patrones similares. Todo esto, con el fin de observar anomalías y señales de fallas de forma gráfica.

Se realizó el proceso de clusterización para avisos, alarmas y combinado entre avisos y alarmas. A partir de eso, se llegó a grandes conclusiones sobre la caracterización de cada cluster y en qué se diferencian unos de otros, las cuales se presentarán a partir de gráficos.

La lista completa de cada cluster se encuentran en el Anexo (7) y los clusters de avisos y alarmas por separado se encuentran en el Anexo (8).

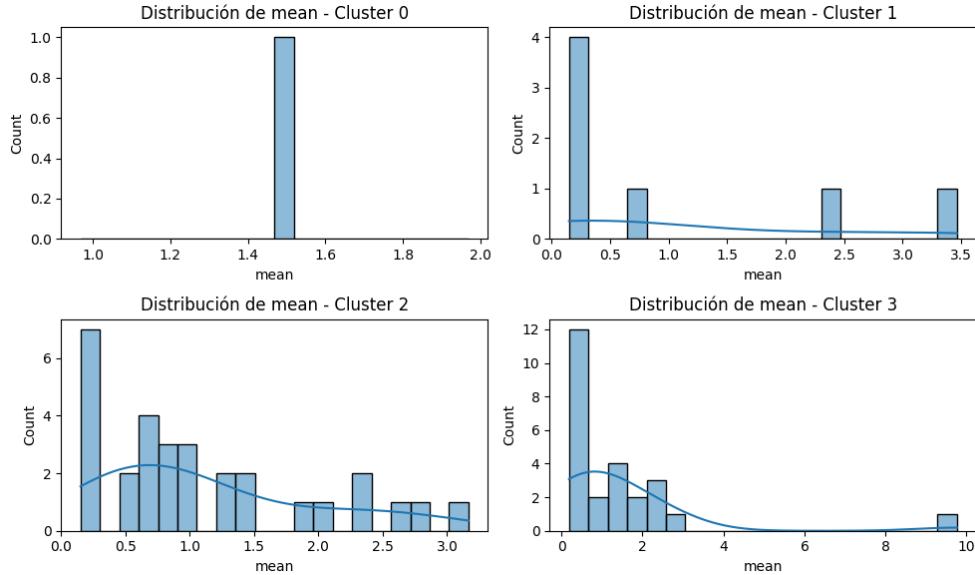


Figure 14: Distribución de cada cluster. Fuente: Elaboración Propia

A partir de la Figura 14, se observan los tiempos promedios de 'STOP' para cada cluster, mostrando en el eje vertical el conteo de estos. Se deduce que cada cluster presenta un comportamiento distintivo, permitiendo identificar una función de distribución que representa la probabilidad de ocurrencia de determinados tiempos promedios. Como aproximación, se destaca que el cluster 0 sigue una distribución normal, mientras que el cluster 1 parece seguir una distribución bimodal. En contraste, los últimos dos cluster exhiben una distribución sesgada, evidenciada por una mayor concentración de datos en los extremos o una inclinación hacia un lado, indicando una densidad de probabilidad más elevada en estas áreas.

Adicionalmente, se presentan los tiempos promedios de 'STOP' para avisos, alarmas y, específicamente, para las fallas. Estos gráficos aportan información valiosa sobre la distribución de los datos en relación con los tiempos promedios de 'STOP', revelando características adicionales de cada cluster.

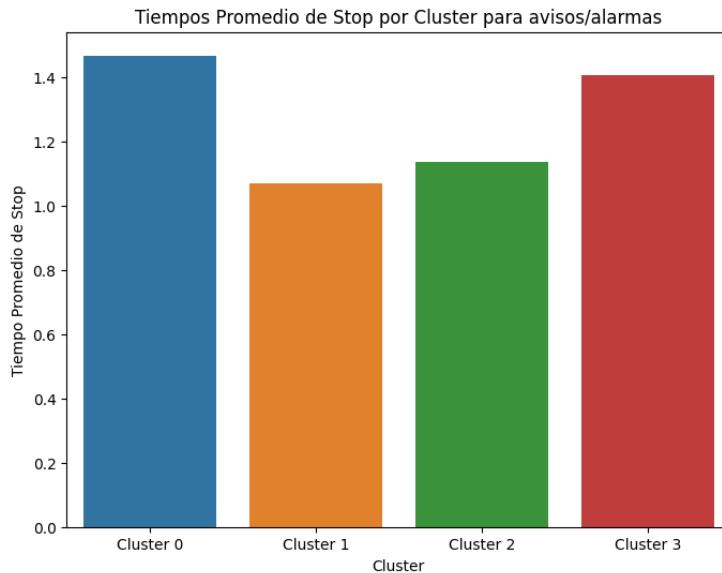


Figure 15: Tiempos promedio de stop avisos y alarmas. Fuente: Elaboración Propia

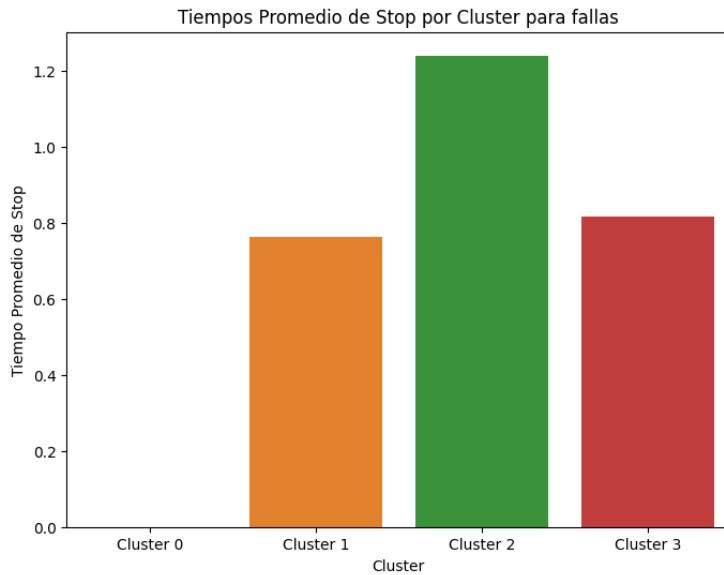


Figure 16: Tiempo promedio de STOP solo fallas. Fuente: Elaboración Propia

Se puede observar que el cluster 0 presenta un alto tiempo de STOP pero no corresponde a fallas debido a que este no cuenta con ninguna. Por otro lado, al filtrar los datos por fallas, se resalta que el cluster 2 posee el tiempo promedio de detención más alto, mientras que los cluster 1 y 3 muestran tiempos promedios de detención similares.

En lo que sigue de la sección se analizarán los resultados importantes obtenidos para el desarrollo de un caso específico de manera analítica como trasfondo y fundamento de la importancia del uso de clusters, el resto de los gráficos se encontrarán en la sección Anexos:

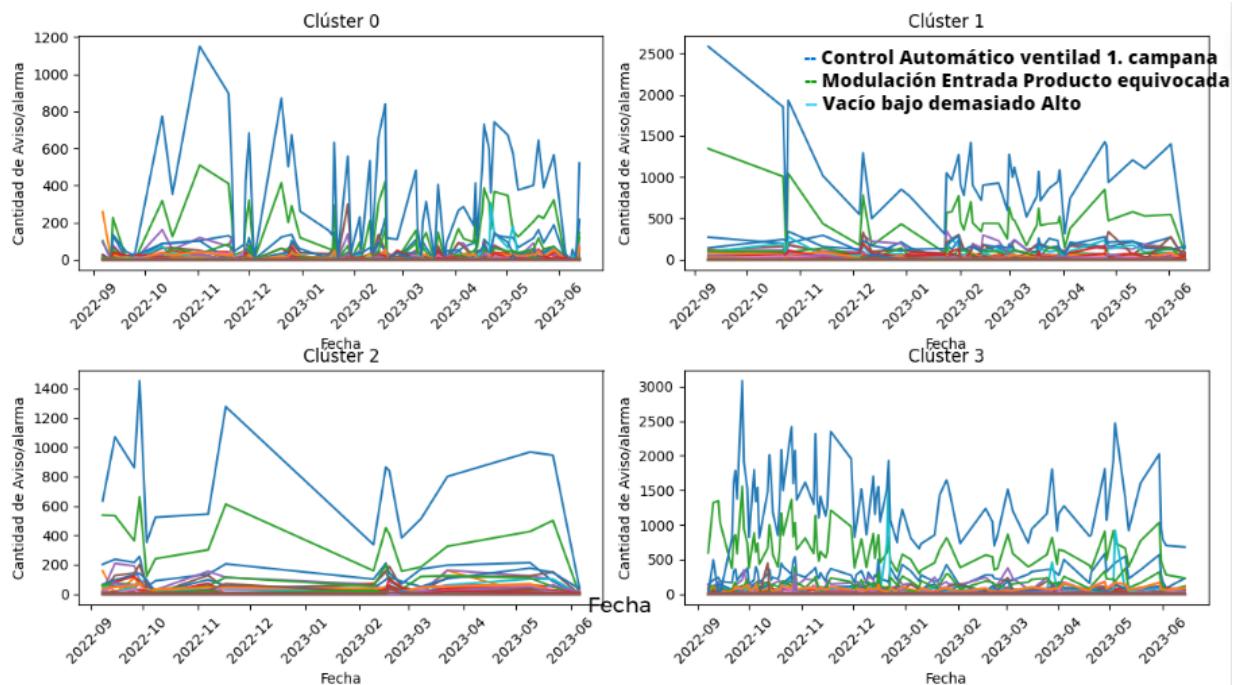


Figure 17: Clusters Avisos/Alarmas. Fuente: Elaboración Propia

A partir de los gráficos se puede identificar que se emplea un enfoque basado en la recopilación y análisis de una cantidad significativa de avisos generados por la máquina, junto con la identificación de patrones de comportamiento similares a través de los clusters. La premisa fundamental radica en la idea de que las fallas en la línea de embotellado suelen precederse de patrones específicos de comportamiento anómalo, que pueden ser detectados a partir de estos avisos y alarmas que han sido registrados. Al analizar y agrupar estos datos, el modelo busca identificar correlaciones y tendencias que permitan predecir cuándo es más probable que ocurra una falla. En el Anexo (9) se encuentran las matrices de correlación por cluster. En cuanto a la correlación entre la frecuencia de avisos y alarmas por cluster y las mantenciones preventivas, detallada en el Anexo (10), se observa que las actuales estrategias de mantenimiento preventivo no ejercen un impacto significativo en la tasa de avisos y alarmas por cluster, esto se puede ver de forma gráfica, dejando en evidencia la necesidad de una revisión del mantenimiento preventivo, el cual se debería enfocar en la implementación de mantenimientos que logren un efecto considerable en las futuras alarmas, avisos y por sobre todo fallas.

## 7.2 Modelo de predicción

Este apartado se enfoca en la elaboración del modelo predictivo, por lo tanto es fundamental el uso de la información obtenida y la elección de una metodología explorada que se ajuste a las características singulares de nuestra problema. La metodología para la elección del modelo predictivo se basó en iterar en los siguientes 3 pasos:

- (a) Desarrollo modelo
- (b) Entrenamiento modelo

El flujo general del modelo final explicado para una falla consiste en primer lugar de una predicción mediante RNN de los avisos y alarmas que condicionan el comportamiento de la falla, esto a través de los datos del clusters. Esto permite extender los datos en 5 días y así conocer el comportamiento de los avisos y alarmas en los 5 días futuros de interés. El resultado de predicción se almacena y se utiliza como input para una segunda RNN que tiene por objetivo determinar la probabilidad de falla en los próximos 5 días, esto lo hace a través del estudio de los patrones de los avisos y las alarmas, identificando las condiciones de estos que originan la falla, este dato es el esperado para entregar la política de mantenimiento a la empresa.

### 7.2.1 Predicción Mediante RNN

Respecto al uso de los datos, los cuales fueron entregados por el equipo docente, se tomaron consideraciones para la utilización y manejo, con el objetivo de evitar sobreajustes y mejorar la precisión del modelo. Para efectos de esta etapa, se optó por tomar un 65% de los datos para entrenar al modelo, y el 35% restante para validación, esto debido a que se probaron distintos valores, y para los datos con los que se cuenta, esta es la proporción más adecuada.

### 7.2.2 Desarrollo de modelo base

Inicialmente se decidió explorar el modelo a partir del aviso “Máquina Ligeramente Desfasada” y su falla relacionada, “Máquina Desfasada”. Se creó una base de datos que incluía únicamente información de estas series de tiempo. Así, se creó el modelo de RNN, con una cantidad arbitraria de epochs y batchs. Se corrió el código varias veces, variando los hiperparámetros hasta llegar a una precisión del conjunto de prueba razonable, tal como se recomendaba en la literatura investigada con respecto a RNN.

En primer lugar, se utilizaron 100 epochs y 3 batchs y se graficaron la perdida y precisión que iba obteniendo el modelo para cada uno.

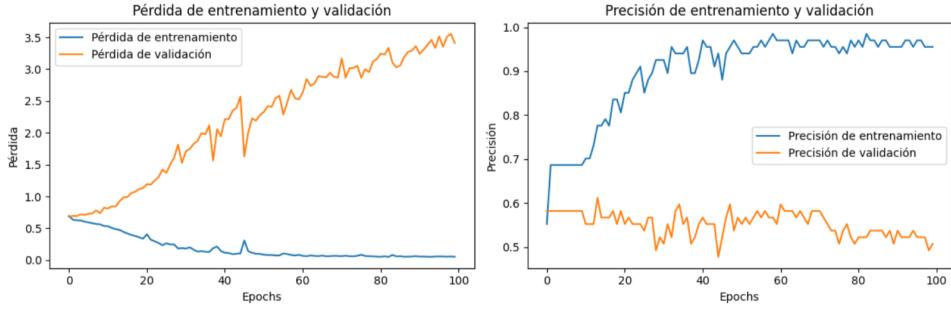


Figure 18: Perdida y precisión de entrenamiento y validación

De la figura anterior, se pudo notar que el modelo se estabiliza a los 40 epochs aproximadamente, por lo que se decidió utilizar esa cantidad. Se mantuvieron los 3 batchs ya que se cuenta con pocos datos para entrenar al modelo, y un número menor introduce más variabilidad en las actualizaciones de pesos en cada paso de entrenamiento. Por esta misma razón, se decidió utilizar en un principio una proporción de 70/30 para entrenamiento y validación, ya que en los últimos datos que se le entrega al modelo, no se contaba con suficiente información para lograr una predicción exacta.

### 7.2.3 Resultado del modelo base

Esta etapa se enfoca en en el análisis y verificación de los resultados obtenidos. El modelo que se utilizó para esta etapa fue de validación cruzada, en su versión más simple, dado la escasez de datos en el caso abordado se utilizó el último 30% de los datos como conjunto de validación. Este conjunto proporciona una estimación imparcial de como se desempeñaría el modelo ante datos desconocidos. Se realizo un gráfico para mostrar como fue la predicción que entrega el modelo en este conjunto para la ocurrencia de “Máquina Desfasada” a partir de la entrega de ocurrencias de “Máquina Ligeramente Desfasada”.

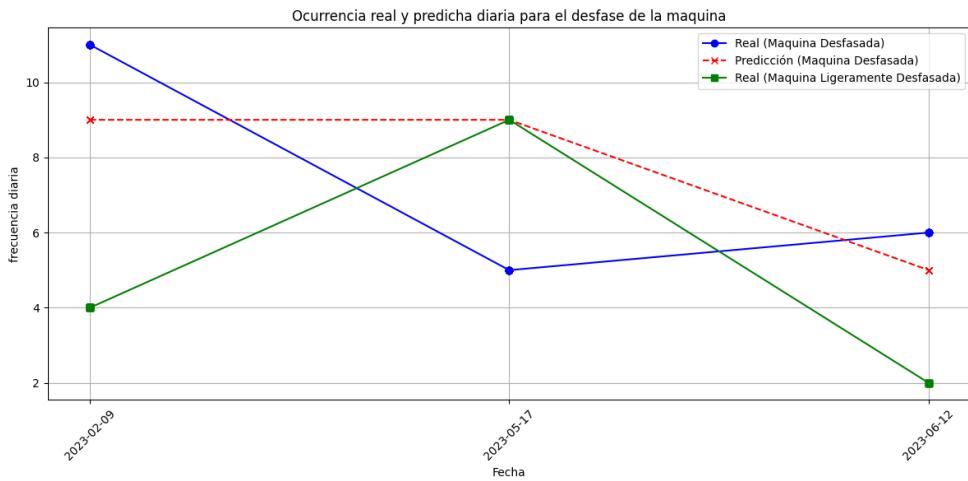


Figure 19: Comparación de ocurrencias reales y predichas por el modelo

Del gráfico obtenido, se observa que existe una diferencia notable entre la predicción del modelo y el escenario real, hecho que no preocupa ya que se sabe que la cantidad de datos que se le entregó al modelo no era suficientemente grande para un entrenamiento óptimo. Aunque no se haya obtenido resultados de utilidad, esto se acerca al modelo de RNN al que se espera llegar, a partir de la unificación de el modelo utilizado con los clusters generados.

Como parte de validación de modelo, se calcularán los KPI's para medir los niveles de desempeño del procedimiento, lo cual para el caso base no es posible debido a la falta de datos y serán calculados una vez el entrenamiento del modelo sea extrapolado a todos los avisos/alarmas.

## 8 Desarrollo del modelo

A partir del modelo obtenido y utilizado en el caso base planteado con los procedimientos mostrados anteriormente, se preparó el modelo para abarcar todos los datos clusterizados, es decir, las predicciones de todas las fallas identificadas dentro de los clusteres obtenidos. Se realiza un proceso de entrenamiento de un modelo de regresión empleando una RNN con el fin de predecir los avisos y alarmas que anticipan las fallas.

Este modelo de RNN se construye a partir de las librerías TensorFlow y Keras, excelentes herramientas en la implementación de modelos de aprendizaje automático y el uso de clusters para poder compilar el modelo y poder realizar un análisis de datos profundo, obteniendo una capa densa con 64 unidades y una capa de salida, ideal para un problema de regresión como es el de predicción de mantenimientos de una línea de embotellado. Para llegar a esto se realiza el siguiente procedimiento: procesamiento de datos, construcción, entrenamiento, evaluación y visualización.

En primer lugar, se pasa por una etapa de pre-procesamiento de los datos, calculando los días transcurridos y adaptando la variable fecha. Luego, se eligen las columnas de entrada y la columna objetivo a predecir la cual representa un aviso o alarma y se define la longitud de la secuencia para tener consideradas las relaciones temporales de los datos al ser series de tiempo. Después de preparar correctamente las secuencias, estas se dividen en entrenamiento y prueba, de las cuales solo un porcentaje está destinado al conjunto de prueba, el cual puede ir variando. Esto se requiere ya que el modelo finalmente solo se evalúa con los datos de prueba para posteriormente calcular la pérdida.

A partir de todo esto, al correr el código se logra visualizar la pérdida de entrenamiento y la validación a lo largo de todas las épocas generando predicciones en el conjunto de datos de entrenamiento que se comparan con los valores reales en función del tiempo y se puede observar en los gráficos obtenidos.

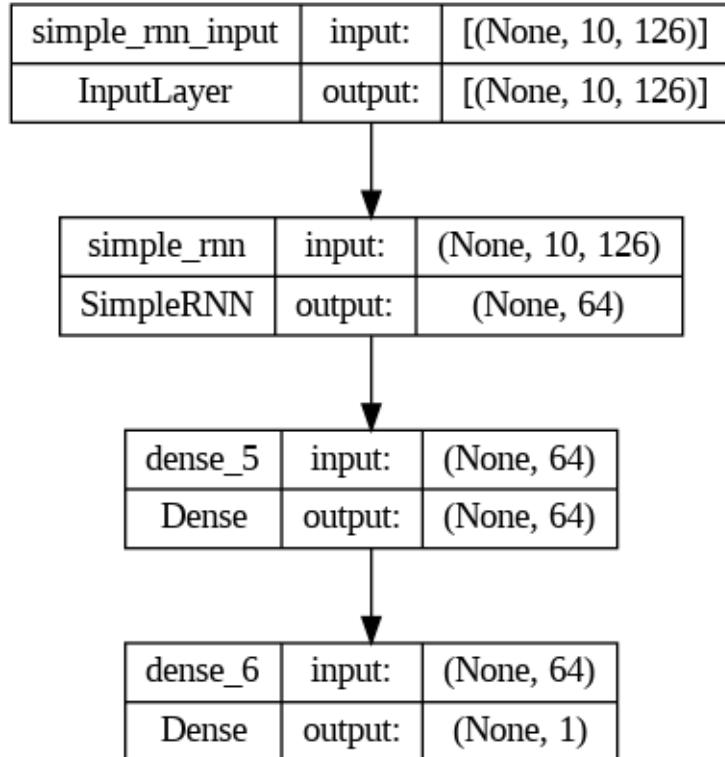


Figure 20: Arquitectura de la Red Neuronal. Fuente: Elaboración Propia

Según la imagen, se observa la arquitectura de la red neuronal empleada a través del uso de 4 dataframes en los cuales se construye un modelo con capas. La primera capa, la simple, tiene 64 unidades y es la encargada de procesar las secuencias de entrada y capturar patrones temporales de las series de tiempo.

Luego viene la capa densa, con 64 unidades y su función es realizar transformaciones no lineales en los datos de salida provenientes de la capa anterior. Por último, la capa de salida también es de naturaleza densa pero tan solo posee 1 unidad, esta capa obtiene la predicción final basada en las características de los datos de entrada. Todo esto es una representación visual de los pasos explicados anteriormente, los cuales son el manejo de datos, procesamiento y refinamiento.

## 8.1 Análisis casos extremos en la predicción

Como con cualquier tecnología de modelado predictivo, las RNN también enfrentan desafíos en su aplicación. Los casos extremos de predicciones son aquellos en donde el modelo se ajusta perfectamente a ciertos avisos o alarmas; o cuando este no logra predecirlas en absoluto. El análisis de estos casos extremos proporciona detalles cruciales en la robustez y la fiabilidad del modelo, y es esencial para entender sus limitaciones y potencial para mejoras.

En el caso óptimo, un modelo RNN que se ajusta perfectamente a los datos puede identificar con precisión las señales de advertencia de avisos y alarmas antes de que ocurran, permitiendo acciones preventivas y reduciendo el tiempo de inactividad. Este tipo de ajuste es el escenario ideal, donde el modelo ha aprendido con éxito las complejas dependencias temporales y es capaz de generalizar este aprendizaje para hacer predicciones precisas en datos no vistos previamente.

En contraste, un modelo que no logra ajustarse adecuadamente puede ser indicativo de varios problemas subyacentes. Esto puede ocurrir debido a una cantidad insuficiente de datos de entrenamiento, llevando a falta de representatividad en los datos, sobreajuste a las particularidades de los datos de entrenamiento (falta de generalización), o una inadecuada arquitectura de red para capturar la complejidad de las señales de alerta.

Dentro del espectro de desempeño del modelo RNN, se encuentra un caso particularmente desafiante: el aviso "Control Protecciones Traseras Abiertas" del cluster 0. Este cluster, caracterizado por tener un volumen reducido de datos, plantea un problema para el aprendizaje efectivo del modelo, reflejado en su incapacidad para predecir correctamente dicho aviso, como se observa en el gráfico a continuación:

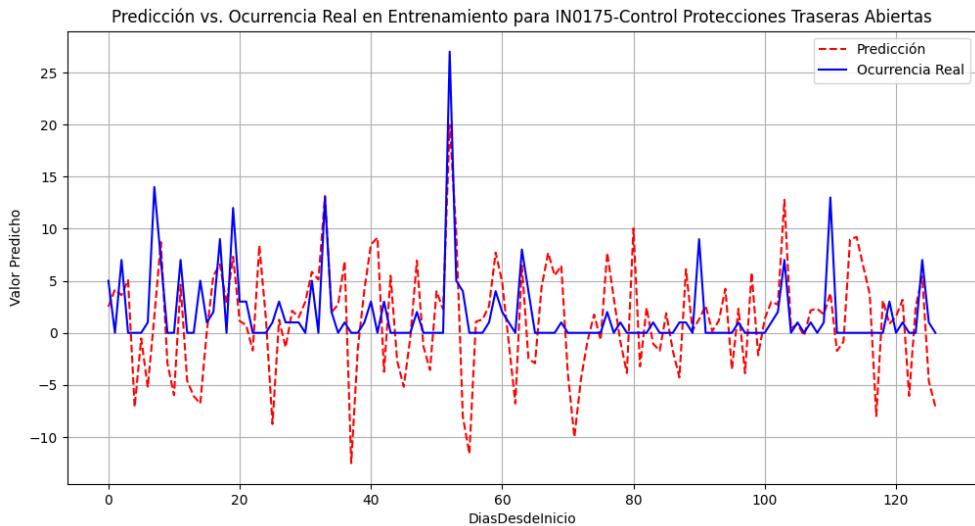


Figure 21: Predicción versus Ocurrencia Real en el Conjunto de Entrenamiento: IN0175 - Control Protecciones Traseras Abiertas

La limitada cantidad de datos de este cluster colleva a que no haya suficientes ejemplos de falla para que el modelo capture y aprenda las complejidades asociadas con este evento. Esta escasez de información puede resultar en una generalización pobre y una capacidad predictiva limitada. La situación se acompleja si consideramos que la data escasa a menudo va acompañada de un ruido considerable y de una falta de señales claras que preceden a una falla.

Por otro lado, si se observa la contrapuesta, al realizar una predicción de un aviso o alarma perteneciente a los clusters 2 o 3. Estos clusters, se caracterizan por su alto volumen de datos, lo que se traduce en un

funcionamiento prácticamente óptimo del modelo. En el siguiente gráfico se puede ver la predicción de la alarma "IN0071-Señal Emergencia Presionada" la cual pertenece al cluster 2.

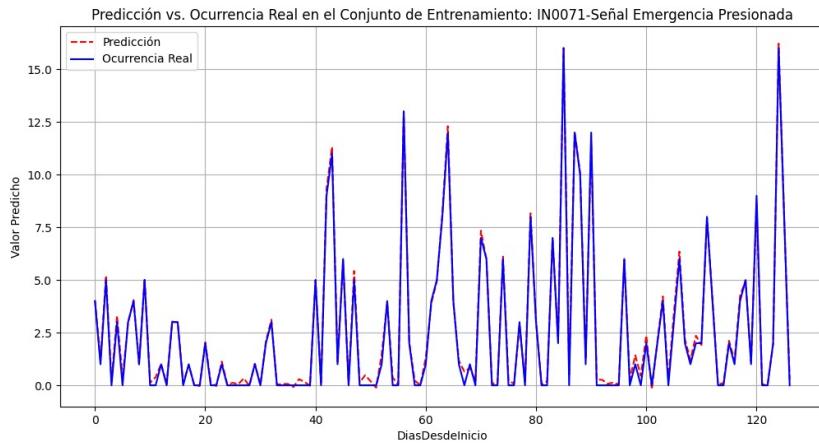


Figure 22: Predicción versus Ocurrencia Real en el Conjunto de Entrenamiento: IN0071 - Señal Emergencia Presionada

En el gráfico anterior, se observa una tasa de aciertos bastante alta, es decir, que esta alarma, puede ser predichas de manera casi perfecta con el modelo. Esto se debe a que la gran cantidad de datos de este cluster permite al modelo aprender y capturar el complejo comportamiento de este tipo de señal, de forma que permite al modelo realizar una predicción acertada y muy similar a la ocurrencia real..

## 8.2 Predicción de probabilidad de falla mediante RNN

Con el objetivo de generar un dato objetivo de predicción, que sea útil para la empresa y en una ventana de temporal que permita a esta reaccionar frente a una falla, se decidió aprovechar la habilidad de la primera RNN para predecir los avisos y alarmas que se relacionan con cada una de las fallas. En esta segunda RNN, se busca predecir la probabilidad de falla a través del aprendizaje del comportamiento de los avisos y alarmas que dan origen a las fallas.

La elección de la estructura de la red se basó en la implementación de un método ya probado en la literatura, específicamente en el libro “Deep Learning for Time Series Forecasting” de Jason Brownlee, en este se enfoca en aplicar técnicas de aprendizaje profundo, incluyendo redes neuronales recurrentes como las LSTM, a problemas de predicción de series de tiempo. Se creó un modelo secuencial de Keras. Este modelo utiliza una capa LSTM con 64 unidades y una capa de salida con una sola neurona y activación sigmoide que permite el resultado en forma de probabilidad.

Algunos beneficios de este tipo de red son:

- Manejo de secuencias largas: Las LSTM (Long Short-Term Memory) son capaces de aprender y recordar patrones en secuencias largas, lo cual es útil en series temporales donde las dependencias a largo plazo son importantes.

- Memoria a largo plazo: Las LSTM tienen una estructura que les permite retener y olvidar información a lo largo del tiempo, lo que es crucial para modelar el comportamiento dinámico y las tendencias en las series de tiempo.

### 8.2.1 Predicciones

Se aplicó la predicción para 2 tipos de fallas que poseen distintas características. La primera falla fue “Máquina Desfasada” esta es una falla que dentro de los datos posee pocos registros de ocurrencia, es-

pecíficamente 6 en todo el lapso temporal, mientras que la segunda es la falla “OUT0210-Válvula Entrada Producto Cerrada por Falta Consentimiento de Centralita C.I.P.” que tiene una alta frecuencia con 113 ocurrencias en el periodo de análisis de 10 meses. Los resultados se muestran a continuación en las siguientes tablas.

Table 4: Probabilidad de ocurrencia de Maquina Desfasada

	10/06/2023	11/06/2023	12/06/2023	13/06/2023	14/06/2023
Ocurrencia Real	0	0	4	0	0
Probabilidad de falla	0.39473101	0.29298905	0.6966645	0.29298905	0.318477

La tabla muestra un periodo de tiempo de 5 días donde en uno de ellos específicamente el 12 de junio ocurrió la falla del tipo “Máquina Desfasada”, el modelo predictivo es el encargado de entregar la probabilidad de falla obteniendo un valor de 0.69 cuando la falla realmente ocurrió y valores menores cuando no ocurre.

Table 5: Probabilidad de ocurrencia de OUT0210-Válvula Entrada

	10/06/2023	11/06/2023	12/06/2023	13/06/2023	14/06/2023
Ocurrencia Real	34	0	1	3	19
Probabilidad de falla	1	0.20889743	1	1	1

Al contrario de la tabla anterior, para la falla de “OUT0210-Válvula Entrada Producto Cerrada por Falta Consentimiento de Centralita C.I.P”, existe de los 5 días un solo día donde esta no ocurre. Esto quiere decir que el día 10/06/2023 la falla ocurrió con una probabilidad de 100%, mientras que la probabilidad para el día siguiente, dado que ya ocurrió el día anterior disminuye a un 20,8%. Esto indica las altas probabilidades que muestran cuando ocurren y el descenso cuando no ocurre.

## 9 Análisis de Sensibilidad

### 9.1 Batch size y Epoch

La función utilizada para entrenar la RNN depende de la caracterización de 2 parámetros los cuales son Batch size y Epoch (sus características están definidas en el apartado 3.2 de Definiciones). La definición de estos parámetros se debe realizar previo a la aplicación del modelo debido a que afectan en la calidad de la predicción, por esto se analizó como responde la predicción de cada uno de los avisos y alarmas en base a la modificación de estos parámetros.

La elección de estos parámetros dependen de varios factores incluyendo el tamaño del conjunto de datos, la arquitectura del modelo y la cantidad de memoria disponible. En base a la división de cluster como estructura de datos que son ingresados al modelo de predicción, es que se analizó en base al tamaño de cada uno de ellos la definición de los parámetros a implementar.

El tamaño de cada uno de los clusters se especifican a continuación:

- Cluster 0: 212x6
- Cluster 1: 212x16
- Cluster 2: 212x58
- Cluster 3: 212x50

Por convención algunos de los valores comunes para los parámetros son para batch size 32, 64, 128 o 256 mientras que para epoch son 10, 20, 50, 100 o 200. Para cada una de las fallas pertenecientes a cada cluster se realizaron las siguientes variaciones de los parámetros hasta observar gráficamente un gráfico de predicción donde el comportamiento sea igual o lo más cercano posible a la ocurrencia real. A continuación se detallan los valores de referencia que se utilizaron para variar los parámetros según cada uno de los clusters.

	Batch size	Epoch
Cluster 0	32	50
Cluster 1	64 - 128	50 - 100
Cluster 2	128 - 256	100 - 200
Cluster 3	128 - 256	100 - 200

Para cada una de las fallas se utilizaron los siguientes valores para la cantidad de épocas de entrenamiento (epochs) y tamaño de lote (Batch Size).

Codigo	Cluster	Frecuencia	Epoch	Batch Size	R <sup>2</sup>	MSE
8	Cluster 1	157	92	64	0.7270	0.6357
9	Cluster 2	1224	196	128	0.9527	0.4310
10	Cluster 3	2522	100	128	0.9771	4.3021
11	Cluster 3	173	196	128	0.9916	0.0673
11	Cluster 2	1886	120	150	0.8889	20.3907
12	Cluster 2	394	196	128	0.9987	0.0662
12	Cluster 3	1370	192	128	0.9521	1.0687
13	Cluster 3	430	150	145	0.9802	0.3081
16	Cluster 3	9210	200	128	0.9909	0.0465
17	Cluster 3	20059	150	100	0.9998	0.8785
20	Cluster 3	38911	150	100	0.9989	2.7815
21	Cluster 2	12	200	128	0.9027	0.1206
21	Cluster 2	17757	150	145	0.9764	7.0046
22	Cluster 2	14	200	128	0.7187	0.1443
23	Cluster 2	10532	150	145	0.9964	1.4089
25	Cluster 2	1154	150	145	0.9999	0.1057
30	Cluster 3	107	180	128	0.9998	0.0096
31	Cluster 3	2048	120	130	0.3919	5.5918
32	Cluster 2	82	100	128	0.2511	0.8058
46	Cluster 2	1718	150	160	0.9995	0.1504
73	Cluster 3	1950	150	100	0.9993	0.2021
74	Cluster 3	345	150	100	0.9576	0.4934
77	Cluster 3	198	100	100	0.6817	0.7071
81	Cluster 2	95042	150	100	0.9611	48.6438
86	Cluster 2	74	100	128	0.7745	0.3649
112	Cluster 2	266	100	128	0.9914	0.1079
124	Cluster 1	35	50	64	0.9687	0.2082
129	Cluster 3	2	200	130	0.7037	0.0093
145	Cluster 2	99	150	130	0.8309	0.8109
194	Cluster 2	743	100	120	0.9995	0.0601
203	Cluster 3	44	100	130	0.0000	5.4896
205	Cluster 3	2632	150	130	0.9961	3.0629
206	Cluster 3	15796	100	128	0.958	3.56
237	Cluster 2	1251	200	256	0	1.6856
257	Cluster 1	55	80	65	0.9746	0.1846
285	Cluster 3	5	150	230	0.6500	0.0745
286	Cluster 2	1	100	130	0	1.3369
293	Cluster 3	24	180	200	0.8999	0.0987
434	Cluster 2	829	180	200	0.9550	0.7470

A partir de la tabla, se observa el código del aviso/alarma, el cluster al cual pertenece, la frecuencia, Epoch, Batch Size, coeficiente de determinación ( $R^2$ ) y el error absoluto medio (MSE). Todo esto se obtuvo o se utilizó como información para el código del entrenamiento del modelo. Los dos últimos indicadores son cruciales para determinar el nivel de predicción que se está logrando.

El coeficiente de determinación ( $R^2$ ) es una medida que calcula la proporción de la varianza en la variable dependiente, esto con el fin de identificar la variabilidad de la variable de interés, en otras palabras, es la bondad de ajuste. Pueden haber 3 opciones, primero que el coeficiente sea igual a 0, lo cual se puede

observar en la tabla anterior. Esto significa que el modelo no explica nada de la variabilidad en los datos, por lo tanto, sería necesario analizar esos avisos/alarmas y volver a plantear la definición de fallas, ya que estas pueden que no sean relevantes para la resolución del problema planteado. Por otro lado, la segunda opción es que el coeficiente sea entre 0 y 1, lo cual representa la fracción de la variabilidad de los datos, la idea es que sea lo más cercano a 1. Por último, si el coeficiente es 1, esto explica que el modelo explica perfectamente la variabilidad en los datos, donde cada punto se encuentra dentro de la línea de regresión y el modelo tiene un gran ajuste. Esto ocurre casi perfectamente con la alarma: "IN0467-Control Temperatura No Adec. Pistola Pegamen."

El error absoluto medio, MSE (por sus siglas en inglés), es una medida que calcula la diferencia promedio entre los valores reales y los valores predichos. En general, mientras menor sea el MSE, mejor es la predicción pero esto también tiene que tomar en cuenta los factores de frecuencia, dependiendo directamente en la data recolectada.

En el Anexo (11) se encuentran graficadas las predicciones realizadas para cada aviso y alarma versus la ocurrencia real de estas mismas, cada una con el gráfico de perdida de entrenamiento y validación para cada época del modelo. Los valores de la tabla anterior se obtuvieron mediante varias pruebas hasta llegar a una predicción más cercana a la realidad, la que se puede apreciar en el gráfico obtenido para cada uno de estos, donde la linea punteada roja corresponde a la predicción a la que se llegó.

En el caso de la segunda RNN, se observa que para la falla "OUT0210-Válvula Entrada Producto Cerrada por Falta Consentimiento de Centralita C.I.P." existe una mejora en su rendimiento y en la probabilidad de ocurrencia de la predicción en el momento en el que aumentan los hiperparámetros de epoch y batch size. Esto viene dado por el hecho de que aumentar los hiperparámetros de epoch y batch size permite al modelo de la RNN procesar más información en cada iteración y durante un período de tiempo más prolongado. Esto mejora la capacidad del modelo para aprender patrones más complejos y sutilidades en los datos.

A medida que se incrementan los epochs, el modelo tiene más oportunidades para ajustar sus pesos, lo que conduce a una mejora en la precisión de las predicciones. Por otro lado, un batch size mayor significa que el modelo puede procesar un mayor volumen de datos en cada iteración, lo cual es particularmente útil en situaciones donde la complejidad de los datos es alta o cuando los patrones son difíciles de discernir.

En el caso específico de la falla mencionada, es probable que los patrones relacionados con esta falla sean complejos y requieran de un análisis más detallado, lo cual se facilita con un aumento en los hiperparámetros mencionados, lo que se observa en la tabla a continuación, donde un aumento en los valores de los hiperparámetros mejora la probabilidad de la predicción.

batch_size , epoch	10/06/2023	11/06/2023	12/06/2023	13/06/2023	14/06/2023
Ocurrencia real	34	0	1	3	19
32 , 50	1	0.46915692	1	1	1
64-128 , 50-100	1	0.2060151	1	1	1
128-256 , 100-200	1	0.19002	1	1	1

Table 6: Análisis de sensibilidad de la probabilidad de ocurrencia de OUT0210

A diferencia de la falla anterior hay casos en los que aumentar el batch size y la cantidad de generaciones puedes ser perjudicial para la predicción, como se puede observar en la siguiente tabla para la falla "Máquina Desfasada":

batch_size , epoch	10/06/2023	11/06/2023	12/06/2023	13/06/2023	14/06/2023
Ocurrencia real	0	0	4	0	0
32 , 50	0.64856434	0.05227916	0.9920787	0.05227919	0.20016035
64-128 , 50-100	0.30257407	0.1491163	0.30185017	0.49461216	0.22853653
128-256 , 100-200	0.02280978	0.07344899	0.07028946	0.02152944	0.07735023

Table 7: Análisis de sensibilidad de la probabilidad de ocurrencia de Máquina Desfasada

Como se puede observar para predicciones de fallas del cluster 1 como es el caso de "Maquina Desfasada", cluster caracterizado por tener menor información, en comparación con la falla que se mostró en la tabla anterior a esta, la que al ser parte del cluster 2 cuenta con mayor información para realizar la predicción por lo tanto no es tan perjudicial el usar mucha información para la predicción.

## 9.2 Test Size

En la evaluación del desempeño del modelo RNN, un aspecto que emerge como crucial es el tamaño de la muestra de prueba (test size). El análisis de sensibilidad revela que el test size es un parámetro con un impacto diferencial en la predicción, dependiendo significativamente del volumen de datos presentes en cada cluster. Para los clusters con una cantidad abundante de datos, un test size mayor puede ser beneficioso, ya que proporciona una evaluación más robusta y fiable del modelo, permitiendo que el algoritmo se beneficie de un conjunto de entrenamiento más grande para aprender las dinámicas complejas de las fallas.

Por contraste, en los clusters con menos datos, como es el caso del cluster de la falla “Control Protecciones Traseras Abiertas”, la elección de test size se convierte en una decisión más crítica. Un tamaño de prueba demasiado grande podría privar al modelo de los pocos datos de entrenamiento que necesita para identificar patrones significativos. En estos casos, puede ser preferible optar por un test size menor para maximizar el volumen de datos de entrenamiento, aunque esto conlleve una evaluación menos extensa de la capacidad predictiva del modelo. Esto se puede observar en la variabilidad visual de los gráficos al variar este parámetro a continuación:

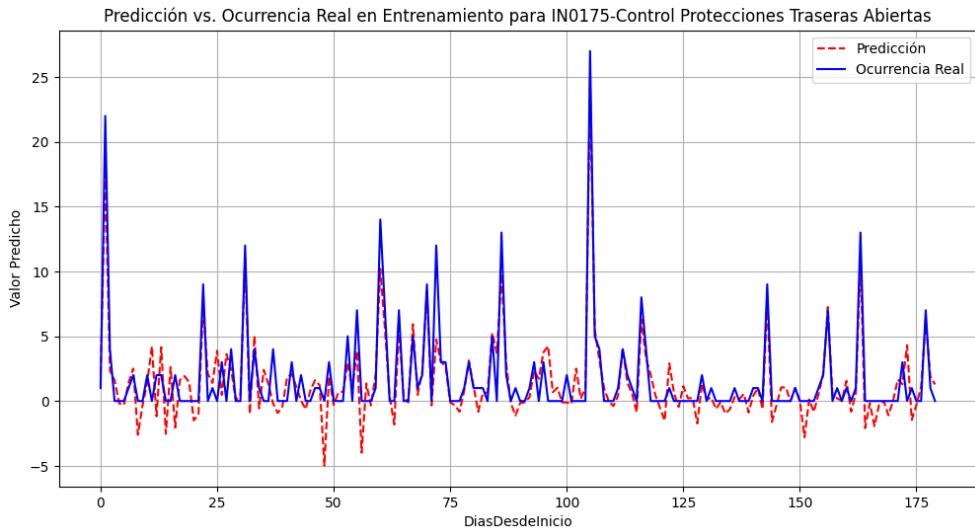


Figure 23: Predicción vs ocurrencia con test size = 0.1. Fuente: Elaboración Propia

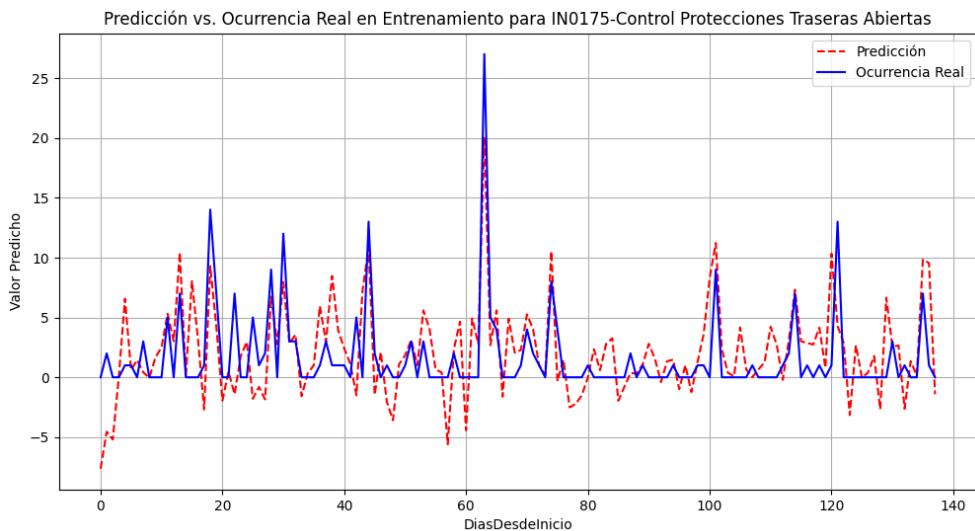


Figure 24: Predicción vs ocurrencia con test size = 0.3. Fuente: Elaboración Propia

Este equilibrio delicado entre la cantidad de datos de entrenamiento y la robustez de la validación es particularmente relevante cuando se habla de eventos de falla anormales o poco frecuentes. La estrategia

de partición de datos debe, por tanto, ser cuidadosamente considerada y posiblemente adaptada para cada cluster individualmente, en lugar de aplicar una regla generalizada.

## 10 Fast Fourier Transform

Un paso que se realizó luego de ejecutar el modelo nombrado anteriormente, fue la utilización del algoritmo Fast Fourier Transform (FFT). Este algoritmo se utiliza para encontrar frecuencias de las fallas a lo largo del tiempo. Aplicar FFT a las fallas de la máquina, entregaría información en el dominio de la frecuencia. La interpretación de los resultados de la FFT puede proporcionar insights valiosos sobre las características de las señales de falla en la máquina. El aplicar este algoritmo puede ayudar a encontrar frecuencias dominantes en los datos. Incluir ambos algoritmos en nuestro modelo, RNN y FFT, puede permitir aprovechar la información tanto en el dominio del tiempo como en el de la frecuencia, puede mejorar la capacidad predictiva y esto podría ayudar a anticipar fallas potenciales en la maquinaria al detectar señales a partir de avisos o patrones ocultos en los datos.

Luego de la implementación de este método, se observó que el residuo obtenido reflejaba una considerable variabilidad o 'movimiento'. Esto sugiere que hay una gran cantidad de datos que parecen ser impredecibles o no están vinculados a patrones específicos de frecuencia identificables mediante la transformada rápida de Fourier (FFT). En otras palabras, aunque se realizó un análisis utilizando la FFT, no se encontraron componentes de frecuencia significativos que pudieran explicar la variabilidad en estos datos.

Para un mayor entendimiento de lo que se realizó en esta etapa, revisar el Anexo (12).

## 11 Evaluación de KPI'S

Inicialmente, se definieron varios indicadores clave de rendimiento (KPIs) para evaluar la efectividad del modelo de mantenimiento predictivo. Aunque en esta fase se ha logrado predecir dos tipos de fallas, el modelo tiene el potencial de extenderse a todas las que se mencionan en el apartado de "Definición de Falla". Basándose en esta premisa, se procede a analizar los KPIs establecidos:

- (a) Tiempo de Actividad en Producción: Este KPI evalúa la eficiencia operativa, particularmente en cuanto al tiempo de actividad de la maquinaria. Inicialmente, se observó que las fallas podían inutilizar la máquina por alrededor de 6 días. Con la implementación del modelo de mantenimiento predictivo, este período se logra reducir significativamente. Al anticipar las fallas en un horizonte de 5 días, se ha conseguido incrementar el tiempo de actividad de la maquinaria.
- (b) Diferencia de Capacidad Real vs. Capacidad Propuesta: Este indicador evalúa la capacidad productiva real de la maquinaria en comparación con la capacidad teórica. Gracias a la disminución de interrupciones por fallas, la capacidad de producción real aumenta. Con la máquina deteniéndose durante seis días, la diferencia entre la capacidad actual y la potencial supera el millón de botellas, donde se destaca la ganancia potencial en términos de producción. Además, el modelo permite anticipar problemas que podrían ocasionar daños mayores, como la rotura de una botella, optimizando así la producción total de la maquinaria.
- (c) Tasa de Avisos y Alarmas Diarias: Este indicador mide la frecuencia con la que se emiten avisos y alarmas por la maquinaria. Se espera que, con la adopción del modelo predictivo propuesto, la tasa de estos avisos y alarmas disminuya progresivamente, lo que indicaría una reducción en la aparición de fallas o problemas en la maquinaria a futuro.

## 12 Propuesta de política

La base de datos es un componente fundamental en las operaciones y toma de decisiones. La integridad de los datos es esencial para garantizar que la organización funcione de manera precisa. Para abordar este importante aspecto, se debe asegurar la implementación de una política que establezca procedimientos claros para la detección y mitigación de fallas en la base de datos. En primer alcance se debe definir explícitamente las fallas en las bases de datos. Actualmente, las bases de datos solo incluyen avisos, alarmas, mantenciones y status, por lo que la información de fallas dentro de esto es nula, y tal como se especificó en las secciones anteriores, hay avisos que causan fallas catastróficas en la línea de embotellado

y no se pueden obviar o solo catalogar como avisos/alarmas. Además, hay algunas ocurrencias que están dentro de ambas bases de datos, lo que dificulta la compresión de la diferencia entre avisos y alarmas. Definiendo bien cada suceso de la máquina, se evita que se ensucien las bases de datos al catalogar sucesos de forma errónea, donde el mejor ejemplo planteado es el aviso de “Máquina Desfasada”.

Por otro lado, nos encontramos con registros de mantenciones erróneas, que dentro de las ocurrencias de la maquina no calzaban las mantenciones realizadas con las fechas que se entregaron en la base de datos de mantenciones, lo que dificultó aun mas encontrar la correlación de mantenciones con cantidad de fallas. Es por esto que se propone a la empresa, además de definir bien los sucesos, contar con un registro efectivo de las mantenciones y revisar su información a diario. De esta manera, el modelo predictivo entregado, tendrá un mejor desempeño y se lograra mejorar aun mas el rendimiento de la maquina, y finalmente, un aumento en la productividad de la empresa.

Con todo el trabajo realizado y con las mejoras que se proponen, se le proporciona a la empresa un mecanismo para anticipar situaciones anómalas o no deseadas para la producción de botellas de vino. Este modelo permite adelantar los comportamientos de sus avisos y alarmas para poder prepararse para la ocurrencia de fallas ligadas a estos sucesos. Para esto se entrega una herramienta de la cual se puede hacer uso diario para obtener las probabilidades de falla de la máquina hasta cinco días a futuro, para poder prepararse para realizar intervenciones correctivas y reducir los tiempos de detención y las pérdidas de producción ligadas a estos lapsos de tiempo.

## 13 Conclusiones

A modo de síntesis, el proyecto de generación de mantenimiento predictivo para una empresa vitivinícola se ha visto marcado por una serie de hitos significativos, impulsado por la implementación efectiva de una carta Gantt (en **Anexos**), la que trazo la ruta de objetivos para llegar a una propuesta. El cambio de enfoque de mantenimiento, pasando de uno preventivo y correctivo a uno predictivo, representa un avance significativo hacia la mejora de las operaciones en la línea de embotellamiento de la empresa. Este cambio significaría un aumento considerable en la eficiencia operativa, disminución de detenciones innecesarias, gracias a la optimización de la gestión del mantenimiento, permitiendo cumplir de mejor manera el objetivo de satisfacer la demanda de manera más efectiva.

En primer lugar, se hizo un análisis exhaustivo de las bases de datos, las que abarcan un período de 10 meses y más de 4 millones de observaciones, lo cual ha sido esencial en este proceso. Mediante la definición de metas y la identificación de indicadores de desempeño, se encontró el camino para evaluar el desempeño de los modelos predictivos que se desarrolle. Sin embargo, es importante recalcar la posible influencia de variables externas en el problema, las cuales pueden llegar a influir en la producción, en las fallas y así mismo en los resultados que proporciona nuestro modelo. Además de ello, se lograron identificar problemas con las bases de datos, tales como traducciones erróneas de los avisos de la maquina, ya que esta es originalmente italiana; problemas en los registros de la mantenciones y clasificación de avisos y alarmas poco explícitas. A partir de eso se estipularon algunas recomendaciones para esta empresa, con el fin de generar un registro efectivo de los acontecimientos de la maquina, y de definir explícitamente que es una falla e identificarlas dentro de sus datos, así poder tener un entrenamiento efectivo del modelo y obtener resultados mas acertados.

El análisis de literatura sobre diversas metodologías, que incluyen redes neuronales recurrentes, métodos de suavización exponencial, el método Holt-Winter, el método de Croston, y el enfoque de SBA y Fast Fourier Transform, nos han brindado una comprensión profunda de cada enfoque en el contexto específico del problema a resolver en la línea de embotellamiento. Entre estas metodologías, se espera que RNN se adapte de mejor manera al problema, gracias a la capacidad que tiene de trabajar grandes volúmenes de datos variados. Esto se pudo comprobar con la comparación de resultados obtenidos con la red, versus los de la FFT, que no llegó a los resultados que se esperaba.

Al entrenar, validar y testear un modelo predictivo de RNN, se pudo identificar diversos patrones de comportamiento, por ejemplo, se identifico que para una falla, que tiene alta frecuencia, la predicción del modelo tiende a ser más cercana a la realidad. Por otro lado también se observo un patrón de comportamiento distinto en las fallas que se encuentran en clusters con pocos datos, como el cluster 0, en estas fallas debido a la limitada cantidad de datos, el modelo tiende a realizar predicciones mas alejadas de la realidad.

En conclusión, tras el análisis de mas de 10 meses de información y datos, este proyecto es un el puntapié inicial en la implementación de un modelo de mantenimiento predictivo con datos categóricos, logrando mejorar la eficiencia y efectividad de las operaciones que realiza la maquina monoblock en el marco de la producción vitivinícola en Chile.

## 14 Referencias Bibliográficas

- Bonet Cruz, I., Salazar Martínez, S., Rodríguez Abed, A., Grau Ábalos, R., & García Lorenzo, M. M. (2007). Redes neuronales recurrentes para el análisis de secuencias. *Revista Cubana de Ciencias Informáticas* 1(4), 48-57. Recuperado de: <https://www.redalyc.org/pdf/3783/378343634004.pdf>
- Gutierrez, J. M. (2021, junio 10). ¿Qué es la autocorrelación? Test de Durbin-Watson. Todo Econometría y ciencia de datos. <https://todoeconometria.com/autocorrelacion-dw/>
- Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, *OTexts: Melbourne, Australia*. <https://otexts.com/fpp2/expsmooth.html> <https://otexts.com/fpp2/holt-winters.html>
- Nieves, D. ¿Qué es el escenario de entrenamiento, validación y prueba de conjuntos de datos en aprendizaje automático?. Recueroado de: <https://es.quora.com/En-el-aprendizaje-automítco>
- Mira Segura, L. L., Trejo Martínez, A., & López Cruz, D. (2018). *Aplicación de Holt-Winters para pronósticos de inventarios*. CIENCIA UANL, 21(90). DOI: <https://doi.org/10.29105/cienciauanl21.90-2>
- López Camuñas, J. M. (2021). Predictive Maintenance Using Deep Learning. Tesis de grado, Grau d'Enginyeria Informàtica, Facultat de Matemàtiques i Informàtica, Universitat de Barcelona. Recuperado de: [https://deposit.ub.edu/dspace/bitstream/2445/182392/3/tfg\\_josemanuellopezcamu](https://deposit.ub.edu/dspace/bitstream/2445/182392/3/tfg_josemanuellopezcamu)
- Pérez Ortiz, J. A. (2002). Modelos predictivos basados en redes neuronales recurrentes de tiempo discreto. *Tesis doctoral, Universidad de Alicante* Recuperado de: [https://www.dlsi.ua.es/\\_japerez/publications/tesi2002.pdf](https://www.dlsi.ua.es/_japerez/publications/tesi2002.pdf)
- Rao, K. R., Kim, D. N., & Hwang, J.-J. (2010). *Fast Fourier Transform - Algorithms and Applications*. Springer Dordrecht. DOI: <https://doi.org/10.1007/978-1-4419-6629-0>
- Santa Cruz, R. & Correa, C. (2017). Previsión de demanda intermitente con métodos de series de tiempo y redes neuronales artificiales: Estudio de caso. DOI: <https://doi.org/10.15446/dyna.v84n203.63141>
- Soni Madhulatha, T. (2012). An overview on clustering methods. SR Journal of Engineering, 2(4), 719-725. Recuperado de: <https://arxiv.org/ftp/arxiv/papers/1205/1205.1117.pdf>
- Torres, J. (2019). Redes Neuronales Recurrentes. Recuperado de: <https://torres.ai/redes-neuronales-recurrentes/>
- Vandeput, N. (2021). Data Science for Supply Chain Forecasting (2nd ed.). *De Gruyter*. Recuperado de: <https://towardsdatascience.com/simple-exponential-smoothing-749fc5631bed>

## 15 Anexos

### 15.1 Carta Gantt

Para ver la Carta Gantt se debe presionar aquí.

### 15.2 Anexo 1

Codigo	Descripción
0	no asignado
8	IN0070-Señal Bloqueo Inversor Motor Principal
9	IN0071-Señal Emergencia Presionada - vea tab. sinóptica máquina
11	IN0086-Control Sobretemperatura Armario
12	IN0042-Señal Paro Ftc Acum. Bot. Salida
15	IN0096-Control Automático Alimentación Ctrl Nivel/Tapa
16	IN0081-Control Presostato No Hay Aire en Red
21	Error Acoplamiento Tapadora 2
22	Temperatura Máx. Cuba Bombas
24	Conteo Botellas Alcanzado
27	Mantenimiento Programado
30	Error Acoplam. Enjuagadora
31	IN0054-Señal Ftc Presencia Botella Caída
32	IN0064-Señal Mesa Acum. Bot. Llena
35	IN0064-Señal Mesa Acum. Bot. Llena
47	Control Seguridades
48	IN0034-Control Disp. Automáticos Intervenidos Grave
49	IN0170-Control Estrella Entrada
50	IN0171-Control Estrella Salida
52	IN0173-Control Rosca Salida
56	IN0191-Control Automático Extractor Humo
57	IN0180-Control Presostato Tratamiento 1
61	IN0195-Control Auto. Motor Ventilador 1 Campana Enjuagadora
64	IN0196-Control Presostato Correa Ventilador 1 Campana Enjuagadora
65	IN0270-Control Estrella Entrada
68	IN0273-Control Rosca Salida
69	IN0277-Control Botella Colgada
70	IN0294-Control Auto. Bomba Vacío Alto
74	IN0280-Control Cuba Alcohol Desalineada
77	Falta Comunicación PLC Depósito
84	IN0250-Control Presostato Atascamiento Filtro 1 Campana
86	PLC Parado
96	IN0241-Control Presostato Correa Ventilad. 2 Campana
97	IN0470-Control Estrella Entrada
107	IN0454-Control Canal Falta Tapas Pieza Final Paro Máquina
112	IN0467-Control Temperatura No Adec. Pistola Pegamen.
113	IN0570-Control Estrella Entrada
116	IN0573-Control Rosca Salida
121	IN0582-Control Temp. Máx Esterilización Tapas
124	IN0577-Control Botella Colgada

Table 8: Alarmas y sus descripciones parte 1

Codigo	Descripción
126	IN0592-Control Auto. Termorresistant. Esterilización Tapas
129	Botella Rota TV
143	Falta Cera en Botellas
144	Faltan Corchos en Botellas
145	Error Codif. Torreta
153	IN0546-Control Guía Entrada acción.
154	IN0560-Control Patín Gas/Vacío Pos. Trab.
155	IN0517-Control Caja Compresión Abierta
158	Alimentador Tapones Vacío
159	Botella Ausente
200	IN0246-Control Guía Entrada Accion.
205	IN0305-Control Botella Explotada Salida
213	IN0186-Control Presostato Soplado Gotas
218	Dbx 27.1
229	IN0950-Control Alarma Ups
237	INA0000-Falta Ref. Velocidad de Inversor Motor Principal
245	IN0484-Control Bloq. Inversor Motor Cabezas
249	Dbx 31.0
259	IN0954-Control Automáticos Circuitos 480V CA Drive
261	Master
262	Enjuagadora
263	Llenadora
264	Tapadora Rosca
268	Rosca
269	Estrellas Enjuagadora
270	Estrellas Llenadora
282	Dbx 35.1
283	INA0322-Control Seguridad Cubeta
285	IN0320-IN0321-Leva Desc. Conos Delanteros Desalin.
286	IN0230-IN0231-Leva Desc. Conos Recipient. Desalin.
287	IN0230-IN0231-Leva Desc. Conos Recipient. Desalin.
292	IN0234-Control Posición Pirómetro
293	INA0227-Avería Sonda Presión Circuito Vacío Bajo
301	IN0377- Ajuste Auto. Anillo Dispositivos
434	INA0228- Avería Sonda Presión Sonda
444	Generic Cause CG
609	Timeout Solicitado
610	Timeout Fase
613	Volante No En Reposo
628	Temperatura No Mantenida

Table 9: Alarmas y sus descripciones parte 2

### 15.3 Anexo 2

Código	Descripción
0	No Asigando
1	IN0083-Selector Panel de Mando Volante Habilitado
2	IN0081-Control Presostato No Hay Aire en Red
8	Elevación en Curso
9	IN0000-Rotación Solicitud
10	IN0040-Señal Ftc Lenta Ninguna Botella Entrada
11	IN0041-Señal Ftc Paro Ninguna Botella Entrada
12	IN0042-Señal Paro Ftc Acum. Bot. Salida
13	Señal Ftc Lenta Acum. Bot. Salida
16	IN0046-Señal Paro de Línea Salida
17	IN0047-Señal Lenta de Línea Salida
20	IN0052-Señal Bloq. Botellas Activad. de Línea
21	Bloq. Botellas Activado por Tecla
23	Bloq. Botellas Activado por Flujo
25	IN0031-Grasa Bloqueada
26	Vel. Lenta Seleccionada
29	IN0002-Auxiliares No Presentes
30	IN0061-Control Presostato No Gas Inerte
31	OUT0031-Mando Disposit. Engrase
36	CIP-Señal de CIP de Alarma
46	Puertas Desbloq.
49	IN0174-Control Protecciones Delanteras Abiertas
50	IN0175-Control Protecciones Traseras Abiertas
64	OUT0204-Lubricación en Curso
65	IN0291-Control Auto. Motor Ajuste Depósito
67	IN0274-Control Protecciones Delanteras Abiertas
68	IN0275-Control Protecciones Traseras Abiertas
73	INA0200-Nivel Máximo Depósito
74	INA0200-Nivel Máximo Depósito
80	Primer Llenado en Curso
81	Modulación Entrada Producto Equivocada
85	IN0252-Control Automático Ventilad. 1 Campana
97	IN0474-Control Protecciones Delanteras Abiertas
98	IN0475-Control Protecciones Traseras Abiertas
99	IN0476-Control Protecciones Laterales Abiertas
113	IN0594-Control Auto. Motor Ajuste Émbolos
115	IN0574-Control Protecciones Delanteras Abiertas
116	IN0575-Control Protecciones Traseras Abiertas
143	Faltan Corchos en Botellas
194	OUT0210-Válvula Entrada Producto Cerrada por Falta
203	OUT0207-Presión Cilindros Demasiado Baja
205	INA0227-Vacio Bajo Demasiado Bajo
206	INA0227-Vacio Bajo Demasiado Alto
230	IN0951-Control Ups Funcionamiento de Batería
237	Señal Falta Solicitud lista de CIP
239	Mando Manual Habilitado
256	Máquina ligeramente Desfasada
257	Máquina Desfasada

Table 10: Avisos y sus descripciones

#### 15.4 Anexo 3

MACRO	DESCRIPTON	RUTINAL
RUN	”Production Slow caused by selecting the button on operator panel”	NO
RUN	”Production at Nominal Speed”	NO
STOP	”Machine stopped with No anomaly and ready to start”	NO
STOP	”Alarm caused by external” cause	NO
STOP	”Machine on alarm”	NO
STOP	”Alarm that are not categorized as previous anomaly”	NO
RUN	”Production slow caused by Infeed conveyor request”	NO
PAUSA	”Anomaly caused by operator action”	NO
STOP	”Machine stopped because of outfeed Full”	NO
PAUSA	”No phase selected”	NO
PAUSA	”Production ending”	SI
PAUSA	”Time used to accesorie the machine for washing (Dummy Bottle etc)”	SI
PAUSA	(DESCRIPCION vacia)	SI
PAUSA	”Production Preparation”	SI
PAUSA	”Machine preparation at the production phase”	SI
STOP	”Machine stopped because of empty inlet”	NO
PAUSA	”Time used for the change over part (changing format, high regulation etc)”	SI
PAUSA	”Washing stopped because of an anomaly”	NO
PAUSA	”Effective washing Time”	SI
PAUSA	”Waiting time for the CIP Unit”	SI
PAUSA	”Waiting for Morning Restart from CIP Unit”	SI

Table 11: Descripciones status máquina

## 15.5 Anexo 4

### Método de Redes Neuronales Recurrentes (RNN)

Una red neuronal recurrente (RNN) es un tipo de modelo de redes neuronales diseñado para trabajar con datos secuenciales o series temporales en sistemas dinámicos no lineales. Estas redes utilizan datos de entrenamiento junto con técnicas algorítmicas de gradiente descendente para aprender patrones, ya que utilizan información previa para generar respuestas. En otras palabras, las RNN tienen elementos de entrada y salida que están conectados a través de conexiones recurrentes. (Bonet, I., Et. Al., 2007)

Lo que hace que las RNN sean especialmente poderosas es su capacidad para recordar información pasada gracias a su memoria interna, lo que les permite generar el resto de una secuencia incluso cuando no la conocen de manera completa. En el contexto específico de las RNN, cada neurona puede estar conectada con cualquier otra, y estas conexiones recurrentes pueden formar ciclos en el grafo definido por las interconexiones de las unidades de procesamiento. Esto les permite descubrir patrones temporales, como estacionalidades, en las secuencias que procesan. (Pérez, J., 2021)

Cuando se aplica este enfoque a un problema como el de una línea de embotellado, se vuelve especialmente relevante debido a su capacidad para lidar con la no linealidad inherente al proceso de producción, su adaptabilidad para ajustarse a cambios en la operación y su capacidad para tolerar fallas sin interrupciones significativas. Esto se justifica aún más cuando se dispone de datos históricos secuenciales, como registros de avisos y alarmas a lo largo del tiempo, y un conjunto de datos relativamente grande recopilado durante un período de 20 meses permitiendo así estudiar profundamente la dimensionalidad del tiempo.

Según se muestra en la figura 1, existen 4 neuronas interconectadas a lo largo del tiempo, cada una de estas la entrada del nodo anterior, como también, su propia salida del instante de tiempo anterior para generar una salida. Llevándolo al problema de línea de embotellado, se utilizan los datos secuenciales a través del tiempo con las entradas que representarían la información recopilada de avisos y alarmas, las conexiones recurrentes son la retroalimentación y respuestas de salida a través del procesamiento de la información, los cuales actualizan el estado interno de la red para así, una vez que ha sido entrenado, pueda capturar patrones, estacionalidades y anomalías y así predecir procesos importantes de la línea. (Torres, J., 2019)

Sin embargo, es importante destacar que las RNN pueden ser computacionalmente intensivas, lo que significa que requieren recursos significativos de procesamiento y memoria para entrenarse y operar de manera eficiente, además que pueden ser afectadas por los “Exploding Gradients” o “Vanishing gradients” donde los valores de los gradientes para el entrenamiento es excesivamente alto (pendiente) por lo que se produce un error y detención en el proceso de aprendizaje. Dada la gran cantidad de datos con la que se cuenta es

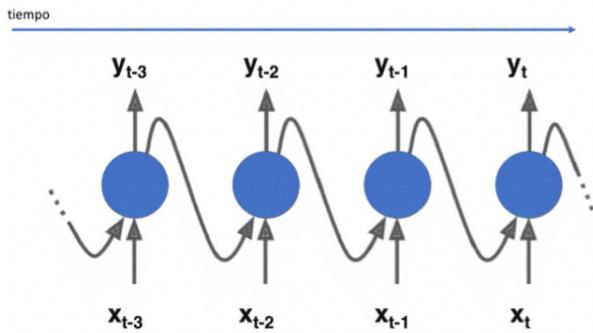


Figure 25: Redes Neuronales Recurrentes en el tiempo. Fuente: Torres, J (2019).

importante evaluar el impacto que estos proceso pueden traer a largo plazo a la eficacia del método.

Como antecedente que respalda y otorga valor a la utilización de esta metodología en el mantenimiento predictivo, se han encontrado estudios y tópicos que abordan la misma problemática en otros ámbitos, como por ejemplo, en la industria de la aviación. A partir del análisis lograron llegar a la siguiente conclusión, “Nuestros mejores modelos han sido capaces de detectar entre un 45-55% de los casos anteriormente mencionados correctamente. Esta cifra debe contrastarse con el hecho de que no podemos asegurar que todos los fallos dentro de nuestro conjunto de datos tengan un origen relacionado con los factores climatológicos estudiados. Por tanto, se ha cumplido el objetivo: demostrar el potencial de predicción de estos datos aplicando Deep Learning.” (López, J.M., 2021). Esto valida justamente que para el estudio de mantenimientos predictivos logra ser una herramienta de gran utilidad para abordarlo por Deep Learning mediante RNN.

## Métodos de Suavización Exponencial

La suavización exponencial es un método de predicción propuesto durante 1950. En este caso las predicciones se llevan a cabo mediante una ponderación exponencial que decae a medida que las observaciones son más antiguas. De esta forma, los datos más relevantes para generar una mejor predicción son los más actuales. La mayor ventaja que tiene SES es que genera rápidamente un pronóstico confiable, lo cual permite tomar, a su vez, decisiones tempranas. (Hyndman, R.J., & Athanasopoulos, G., 2018)

### Método de Suavización Exponencial Simple (SES)

El método de suavización simple se utiliza para series de tiempo sin tendencias evidentes ni estacionalidades periódicas. Se basa en un único parámetro de suavizado llamado ”nivel” y pondera los datos históricos de manera exponencial para lograr alta precisión (Vandeput, 2021).

En el contexto de mantenimiento en la línea de embotellado, este método tiene ventajas y desventajas. Entre sus ventajas se encuentra su amplia utilización en el pronóstico de la demanda debido a su sencillez y eficiencia computacional, lo que permite alertas tempranas para reducir el tiempo de inactividad. También destaca su capacidad de actualización en tiempo real.

Sin embargo, presenta limitaciones cuando se trata de series temporales con patrones complejos o cambios abruptos en los datos, como tendencias no lineales. Este problema se agrava cuando se manejan conjuntos de datos grandes. Además, asume que los datos históricos son estables, lo cual no siempre es cierto en la realidad, y no tiene en cuenta factores externos que pueden afectar el mantenimiento. La elección correcta del valor inicial de suavizado es crucial.

En resumen, este método puede ser efectivo como enfoque inicial para los datos, pero no es sostenible a largo plazo debido a su simplicidad y su incapacidad para identificar las causas subyacentes de las fallas, lo que es esencial para mejorar los resultados con el tiempo.

### Método Holt-Winters

El método Holt-Winters es una adaptación mejorada de la suavización exponencial simple (SES), especialmente efectiva para series de tiempo más extensas y pronósticos a largo plazo. Matemáticamente, emplea una ecuación de pronóstico y tres ecuaciones de suavizado relacionadas con nivel, tendencia y estacionalidad. Cuando las variaciones estacionales son constantes, se considera un método aditivo; cuando cambian proporcionalmente, se usa el método multiplicativo (Hyndman y Athanasopoulos, 2018).

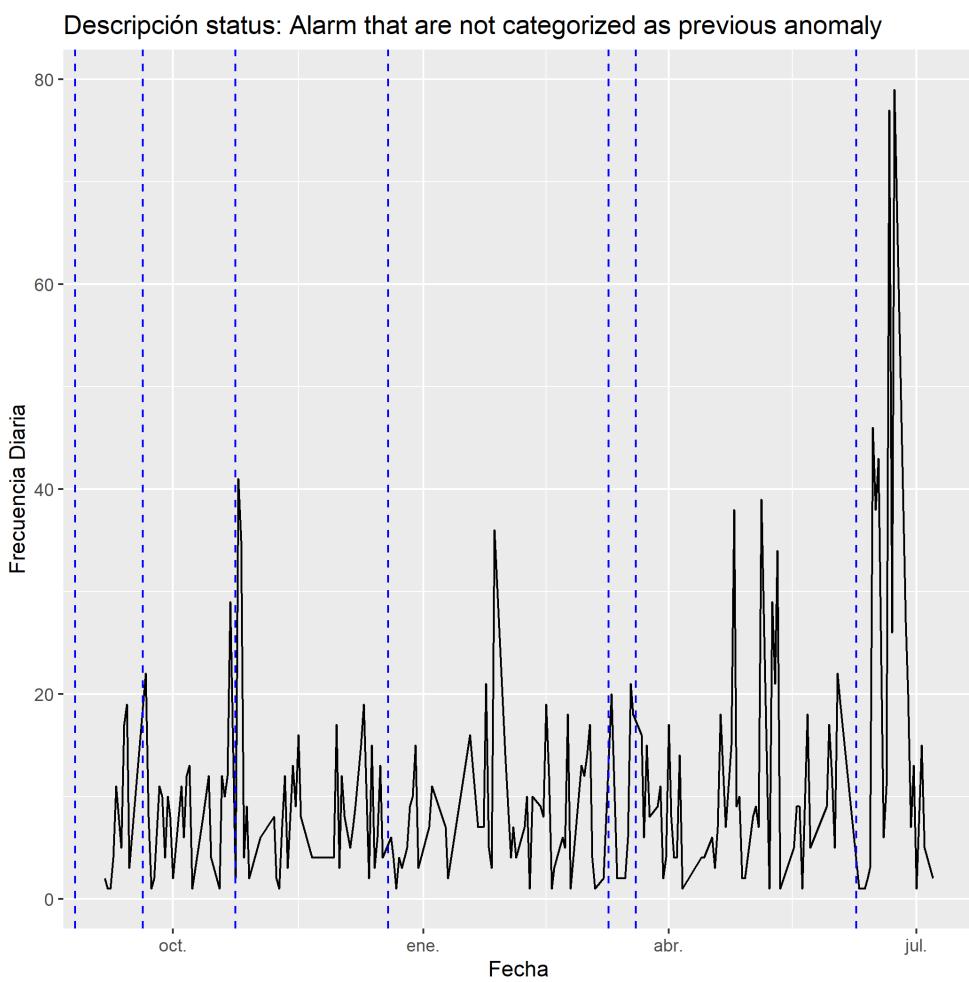
En comparación con SES, Holt-Winters es más complejo y versátil. Proporciona una visión más profunda de las tendencias a largo plazo y permite la elaboración de pronósticos a corto, mediano y largo plazo. Puede adaptarse a diversos tipos de datos, incluyendo datos categóricos secuenciales con estacionalidades marcadas. Sin embargo, como limitaciones, al igual que SES, no maneja cambios abruptos en los datos y a veces requiere ajustes manuales de parámetros. Además, no considera factores externos, que son relevantes en situaciones con múltiples variables influyentes (Mira, 2018).

## **Fast Fourier Transform**

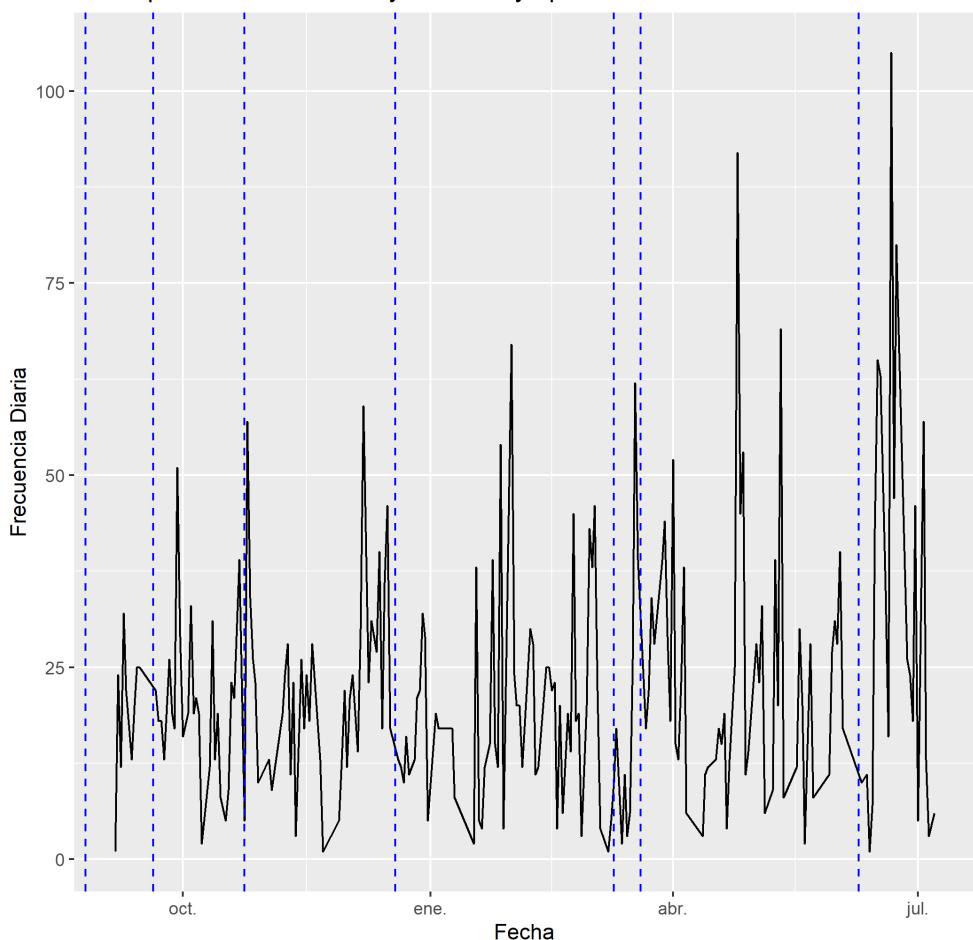
La Transformada Rápida de Fourier (FFT) es una técnica que descompone señales en sus componentes de frecuencia, proporcionando información sobre amplitud y fase. Se convierte una señal de dominio de tiempo en una representación en el dominio de la frecuencia, revelando las frecuencias presentes y su energía, denominado "espectro de frecuencia" (Rao, Kim, Hwang, 2010). Esto es útil para identificar patrones y tomar decisiones en aplicaciones de predicción y procesamiento de datos.

Es esencial entender que la FFT se emplea para analizar series de tiempo y señales, descomponiéndolas en componentes de frecuencia. Aunque no predice, extrae información valiosa de las series temporales, siendo útil en el pronóstico de fallas. En la línea de embotellado, ayuda a identificar patrones de frecuencia, detectar ciclos o patrones estacionales en los datos de fallas y proporciona información sobre eventos repetitivos. Sin embargo, tiene limitaciones en la detección de patrones no lineales y depende de la calidad de los datos, siendo menos confiable con datos incompletos.

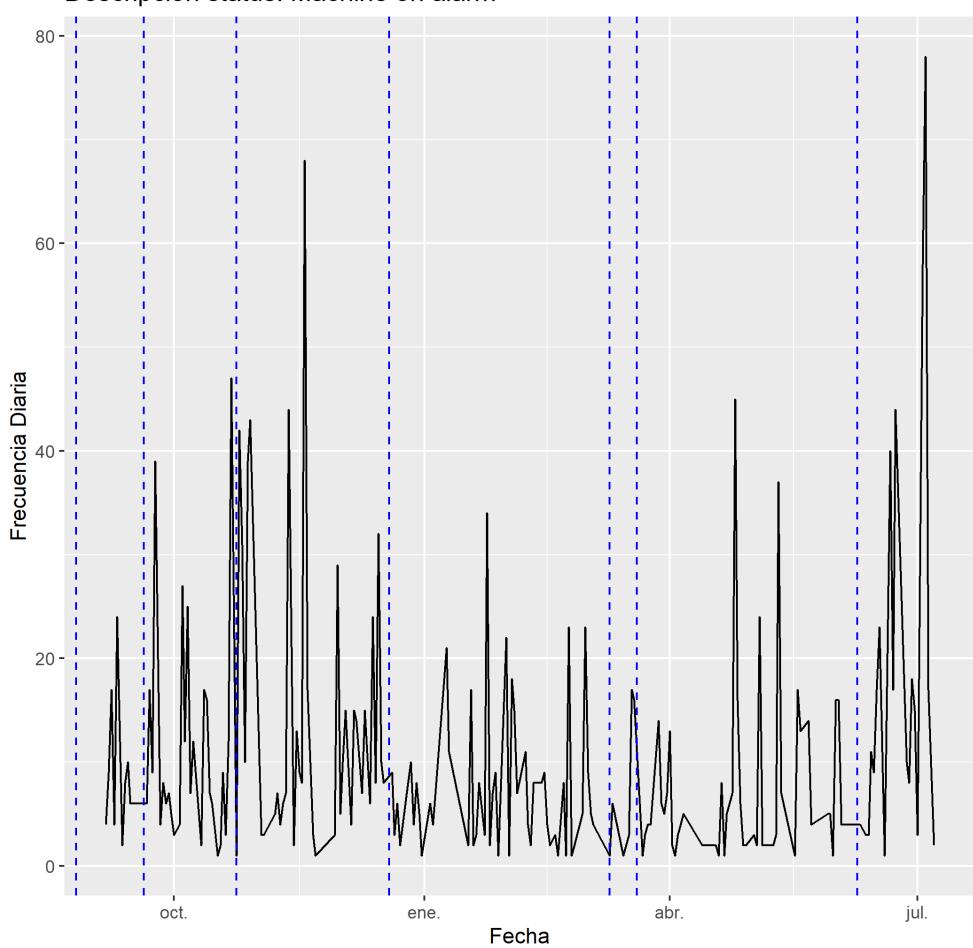
## 15.6 Anexo 5 Gráficos frecuencia status



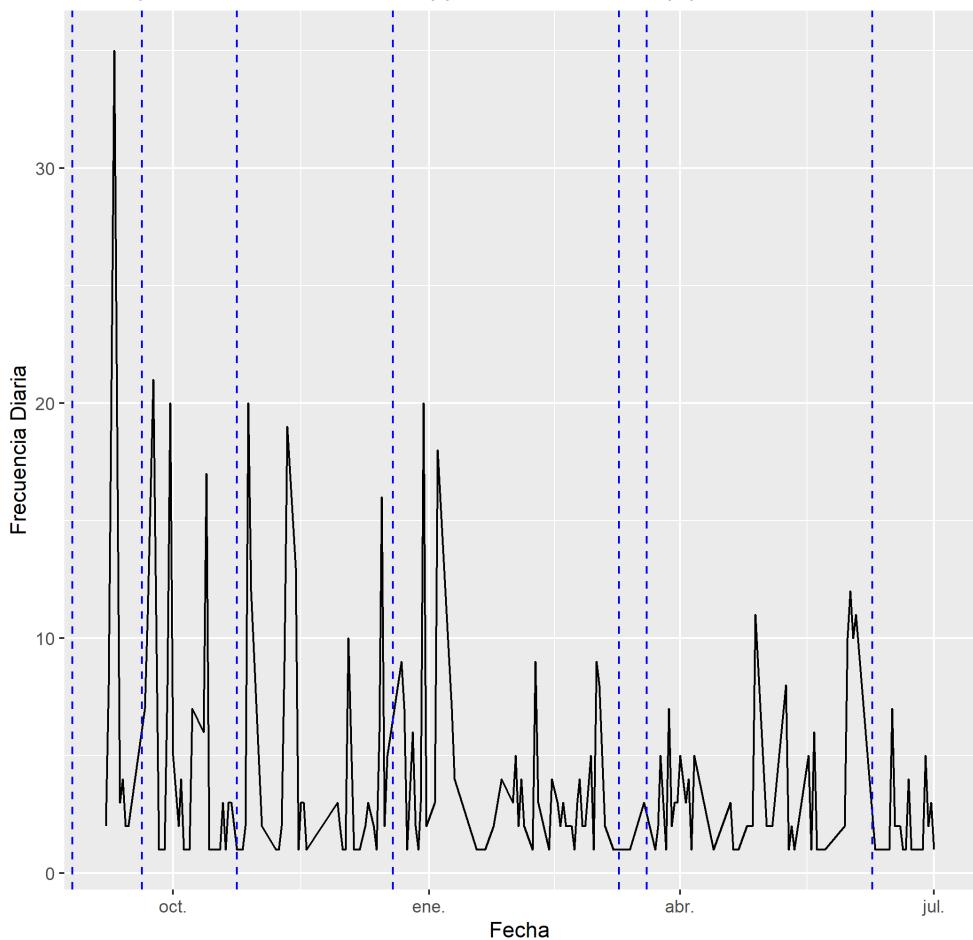
Descripción status: Anomaly caused by operator action



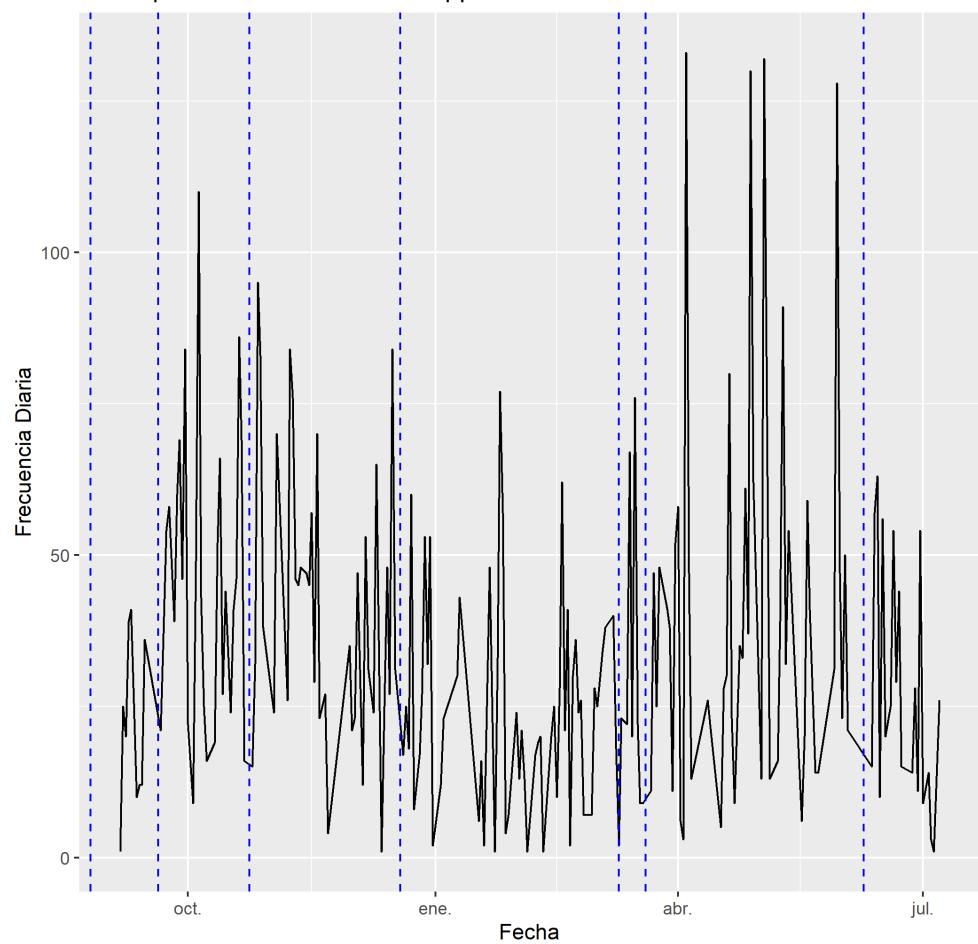
Descripción status: Machine on alarm



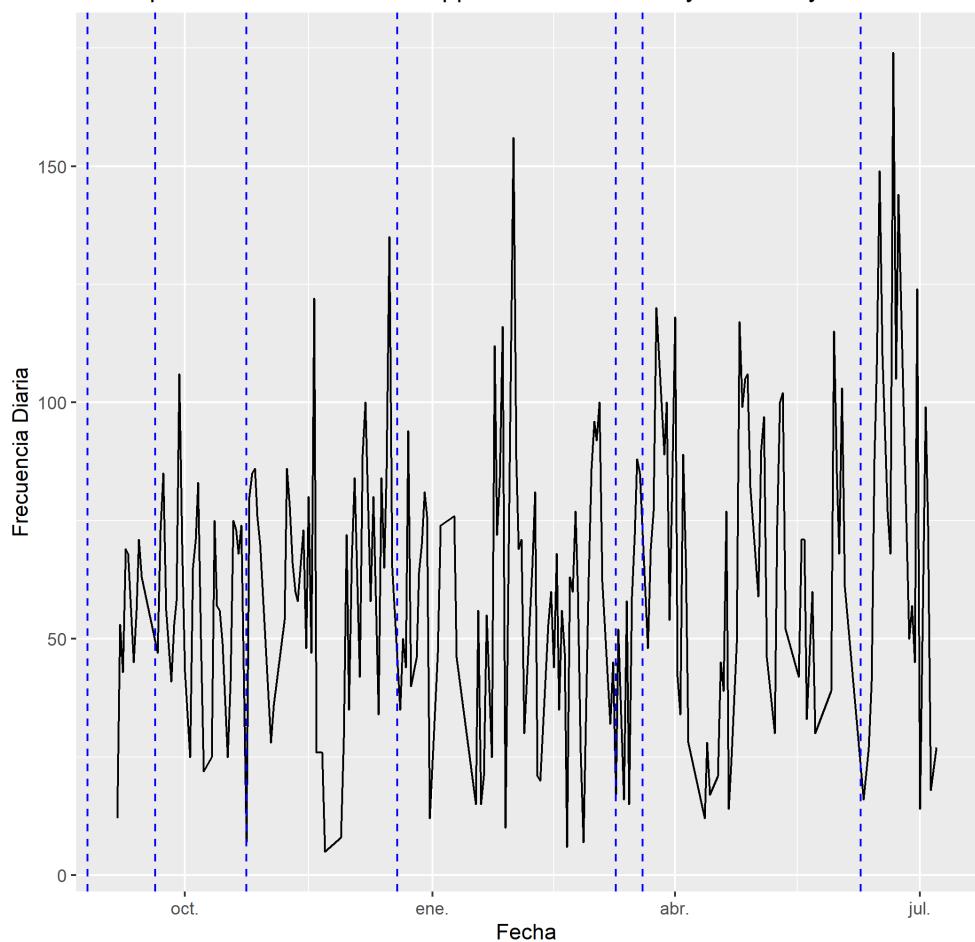
Descripción status: Machine stopped because of empty inlet



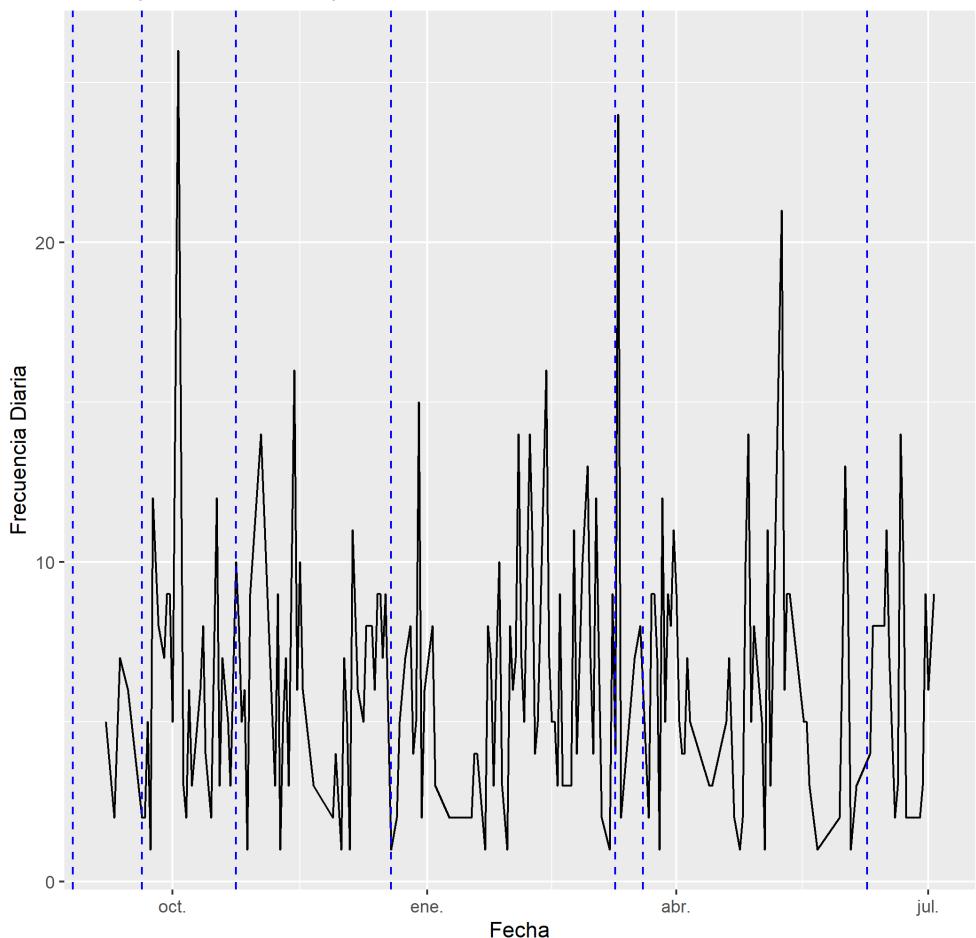
Descripción status: Machine stopped because of outfeed Full



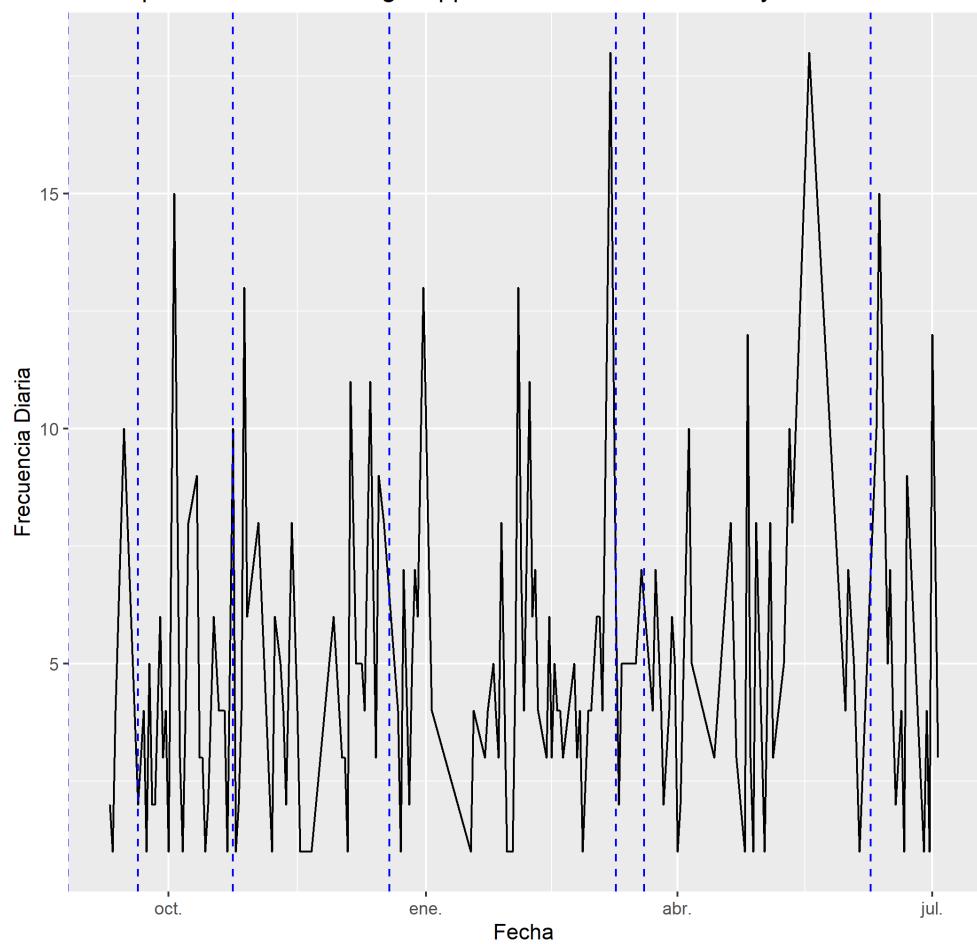
Descripción status: Machine stopped with No anomaly and ready to start



Descripción status: No phase selected



Descripción status: Washing stopped because of an anomaly



## 15.7 Anexo 6 Correlación de series de tiempo para las fallas

### 15.7.1 Fallas y Series Cluster 1

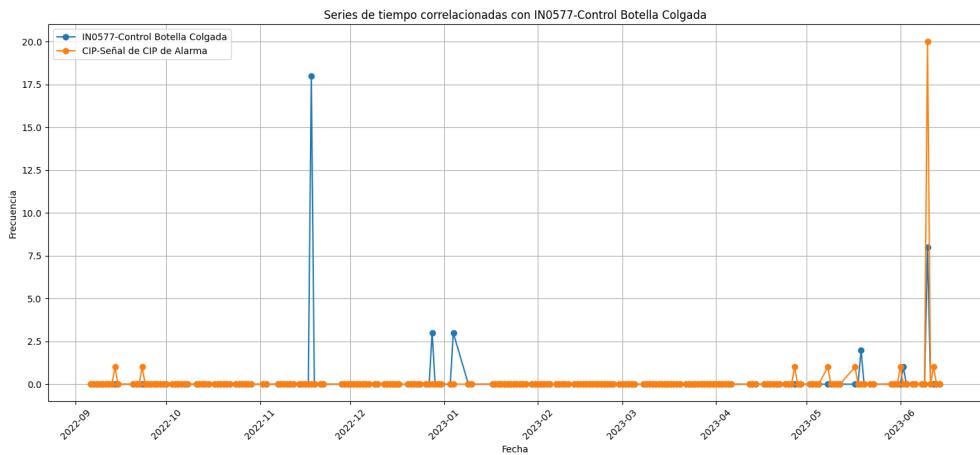


Figure 26: Falla IN0577-Control Botella Colgada y sus series de tiempo más correlacionadas

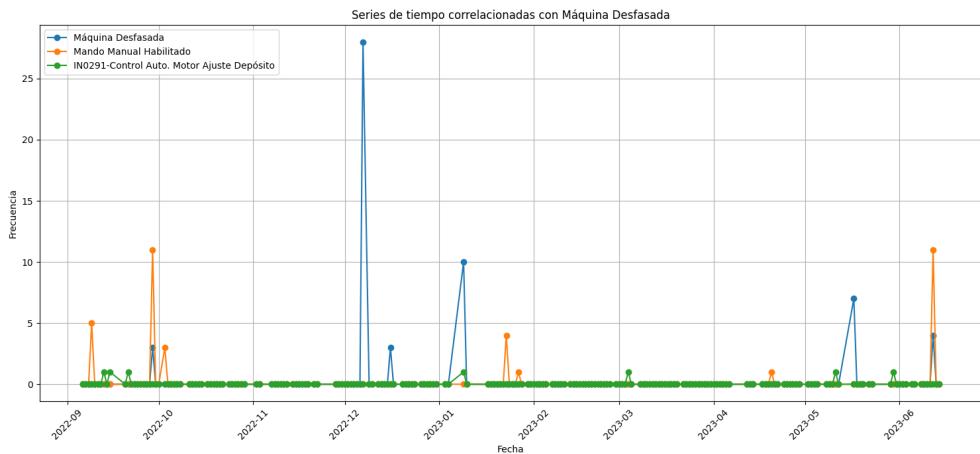


Figure 27: Falla Máquina Desfasada y sus series de tiempo más correlacionadas

### 15.7.2 Fallas y Series Cluster 2

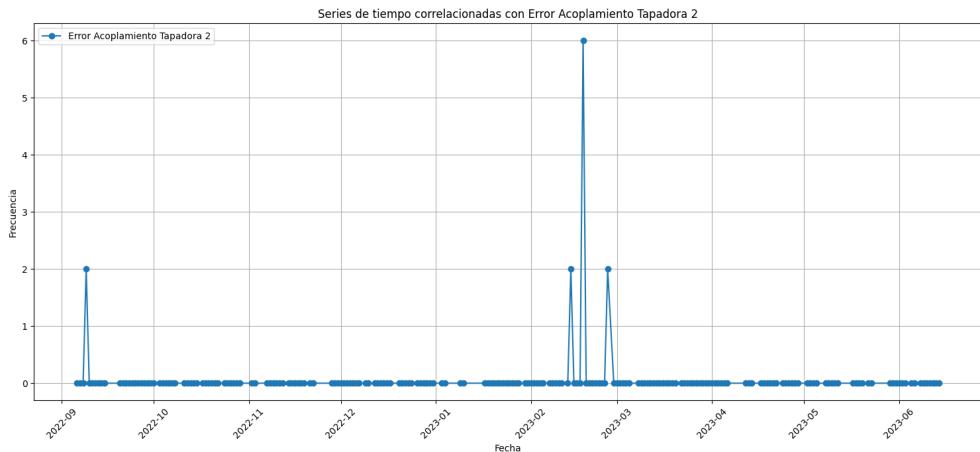


Figure 28: Falla Error Acoplamiento Tapadora 2 y sus series de tiempo más correlacionadas

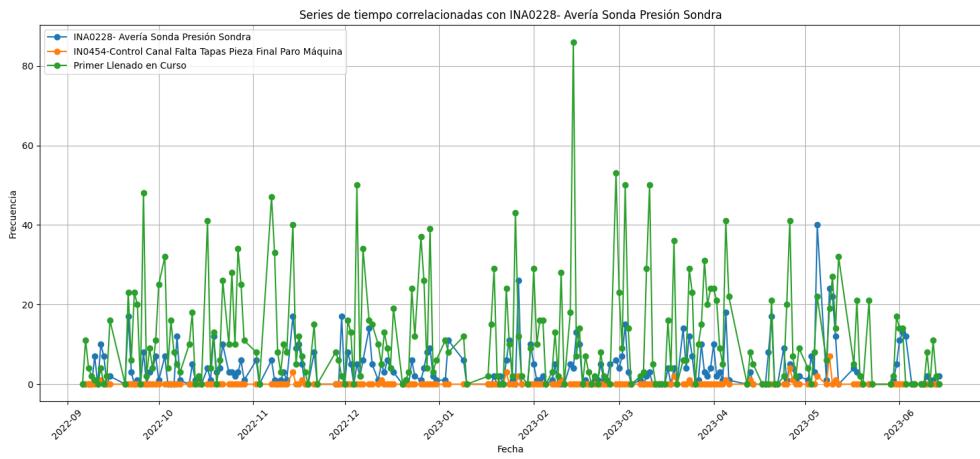


Figure 29: Falla INA0228- Avería Sonda Presión Sondra y sus series de tiempo más correlacionadas

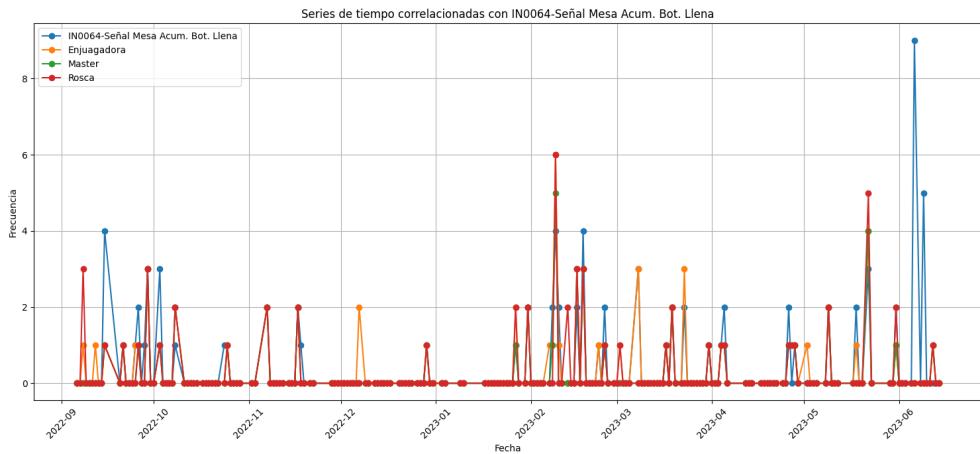


Figure 30: Falla IN0064-Señal Mesa Acum. Bot. Llena y sus series de tiempo más correlacionadas

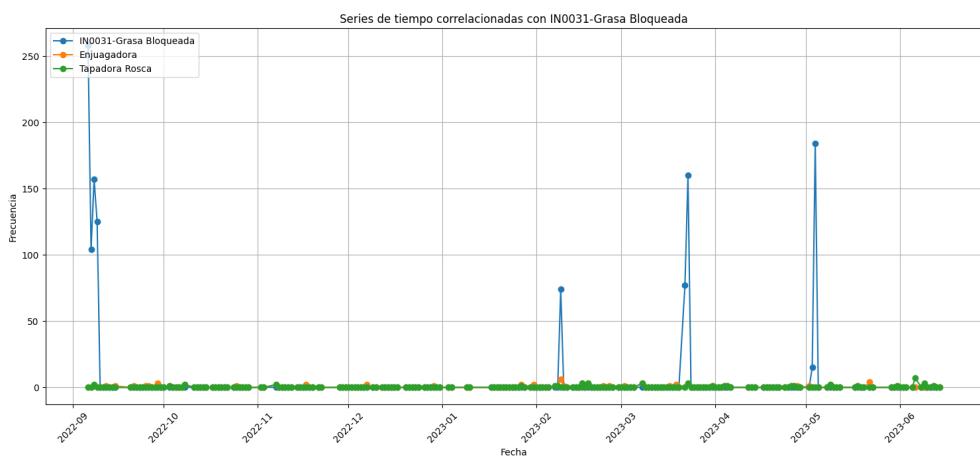


Figure 31: Falla IN0031-Grasa Bloqueada y sus series de tiempo más correlacionadas

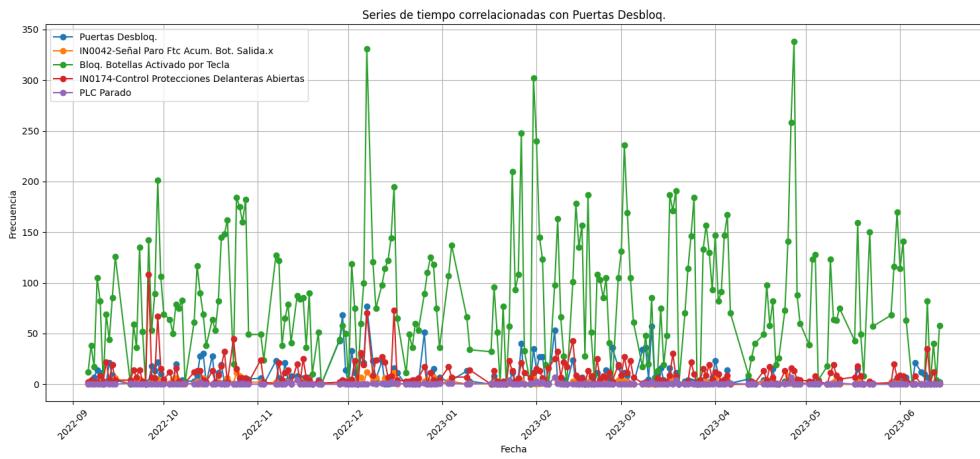


Figure 32: Falla Puertas Desbloq. y sus series de tiempo más correlacionadas

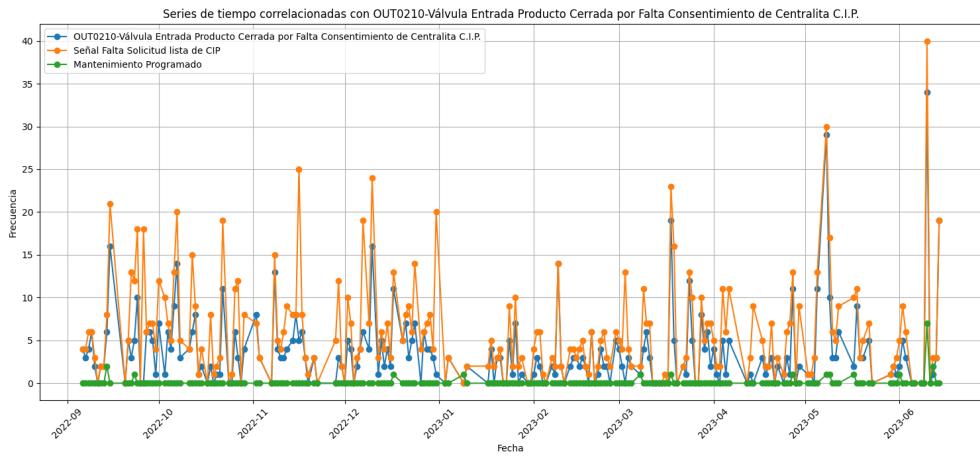


Figure 33: Falla OUT0210-Válvula Entrada Producto Cerrada por Falta Consentimiento de Centralita C.I.P. y sus series de tiempo más correlacionadas

### 15.7.3 Fallas y Series Cluster 3

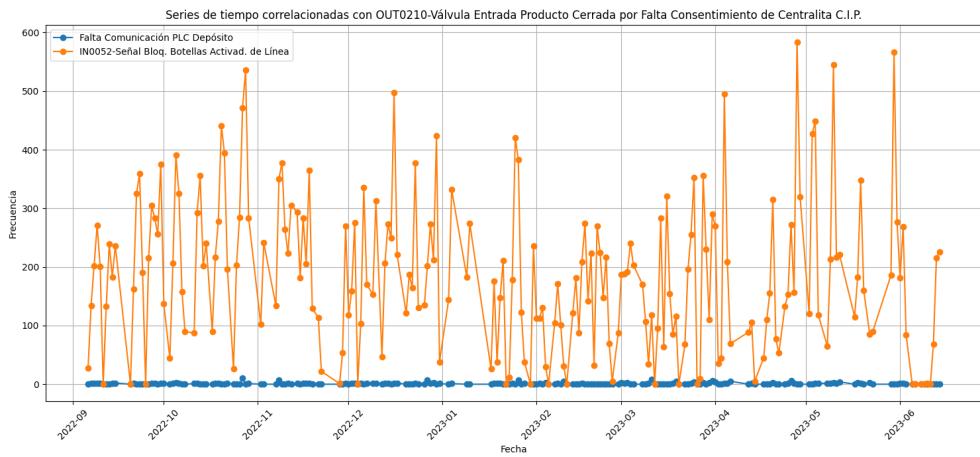


Figure 34: Falla Falta Comunicación PLC Depósito y sus series de tiempo más correlacionadas

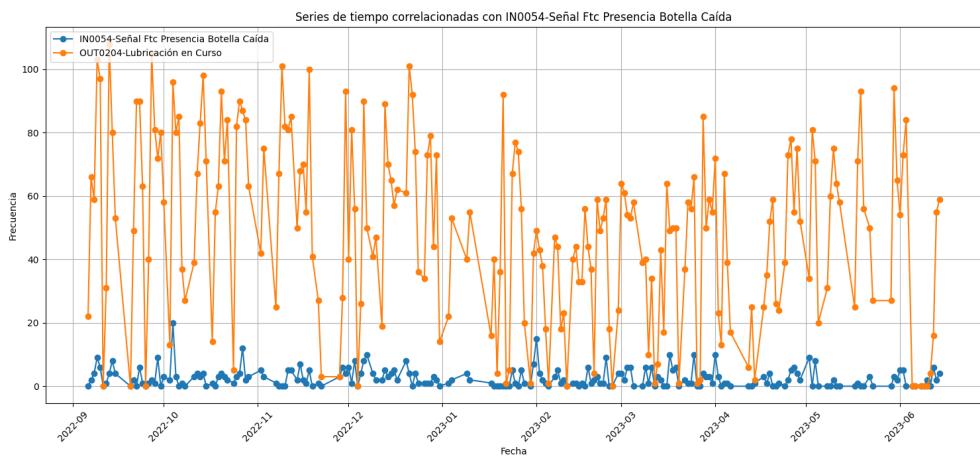


Figure 35: Falla IN0054-Señal Ftc Presencia Botella Caída y sus series de tiempo más correlacionadas

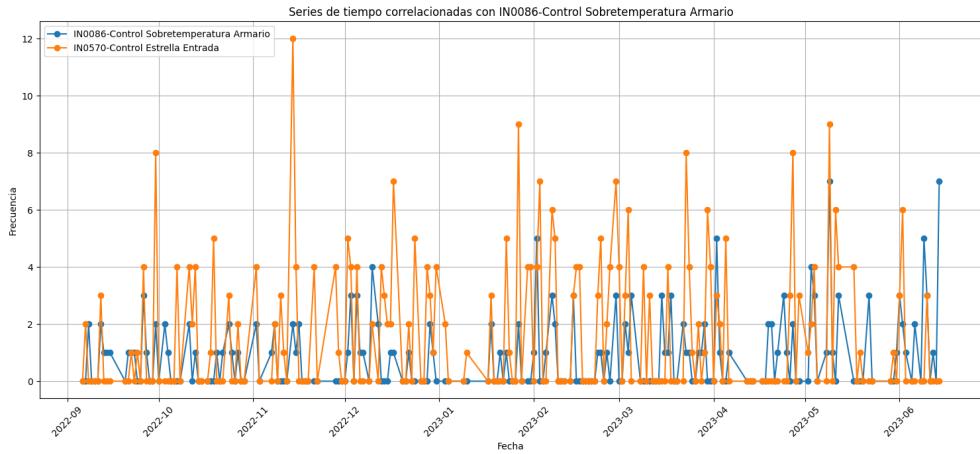


Figure 36: Falla IN0086-Control Sobretemperatura Armario y sus series de tiempo más correlacionadas

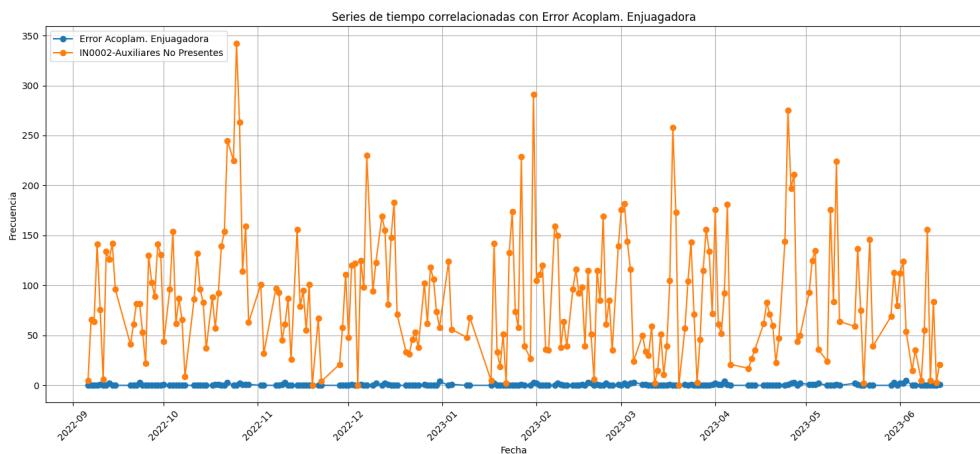


Figure 37: Falla Error Acoplam. Enjuagadora y sus series de tiempo más correlacionadas

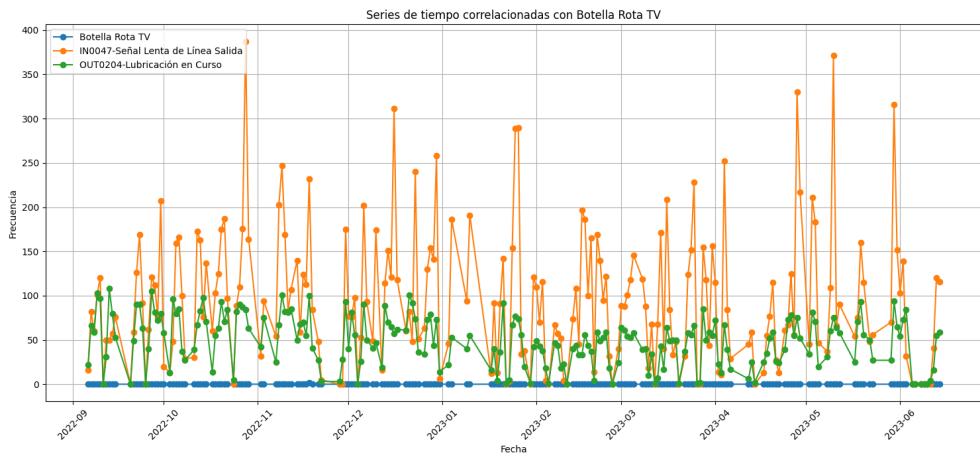


Figure 38: Falla Botella Rota TV y sus series de tiempo más correlacionadas

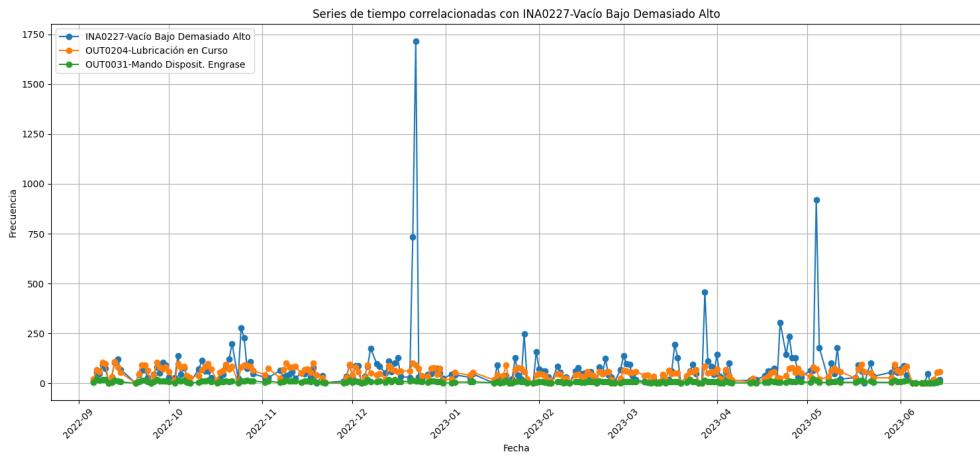


Figure 39: Falla INA0227-Vacio Bajo Demasiado Alto y sus series de tiempo más correlacionadas

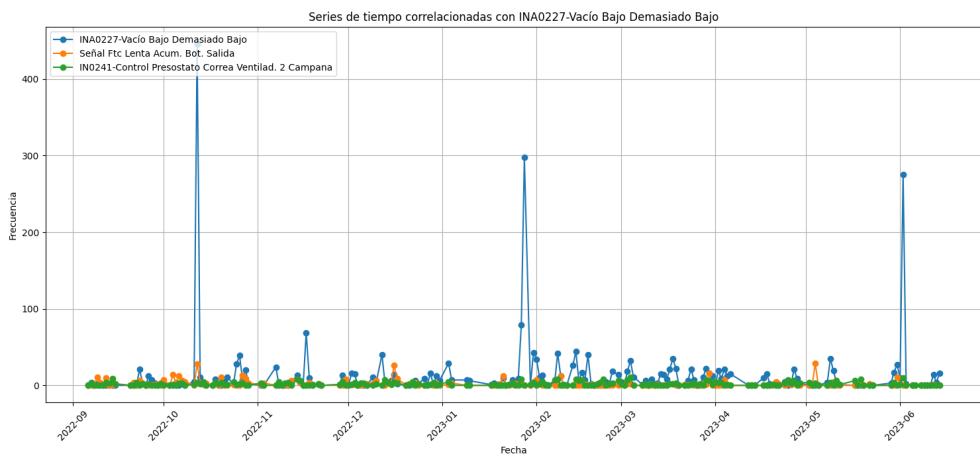


Figure 40: Falla INA0227-Vacio Bajo Demasiado Bajo y sus series de tiempo más correlacionadas

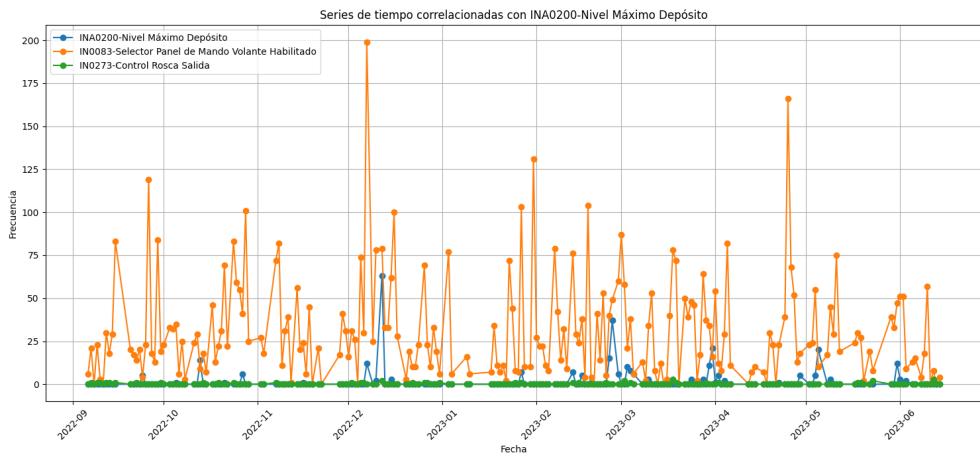


Figure 41: Falla INA0200-Nivel Máximo Depósito y sus series de tiempo más correlacionadas

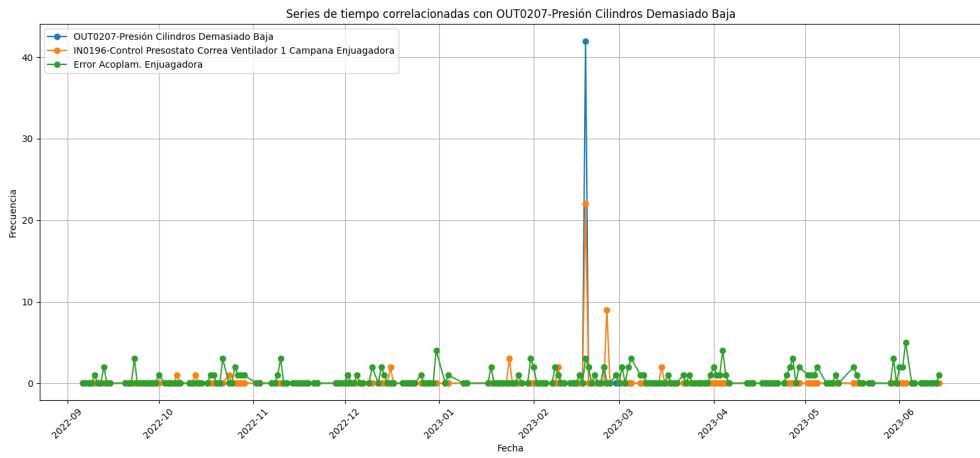


Figure 42: Falla OUT0207-Presión Cilindros Demasiado Baja y sus series de tiempo más correlacionadas

## 15.8 Anexo 7 Listado Clusters

<b>Tipo</b>	<b>Nombre</b>
Alarma	IN0560-Control Patín Gas/Vacío Pos. Trab.
Aviso	IN0000-Rotación Solicitado
Aviso	IN0175-Control Protecciones Traseras Abiertas
Aviso	IN0252-Control Automático Ventilad. 1 Campana
Aviso	IN0951-Control Ups Funcionamiento de Batería

Table 12: Eventos en Cluster 0

<b>Tipo</b>	<b>Nombre</b>
Alarma	IN0470-Control Estrella Entrada
Alarma	IN0517-Control Caja Compresión Abierta
Alarma	IN0070-Señal Bloqueo Inversor Motor Principal
Alarma	Estrellas Enjuagadora
Aviso	Faltan Corchos en Botellas.x
Alarma	IN0250-Control Presostato Atascamiento Filtro 1 Campana
Falla	IN0577-Control Botella Colgada
Alarma	IN0294-Control Auto. Bomba Vacío Alto
Aviso	IN0574-Control Protecciones Delanteras Abiertas
Aviso	IN0575-Control Protecciones Traseras Abiertas
Aviso	Mando Manual Habilitado
Aviso	IN0291-Control Auto. Motor Ajuste Depósito
Aviso	CIP-Señal de CIP de Alarma
Falla	Máquina Desfasada
Aviso	IN0476-Control Protecciones Laterales Abiertas

Table 13: Eventos en Cluster 1

<b>Tipo</b>	<b>Nombre</b>
Alarma	IN0171-Control Estrella Salida
Alarma	Falta Cera en Botellas
Alarma	IN0071-Señal Emergencia Presionada - vea tab. sinóptica máquina para detalles
Alarma	Volante No En Reposo
Alarma	IN0467-Control Temperatura No Adec. Pistola Pegamen.
Alarma	IN0582-Control Temp. Máx Esterilización Tapas
Alarma	IN0573-Control Rosca Salida
Aviso	IN0042-Señal Paro Ftc Acum. Bot. Salida.x
Alarma	IN0280-Control Cuba Alcohol Desalineada
Alarma	Botella Ausente
Alarma	Rosca
Alarma	Tapadora Rosca
Alarma	Enjuagadora
Falla	Error Acoplamiento Tapadora 2
Falla	INA0000-Falta Ref. Velocidad de Inversor Motor Principal
Falla	INA0228- Avería Sonda Presión Sondra
Alarma	Error Codif. Torreta
Alarma	IN0195-Control Auto. Motor Ventilador 1 Campana Enjuagadora
Alarma	IN0246-Control Guía Entrada Accion.
Alarma	PLC Parado
Alarma	IN0454-Control Canal Falta Tapas Pieza Final Paro Máquina
Alarma	IN0173-Control Rosca Salida
Alarma	Dbx 31.0
Alarma	Mantenimiento Programado
Alarma	Conteo Botellas Alcanzado
Falla	IN0064-Señal Mesa Acum. Bot. Llena
Alarma	Llenadora
Alarma	Master
Alarma	Temperatura Máx. Cuba Bombas
Alarma	IN0096-Control Automático Alimentación Ctrl Nivel/Tapa
Alarma	INA0322-Control Seguridad Cubeta
Alarma	IN0950-Control Alarma Ups
Alarma	IN0277-Control Botella Colgada
Alarma	IN0170-Control Estrella Entrada
Alarma	IN0592-Control Auto. Termorresistent. Esterilización Tapas
Alarma	Timeout Fase
Alarma	IN0270-Control Estrella Entrada
Alarma	Temperatura No Mantenida
Alarma	Timeout Solicitado
Alarma	IN0546-Control Guía Entrada accion.
Alarma	Control Seguridades
Alarma	IN0230-IN0231-Leva Desc. Conos Recipient. Desalin.
Alarma	NA
Alarma	IN0232-IN0233-Cubata Desalineada
Falla	IN0031-Grasa Bloqueada
Aviso	Modulación Entrada Producto Equivocada
Aviso	Bloq. Botellas Activado por Tecla
Falla	Puertas Desbloq.
Aviso	Señal Falta Solicitud lista de CIP
Falla	OUT0210-Válvula Entrada Producto Cerrada por Falta Consentimiento de Centralita C.I.P.
Aviso	IN0174-Control Protecciones Delanteras Abiertas
Alarma	Faltan Corchos en Botellas.y
Aviso	Primer Llenado en Curso
Aviso	Bloq. Botellas Activado por Flujo
Aviso	IN0041-Señal Ftc Paro Ninguna Botella Entrada
Aviso	IN0274-Control Protecciones Delanteras Abiertas
Aviso	IN0275-Control Protecciones Traseras Abiertas

Table 14: Eventos en cluster 2

<b>Tipo</b>	<b>Nombre</b>
Aviso	IN0081-Control Presostato No Hay Aire en Red.x
Alarma	IN0191-Control Automático Extractor Humo
Alarma	Dbx 27.1
Falla	Falta Comunicación PLC Depósito
Alarma	IN0273-Control Rosca Salida
Alarma	IN0241-Control Presostato Correa Ventilad. 2 Campana
Alarma	IN0570-Control Estrella Entrada
Falla	IN0054-Señal Ftc Presencia Botella Caída
Alarma	Estrellas Llenadora
Falla	IN0086-Control Sobretemperatura Armario
Alarma	Alimentador Tapones Vacío
Alarma	IN0090-Control Automático Motor Cinta
Alarma	Dbx 35.1
Falla	Error Acoplam. Enjuagadora
Falla	INA0227-Avería Sonda Presión Circuito Vacío Bajo
Alarma	IN0034-Control Disp. Automáticos Intervenidos Grave
Alarma	IN0180-Control Presostato Tratamiento 1
Alarma	IN0186-Control Presostato Soplado Gotas
Alarma	IN0377- Ajuste Auto. Anillo Dispositivos
Alarma	IN0320-IN0321-Leva Desc. Conos Delanteros Desalin.
Alarma	IN0305-Control Botella Explotada Salida
Alarma	IN0196-Control Presostato Correa Ventilador 1 Campana Enjuagadora
Alarma	IN0234-Control Posición Pirómetro
Alarma	IN0484-Control Bloq. Inversor Motor Cabezas (
Falla	Botella Rota TV
Alarma	IN0954-Control Automáticos Circuitos 480V CA Drive
Aviso	OUT0031-Mando Disposit. Engrase
Aviso	IN0052-Señal Bloq. Botellas Activad. de Línea
Aviso	OUT0204-Lubricación en Curso
Aviso	Vel. Lenta Seleccionada
Falla	INA0227-Vacio Bajo Demasiado Alto
Aviso	IN0002-Auxiliares No Presentes
Aviso	IN0047-Señal Lenta de Línea Salida
Aviso	IN0083-Selector Panel de Mando Volante Habilitado
Aviso	IN0474-Control Protecciones Delanteras Abiertas
Aviso	IN0046-Señal Paro de Línea Salida
Aviso	IN0040-Señal Ftc Lenta Ninguna Botella Entrada
Falla	INA0200-Nivel Mínimo Depósito
Alarma	INA0227-Vacio Bajo Demasiado Bajo
Alarma	IN0042-Señal Paro Ftc Acum. Bot. Salida.y
Aviso	IN0475-Control Protecciones Traseras Abiertas
Aviso	Señal Ftc Lenta Acum. Bot. Salida
Aviso	Elevación en Curso
Aviso	IN0081-Control Presostato No Hay Aire en Red.y
Aviso	IN0061-Control Presostato No Gas Inerte
Aviso	INA0200-Nivel Máximo Depósito
Aviso	Máquina ligeramente Desfasada
Aviso	IN0594-Control Auto. Motor Ajuste Émbolos
Falla	OUT0207-Presión Cilindros Demasiado Baja

Table 15: Eventos en cluster 3

## 15.9 Anexo 8 Gráficas Clusters

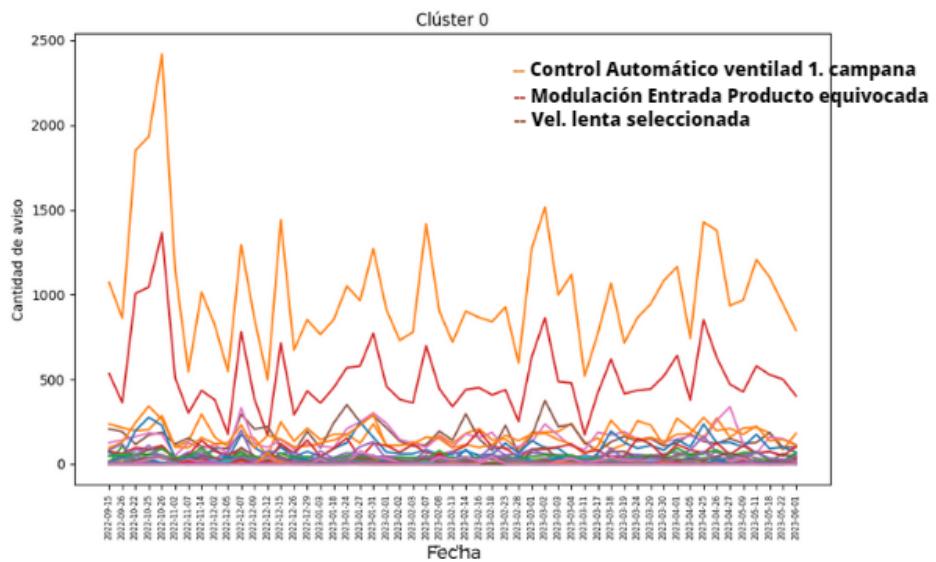


Figure 43: Cluster 0 Avisos. Fuente: Elaboración Propia

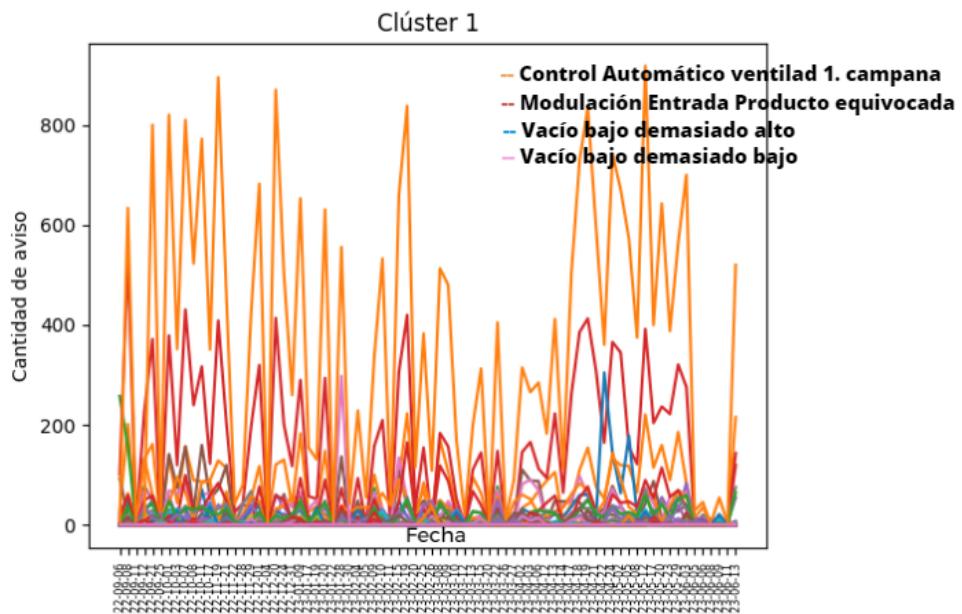


Figure 44: Cluster 1 Avisos. Fuente: Elaboración Propia

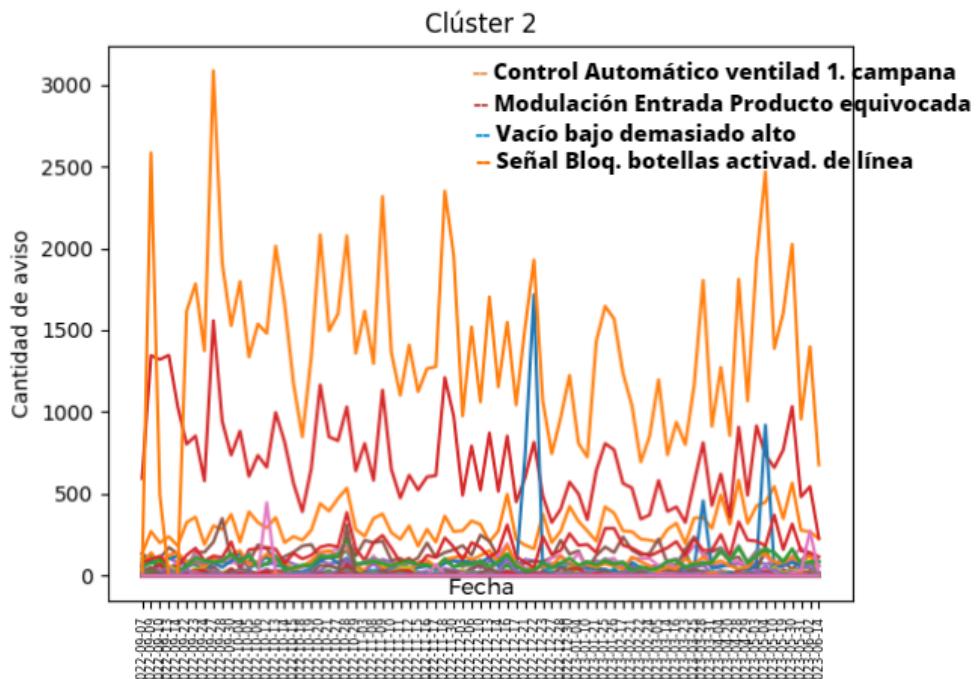


Figure 45: Cluster 2 Avisos. Fuente: Elaboración Propia

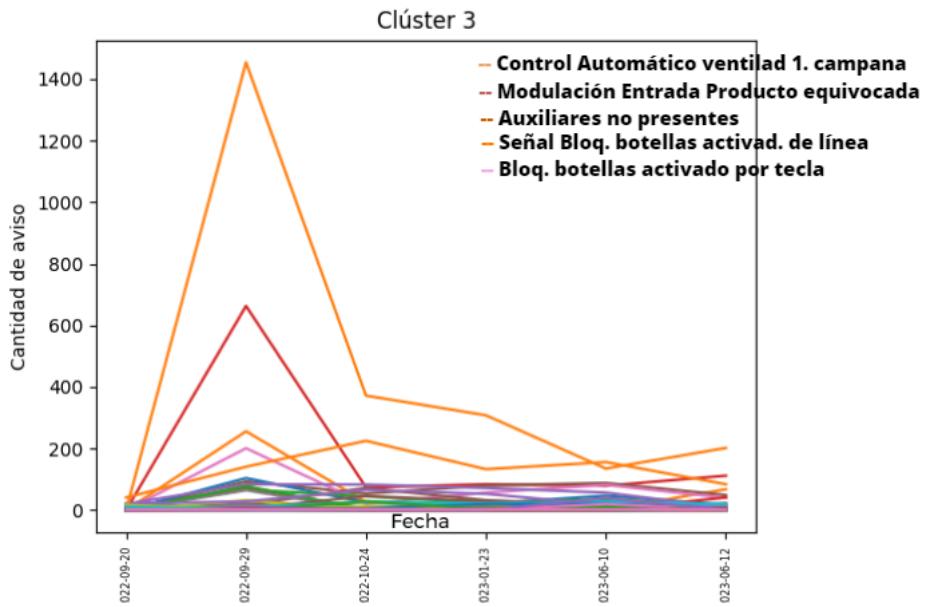


Figure 46: Cluster 3 Avisos. Fuente: Elaboración Propia

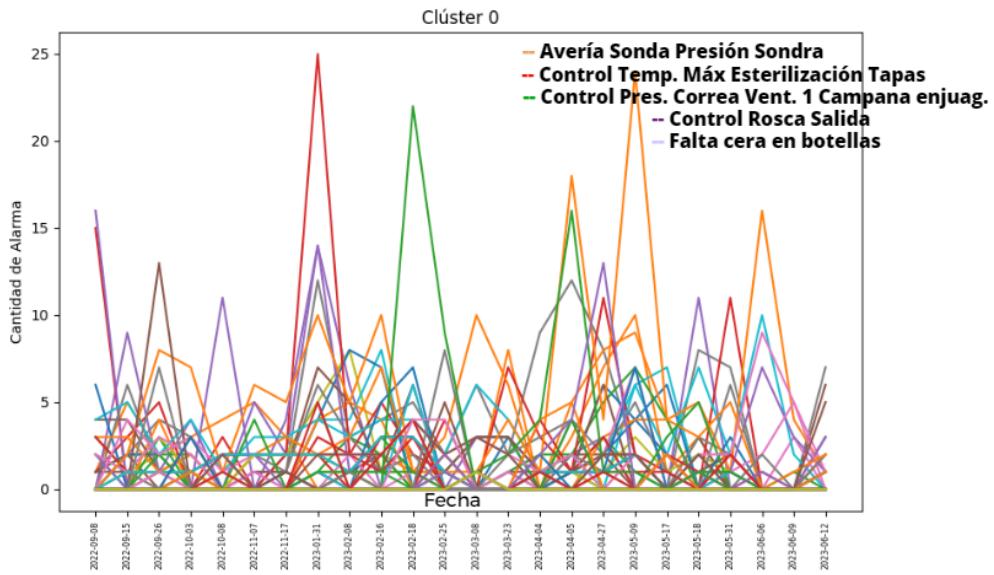


Figure 47: Cluster 0 Alarmas. Fuente: Elaboración Propia

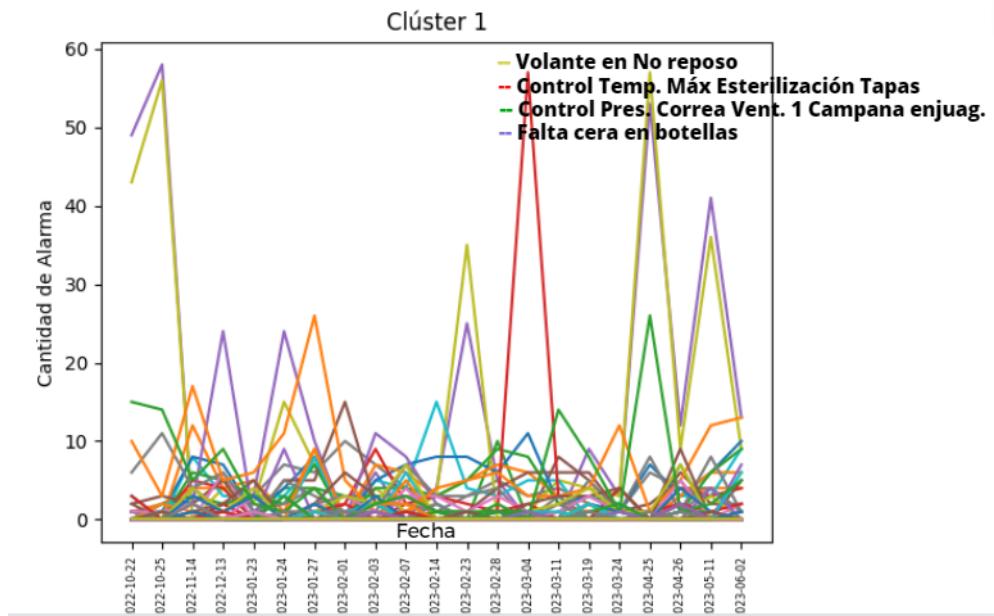


Figure 48: Cluster 1 Alarmas. Fuente: Elaboración Propia

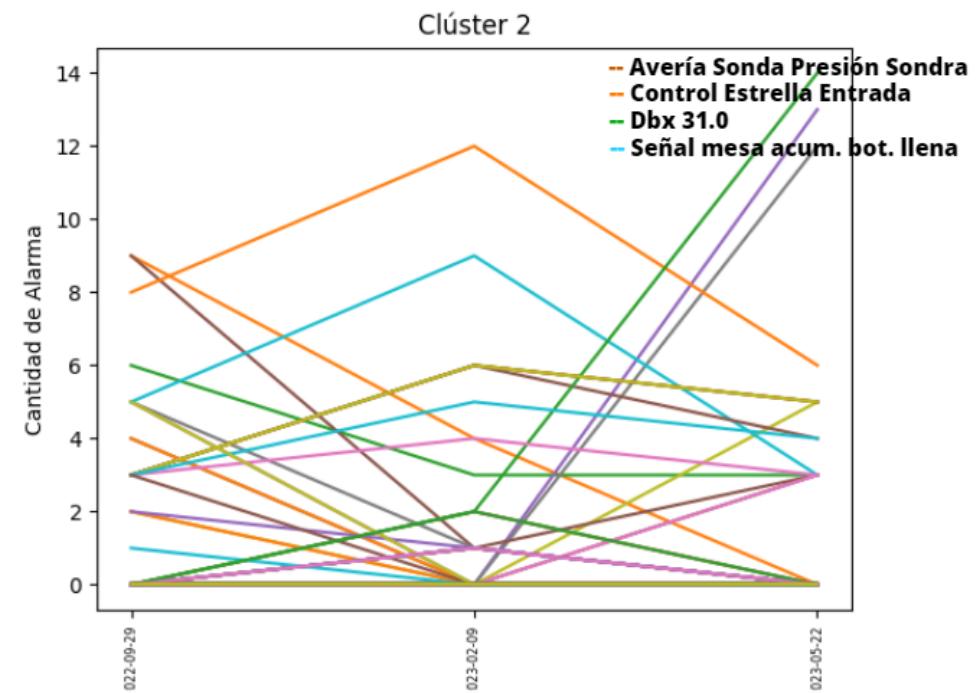


Figure 49: Cluster 2 Alarmas. Fuente: Elaboración Propia

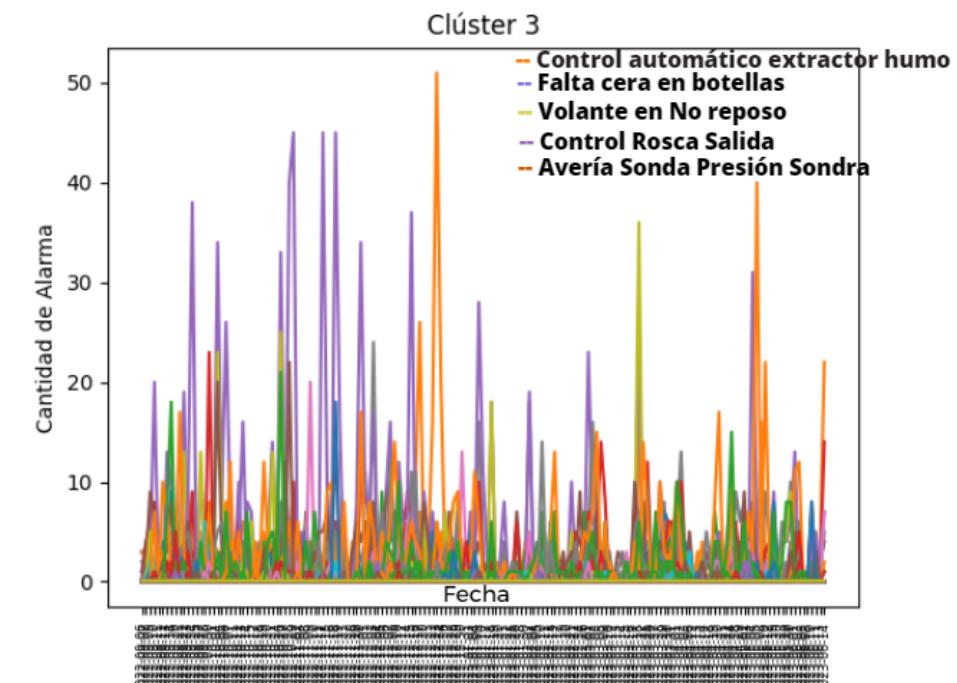


Figure 50: Cluster 3 Alarmas. Fuente: Elaboración Propia

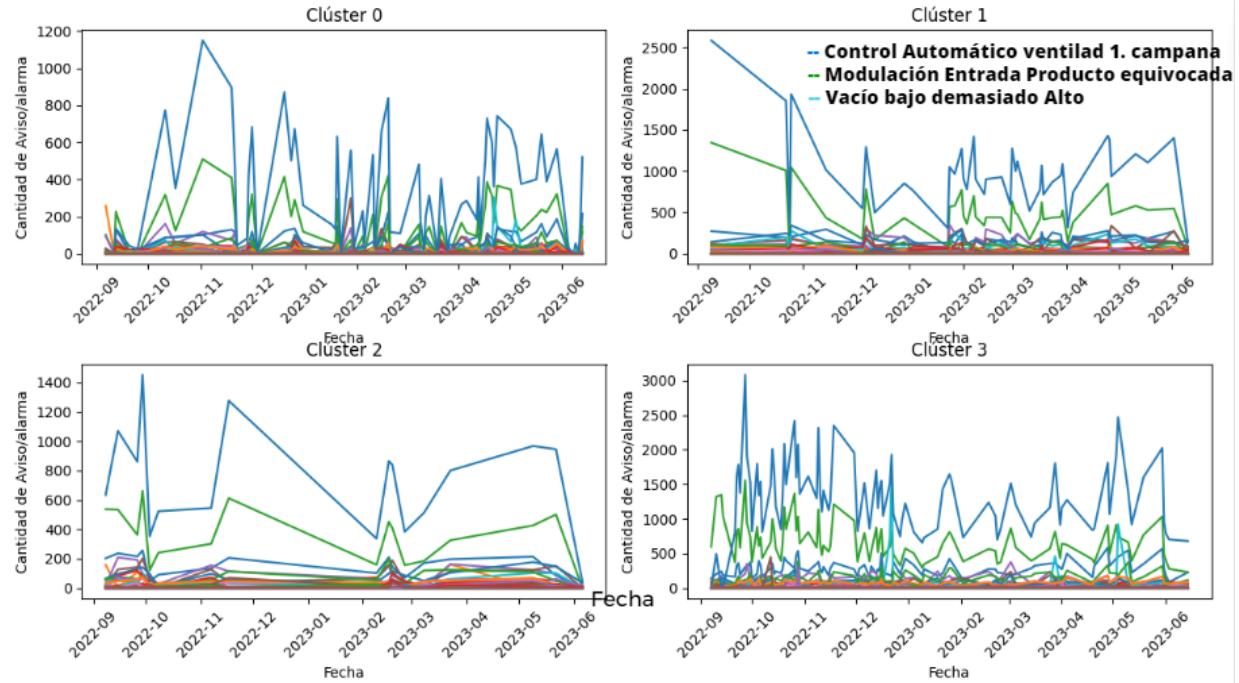


Figure 51: Clusters Avisos/Alarmas. Fuente: Elaboración Propia

## 15.10 Anexo 9 Matrices de Correlación por cluster

	min	max	mean	total_time	total_count	fail_count
min	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	NaN	NaN	NaN	NaN	NaN
total_time	NaN	NaN	NaN	1.0	NaN	1.0
total_count	NaN	NaN	NaN	NaN	NaN	NaN
fail_count	NaN	NaN	NaN	1.0	NaN	1.0

Table 16: Matriz de correlación para Cluster 0

	min	max	mean	total_time	total_count	fail_count
min	1.000000	0.241051	0.794213	0.548559	-0.251107	-0.072556
max	0.241051	1.000000	0.780872	0.855780	0.785229	0.293514
mean	0.794213	0.780872	1.000000	0.890787	0.325543	0.146211
total_time	0.548559	0.855780	0.890787	1.000000	0.576396	0.713183
total_count	-0.251107	0.785229	0.325543	0.576396	1.000000	0.347005
fail_count	-0.072556	0.293514	0.146211	0.713183	0.347005	1.000000

Table 17: Matriz de correlación para Cluster 1

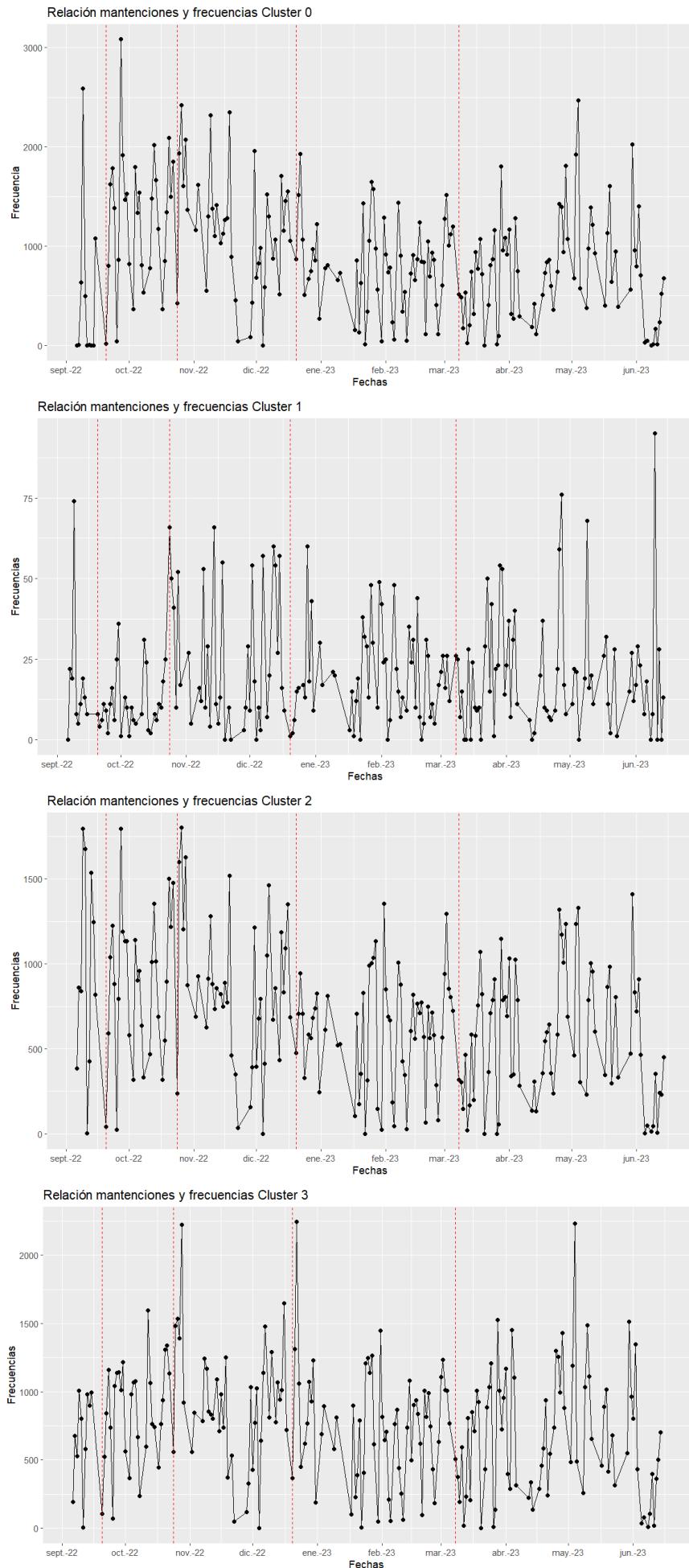
	min	max	mean	total_time	total_count	fail_count
min	1.000000	-0.152544	0.721018	-0.163101	NaN	-0.190785
max	-0.152544	1.000000	0.292324	0.998577	NaN	0.984810
mean	0.721018	0.292324	1.000000	0.283165	NaN	0.255999
total_time	-0.163101	0.998577	0.283165	1.000000	NaN	0.990229
total_count	NaN	NaN	NaN	NaN	NaN	NaN
fail_count	-0.190785	0.984810	0.255999	0.990229	NaN	1.000000

Table 18: Matriz de correlación para Cluster 2

	min	max	mean	total_time	total_count	fail_count
min	1.000000	-0.134849	0.259179	-0.224282	-0.287216	-0.248391
max	-0.134849	1.000000	0.470487	0.853781	0.796446	0.816454
mean	0.259179	0.470487	1.000000	0.070839	-0.018242	0.015375
total_time	-0.224282	0.853781	0.070839	1.000000	0.910173	0.996611
total_count	-0.287216	0.796446	-0.018242	0.910173	1.000000	0.908430
fail_count	-0.248391	0.816454	0.015375	0.996611	0.908430	1.000000

Table 19: Matriz de correlación para Cluster 3

## 15.11 Anexo 10 Mantenciones por Cluster



## 15.12 Anexo 11 Predicción y pérdida para cada falla

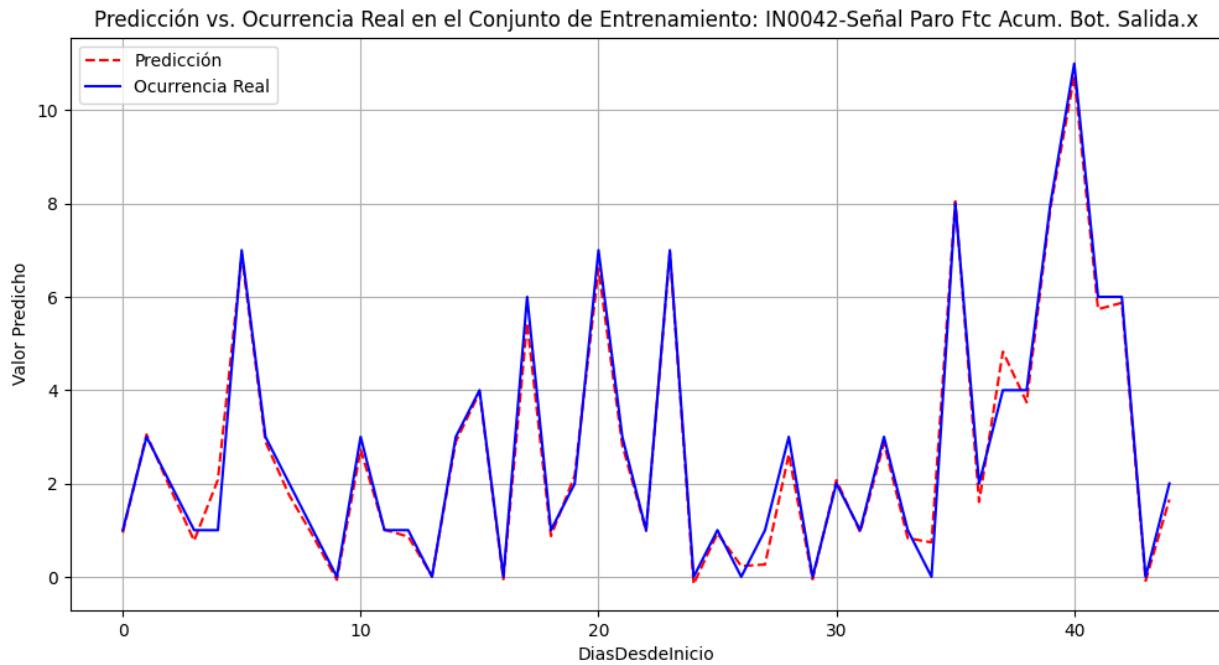


Figure 53: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

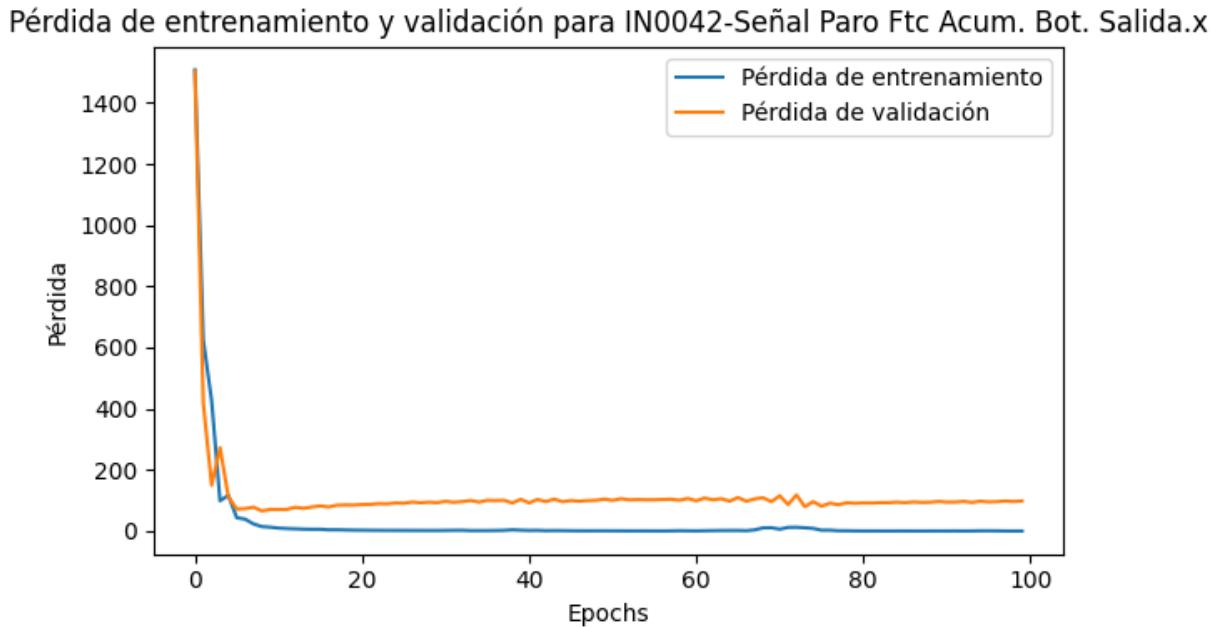


Figure 54: Pérdidas para cada época. Fuente: Elaboración Propia

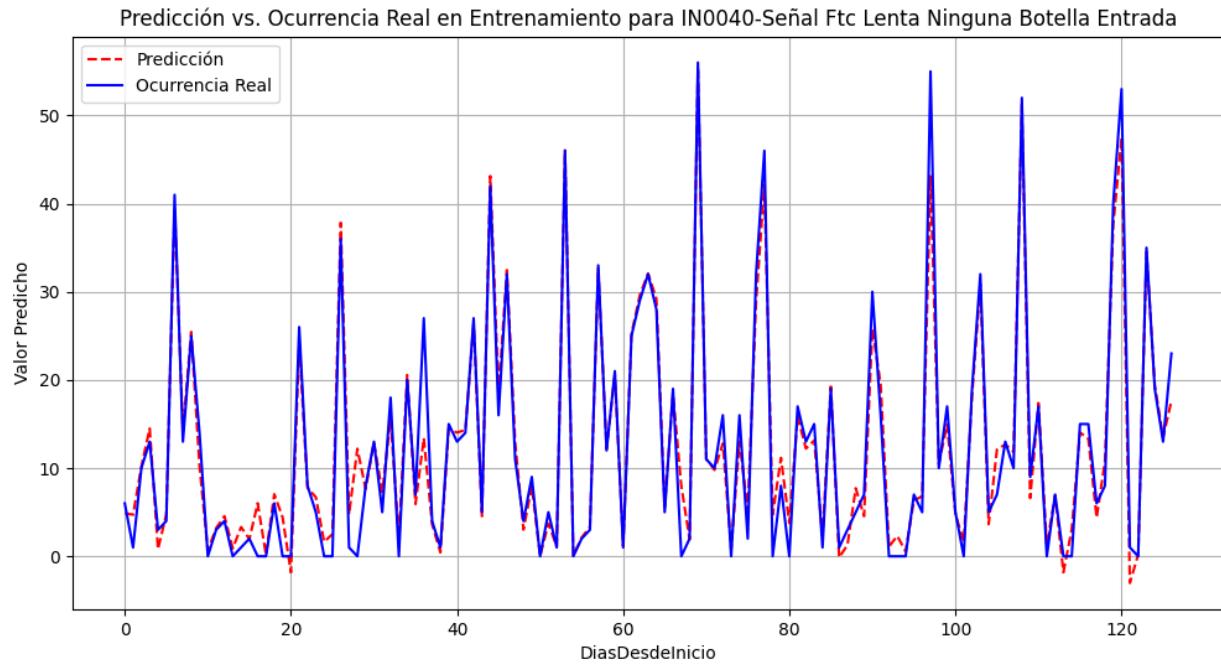


Figure 55: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

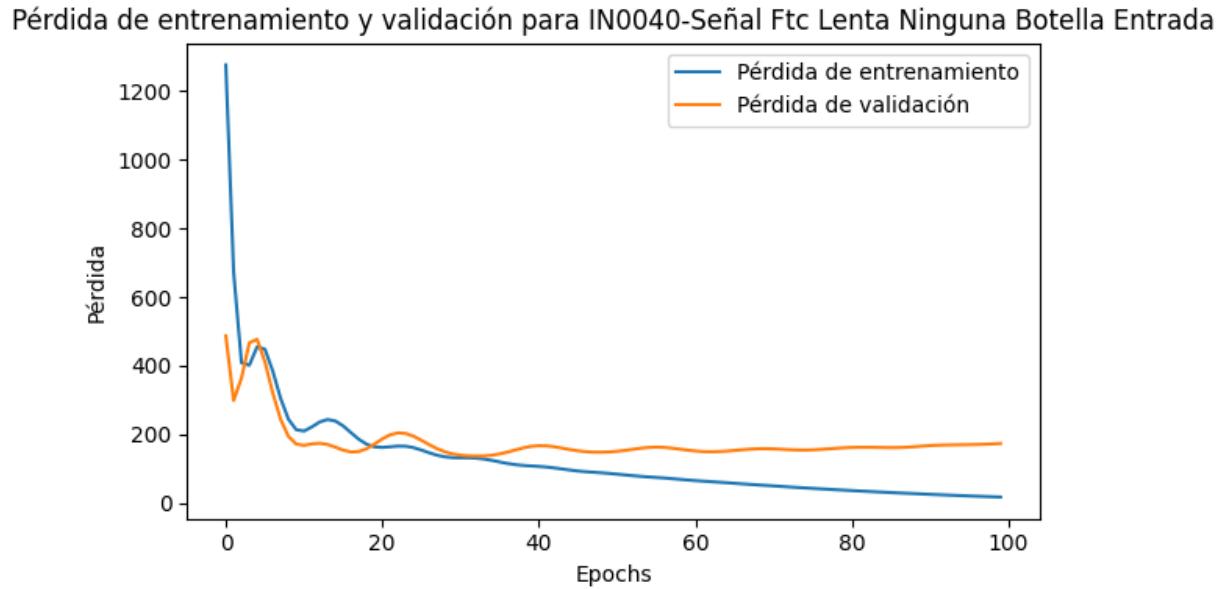


Figure 56: Pérdidas para cada época. Fuente: Elaboración Propia

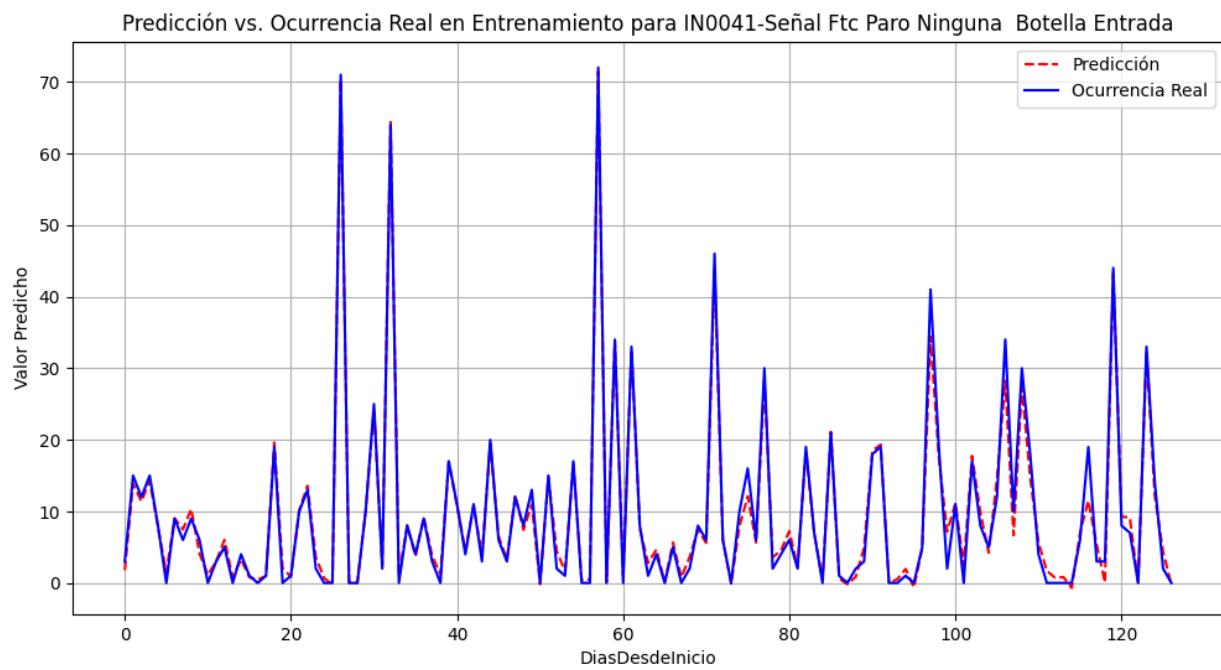


Figure 57: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

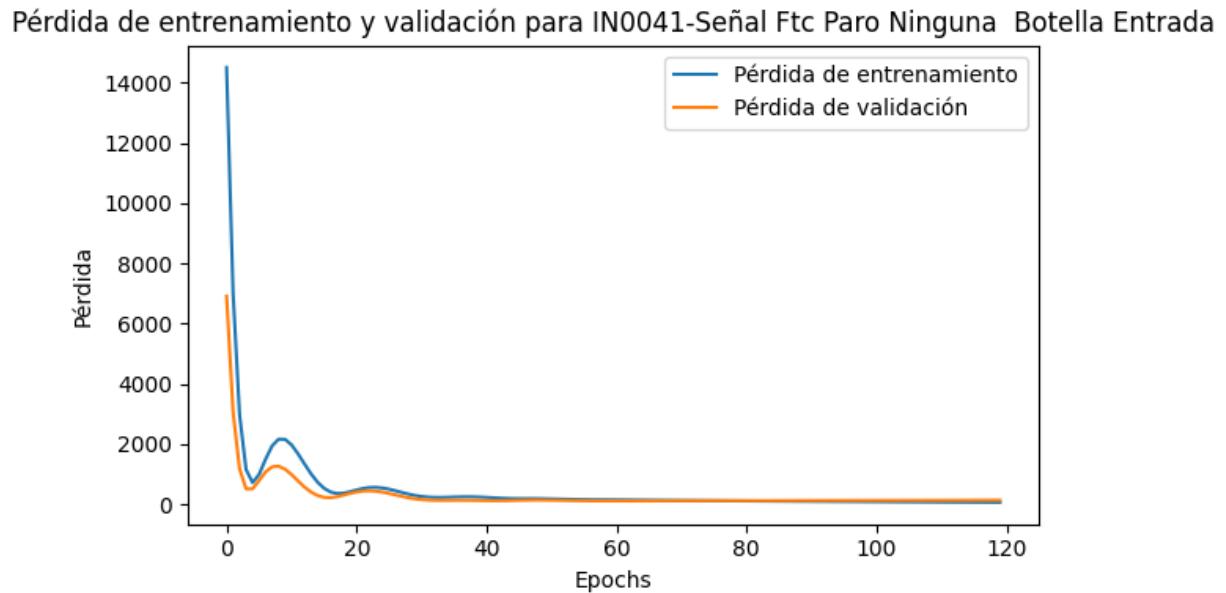


Figure 58: Pérdidas para cada época. Fuente: Elaboración Propia

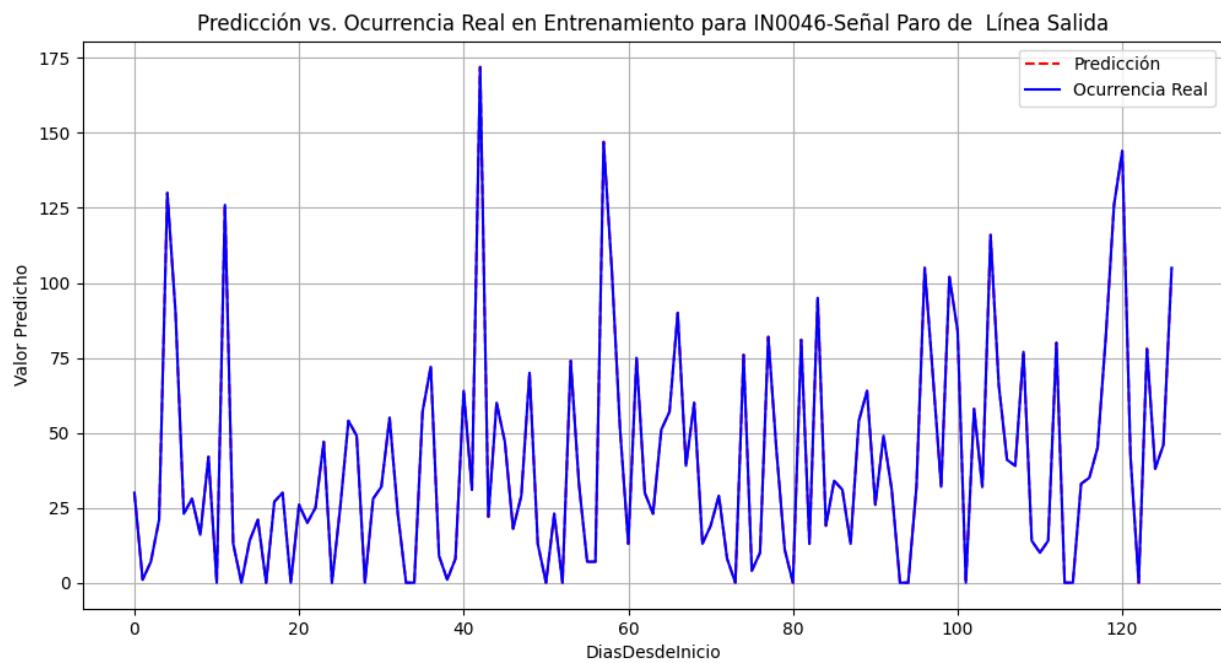


Figure 59: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

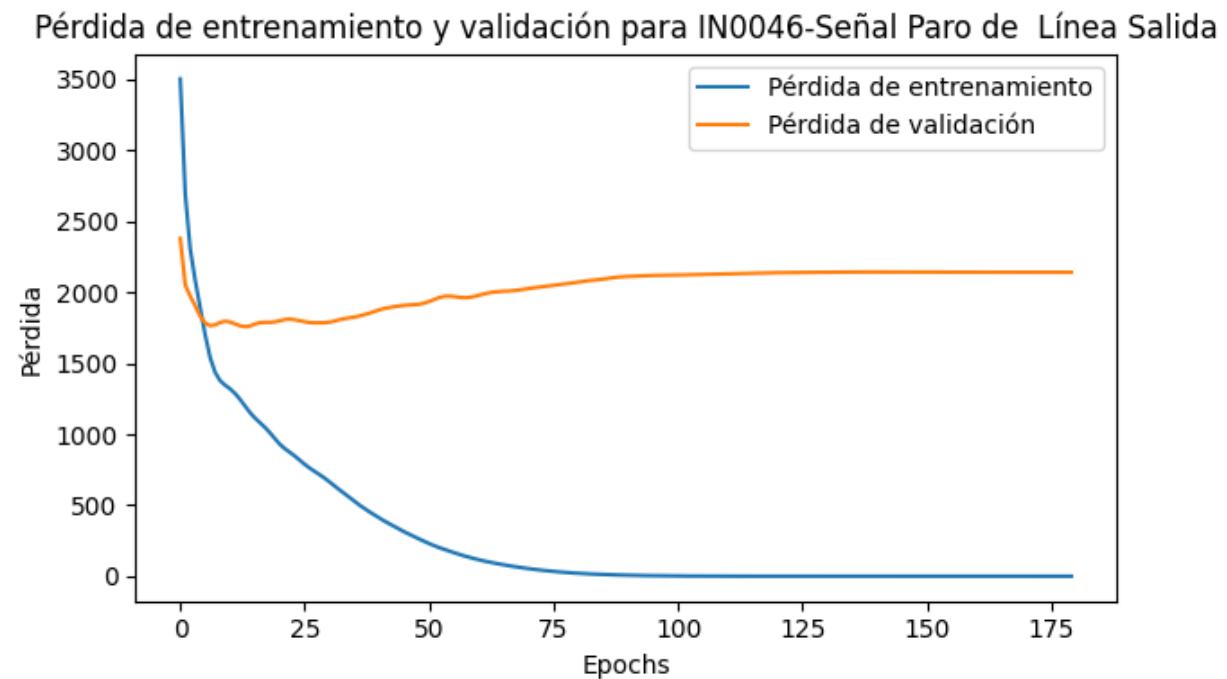


Figure 60: Pérdidas para cada época. Fuente: Elaboración Propia

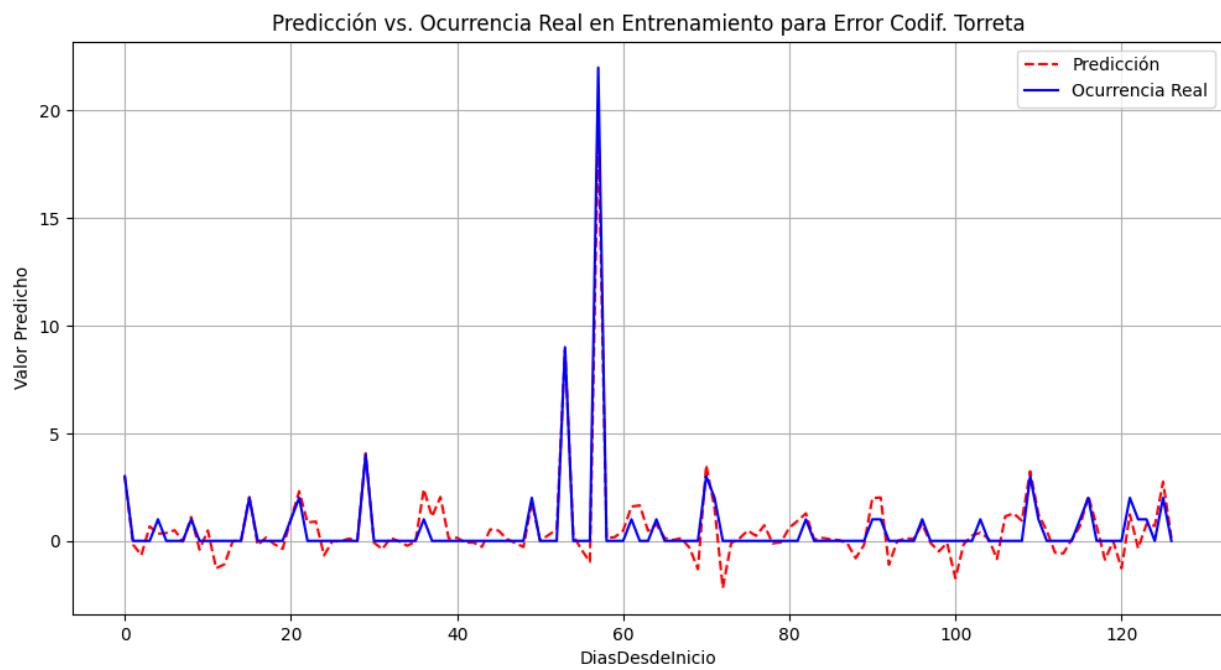


Figure 61: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

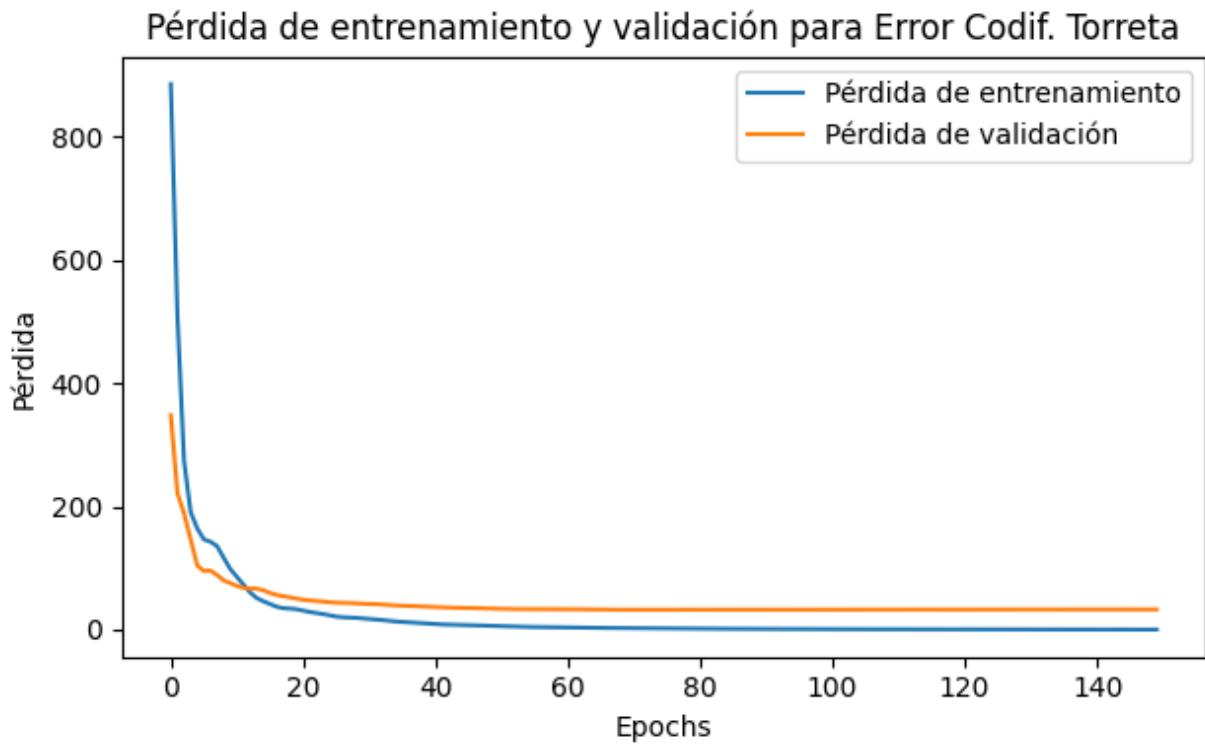


Figure 62: Pérdidas para cada época. Fuente: Elaboración Propia

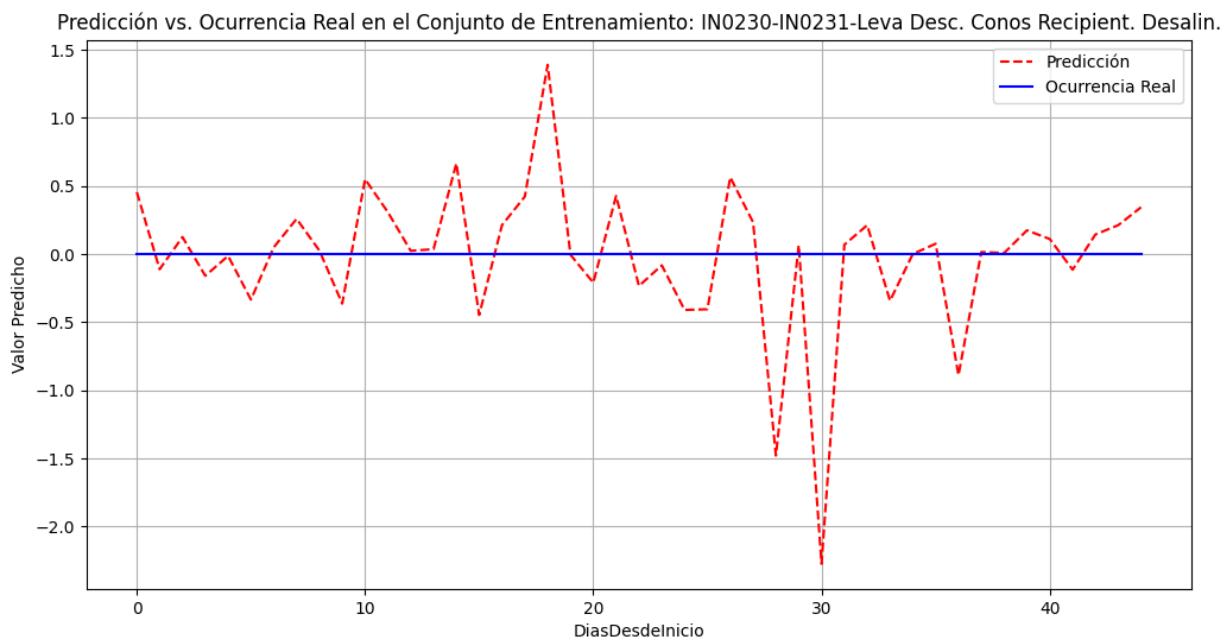


Figure 63: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

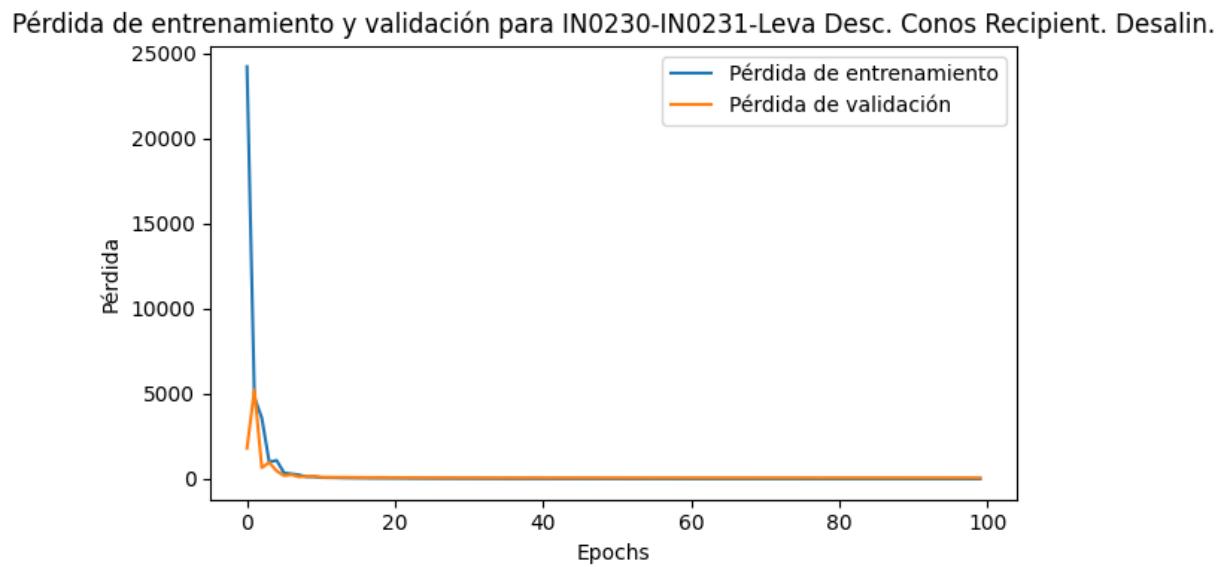


Figure 64: Pérdidas para cada época. Fuente: Elaboración Propia

Predicción vs. Ocurrencia Real en el Conjunto de Entrenamiento: IN0320-IN0321-Leva Desc. Conos Delanteros Desalin.

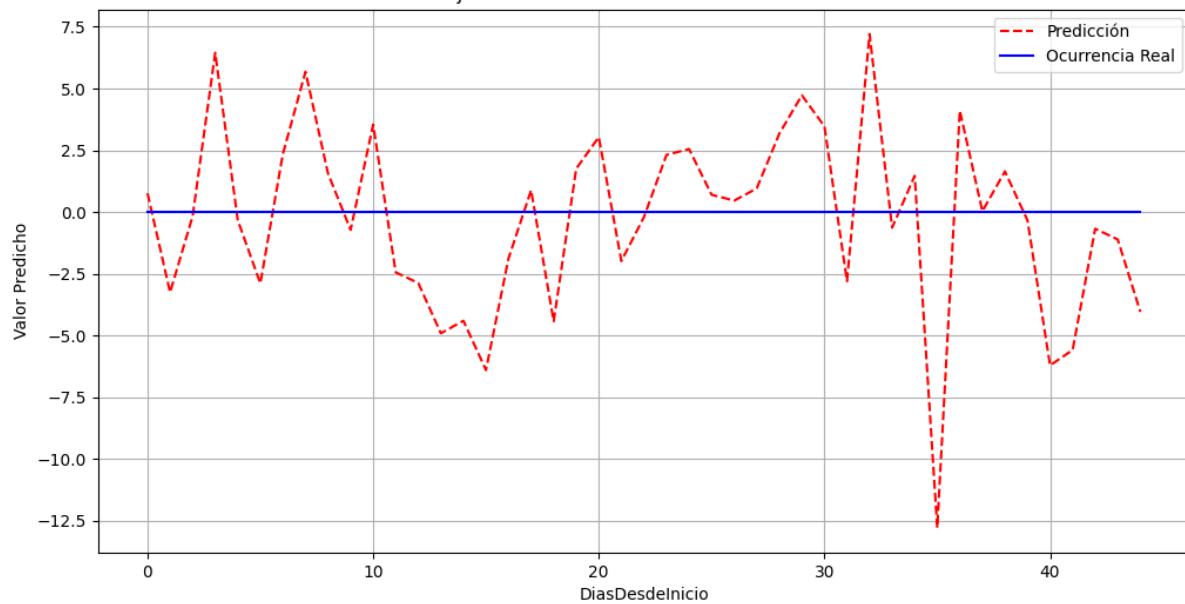


Figure 65: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

Pérdida de entrenamiento y validación para IN0320-IN0321-Leva Desc. Conos Delanteros Desalin.

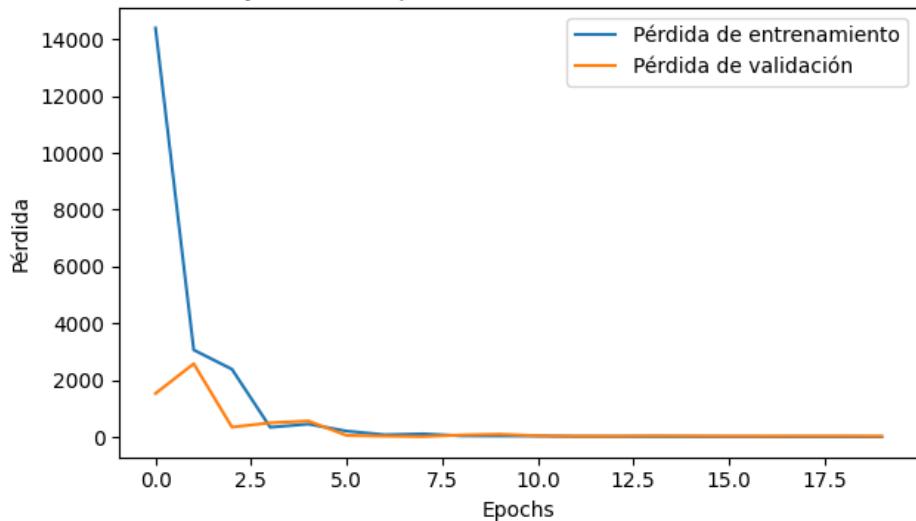


Figure 66: Pérdidas para cada época. Fuente: Elaboración Propia

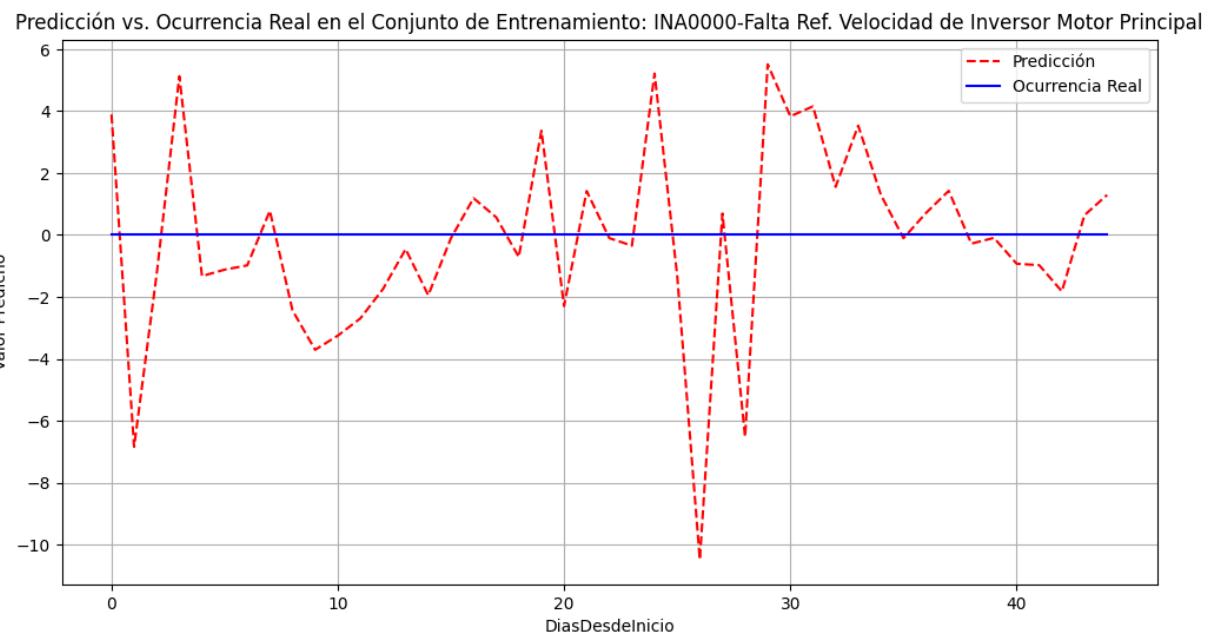


Figure 67: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

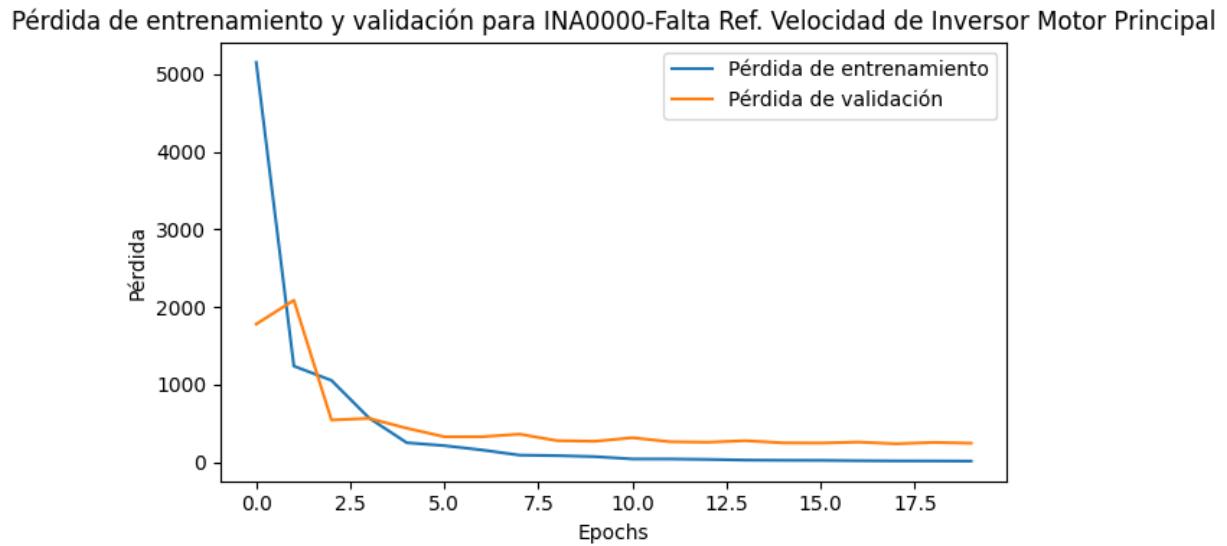


Figure 68: Pérdidas para cada época. Fuente: Elaboración Propia

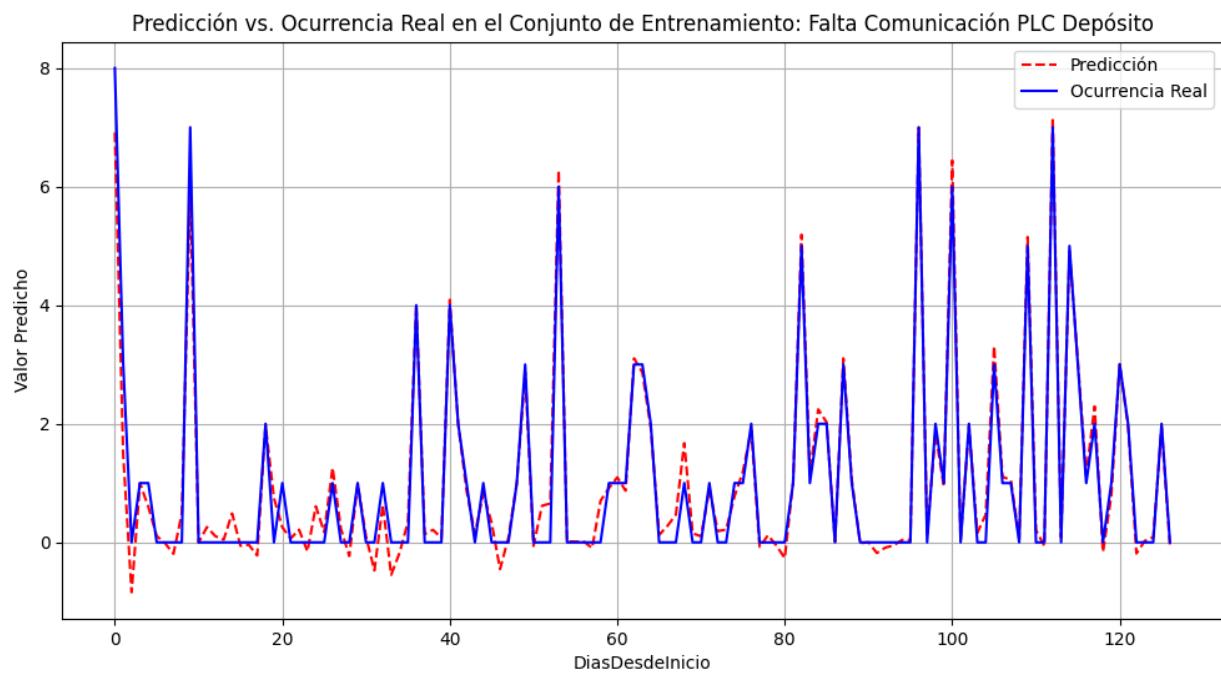


Figure 69: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

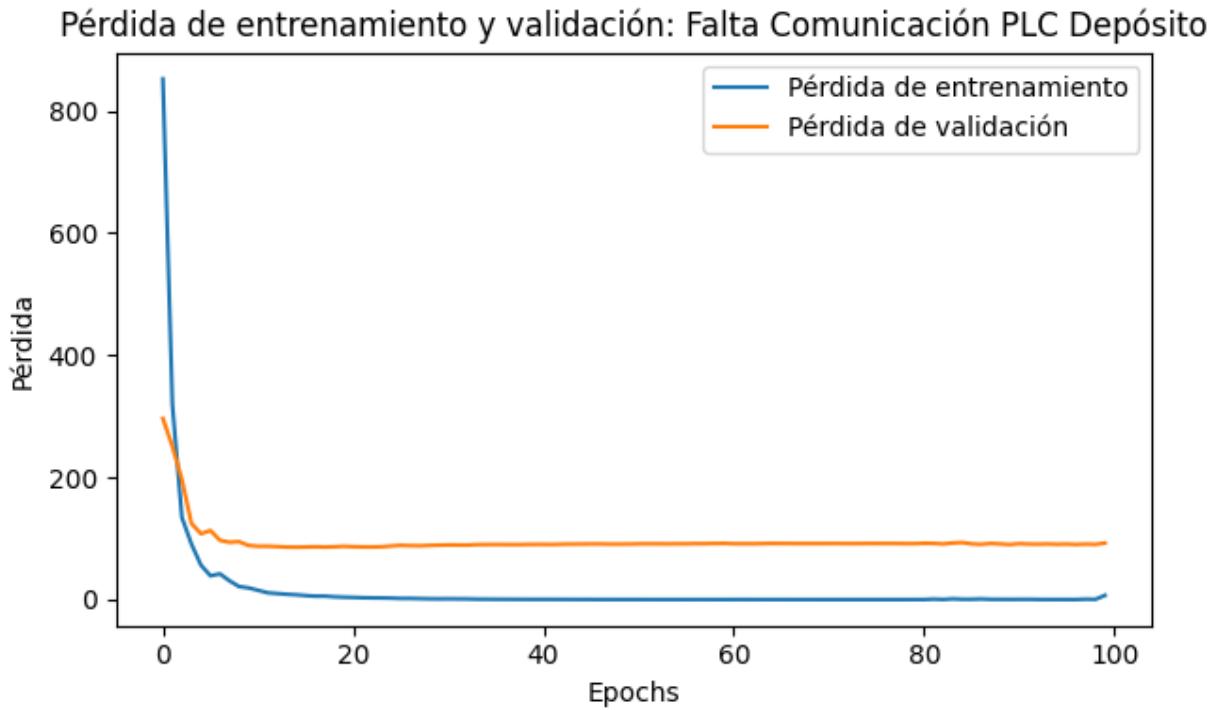


Figure 70: Pérdidas para cada época. Fuente: Elaboración Propia

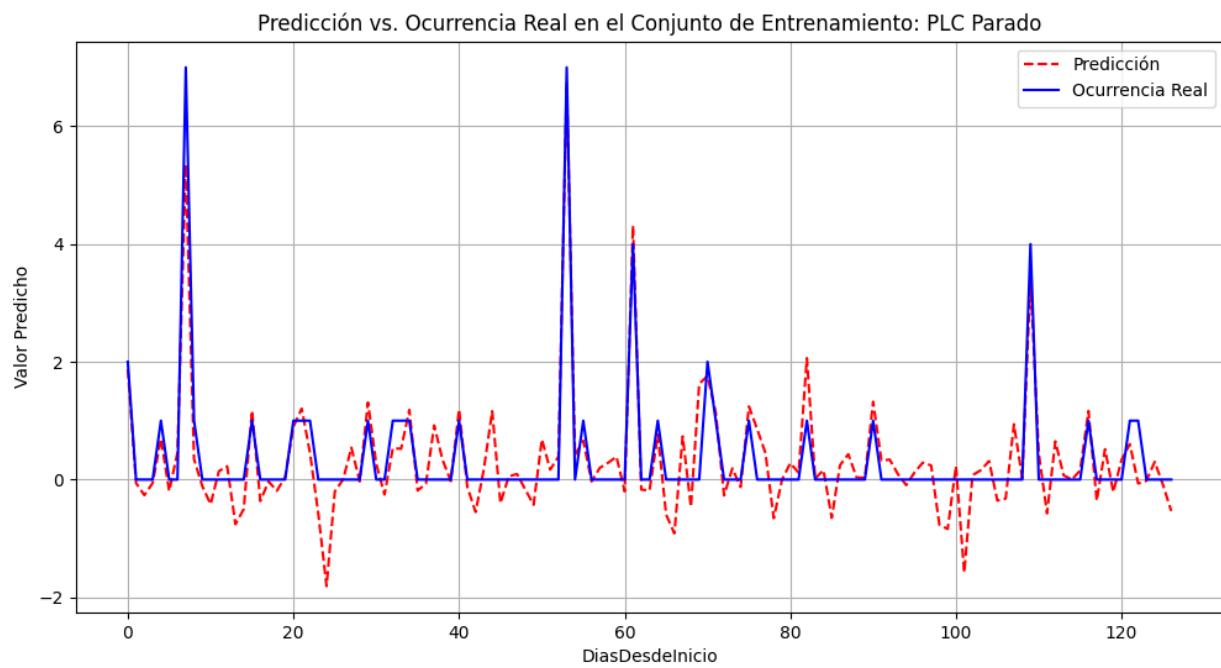


Figure 71: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

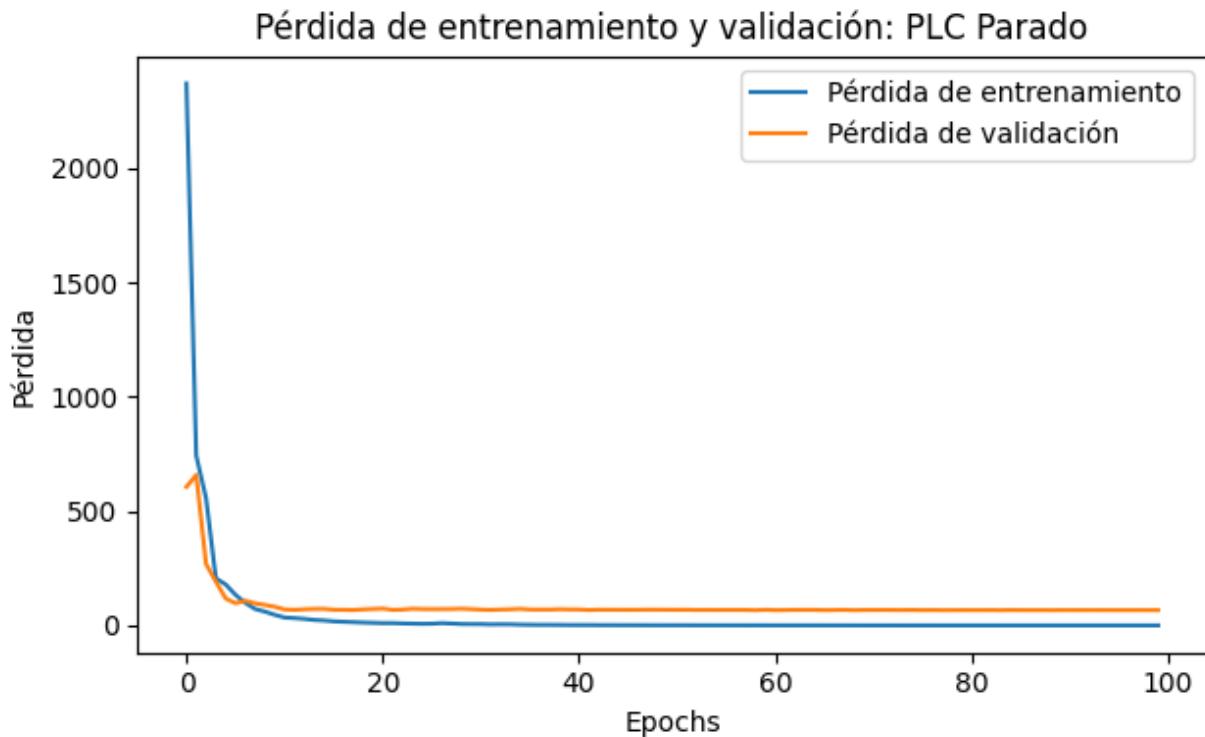


Figure 72: Pérdidas para cada época. Fuente: Elaboración Propia

Predicción vs. Ocurrencia Real en el Conjunto de Entrenamiento: IN0467-Control Temperatura No Adec. Pistola Pegamen.

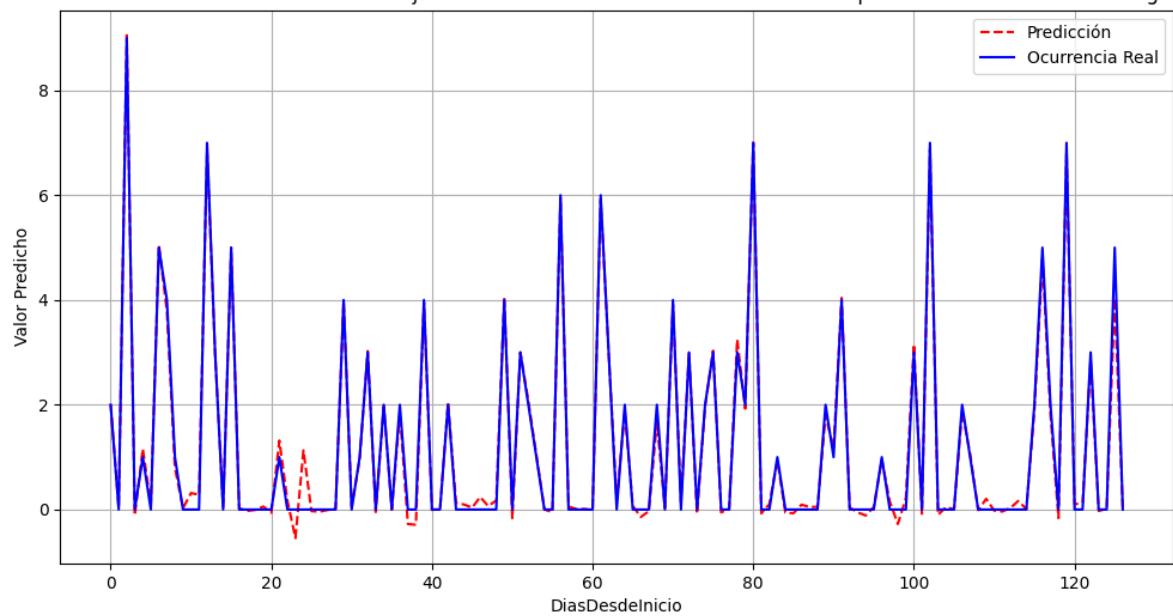


Figure 73: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

Pérdida de entrenamiento y validación: IN0467-Control Temperatura No Adec. Pistola Pegamen.

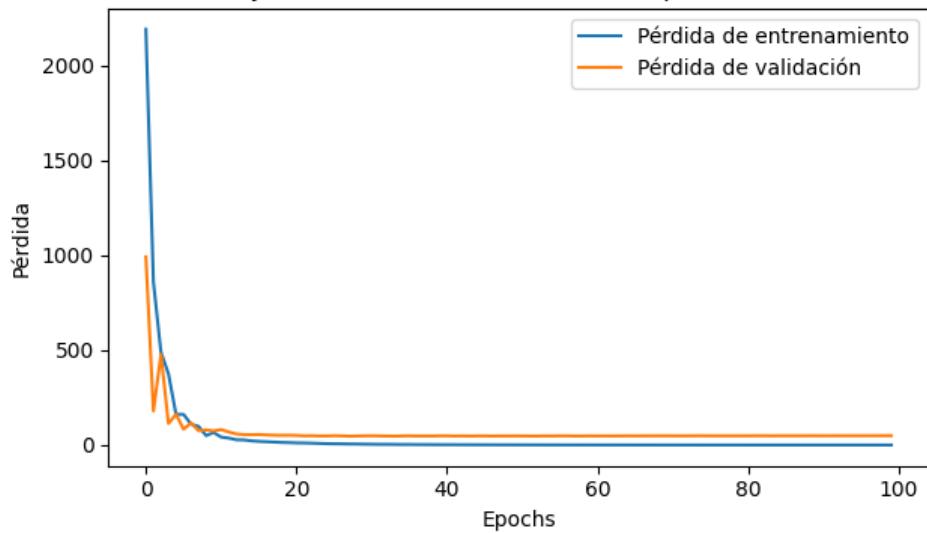


Figure 74: Pérdidas para cada época. Fuente: Elaboración Propia

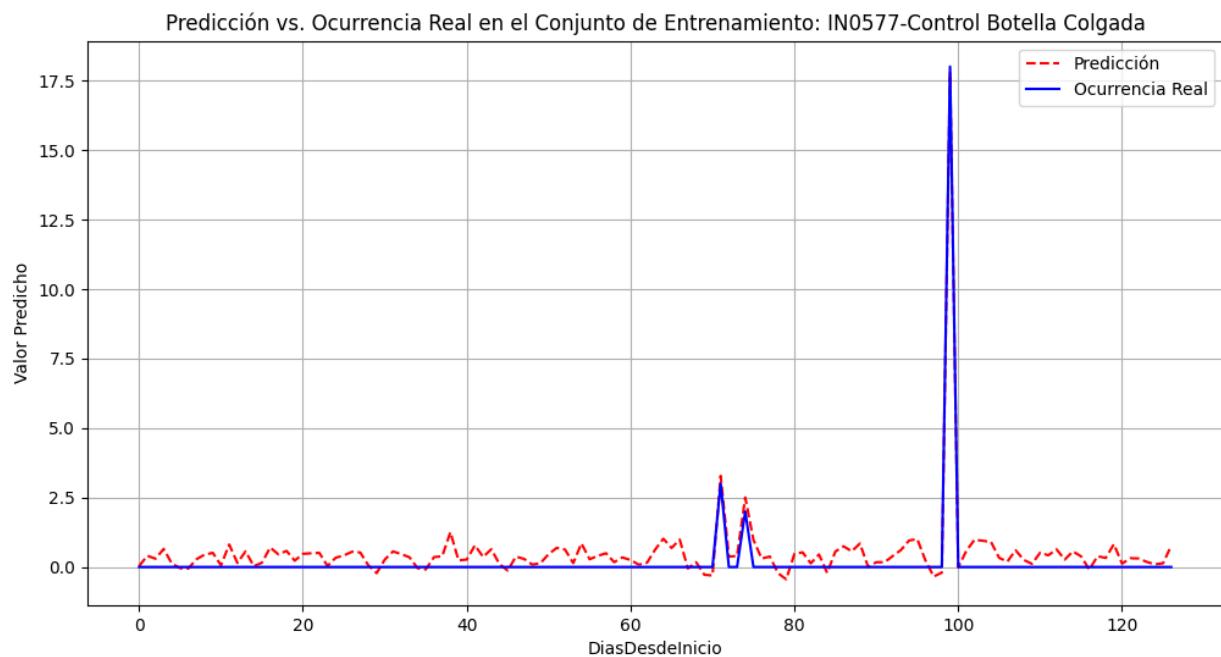


Figure 75: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

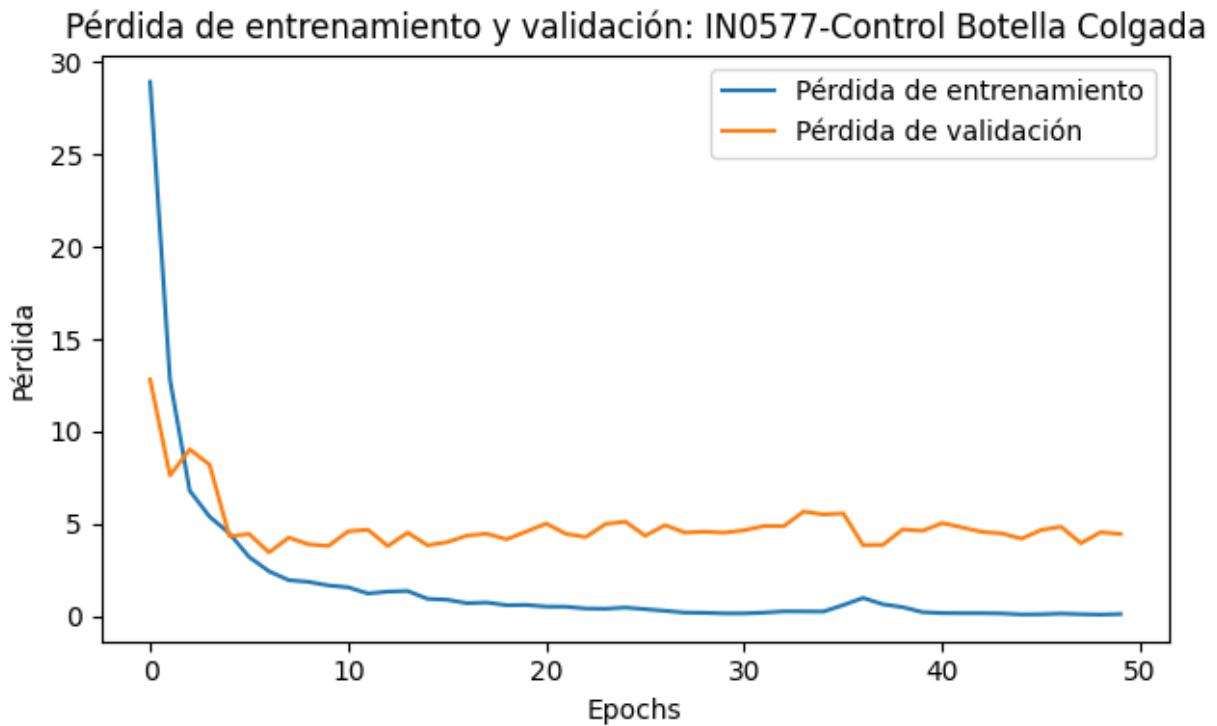


Figure 76: Pérdidas para cada época. Fuente: Elaboración Propia

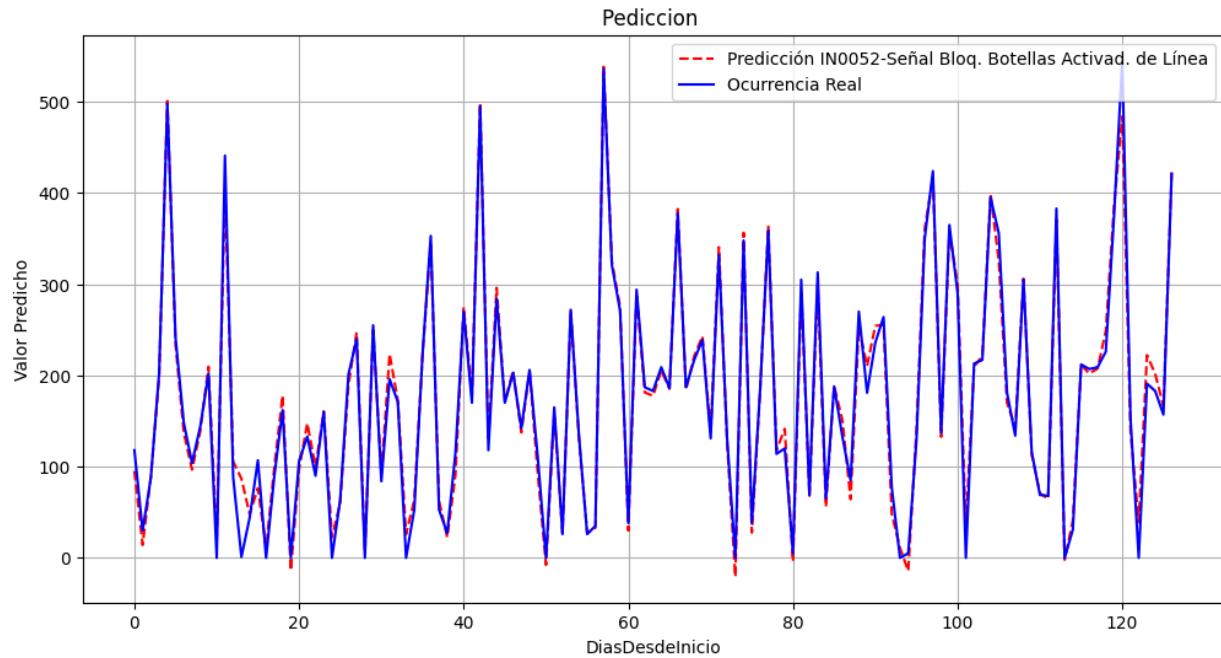


Figure 77: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

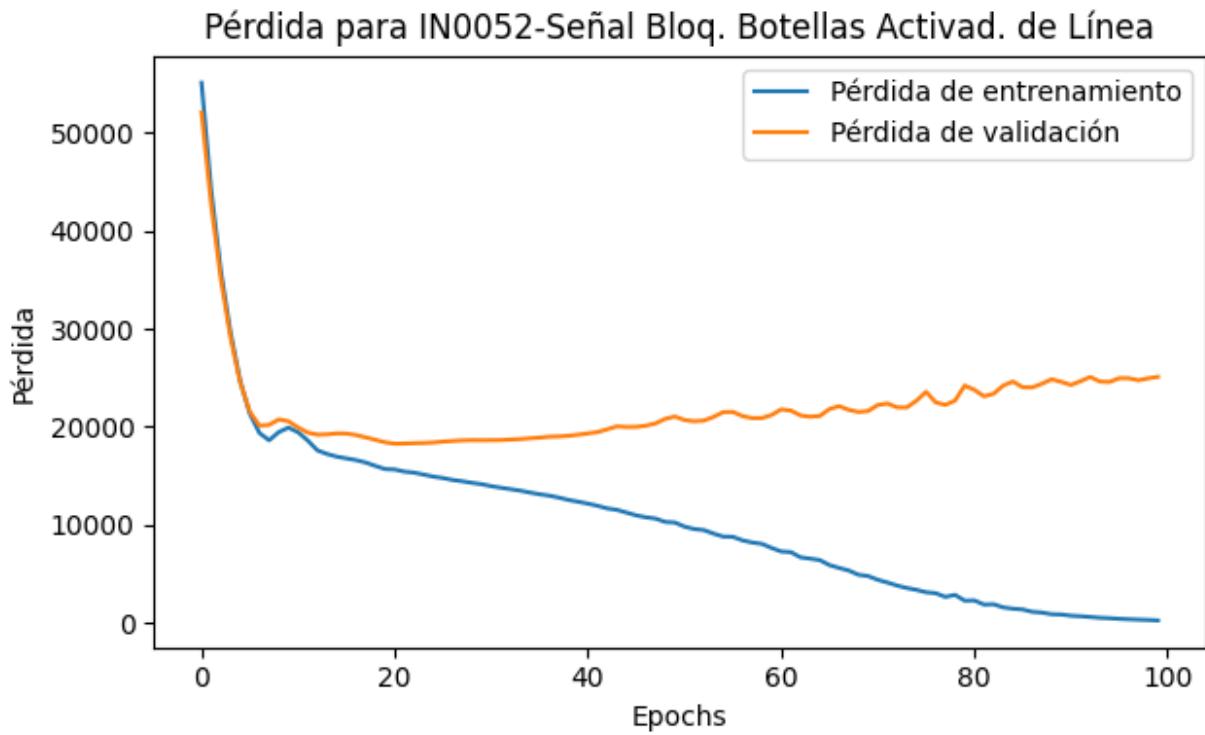


Figure 78: Pérdidas para cada época. Fuente: Elaboración Propia

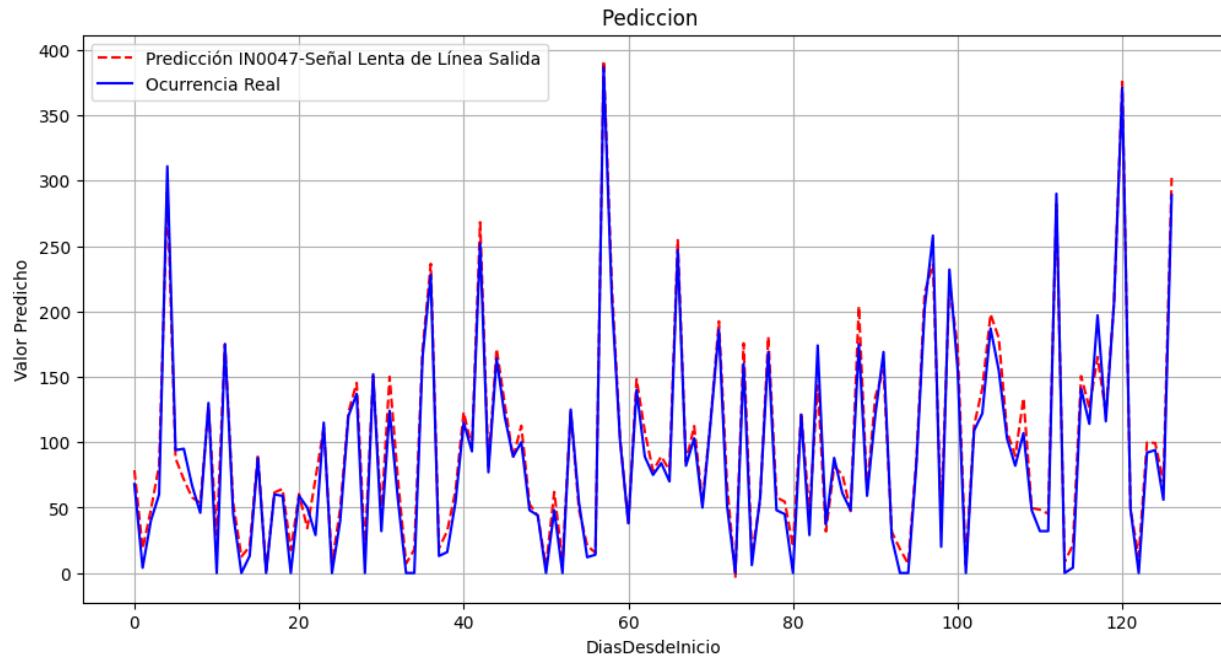


Figure 79: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

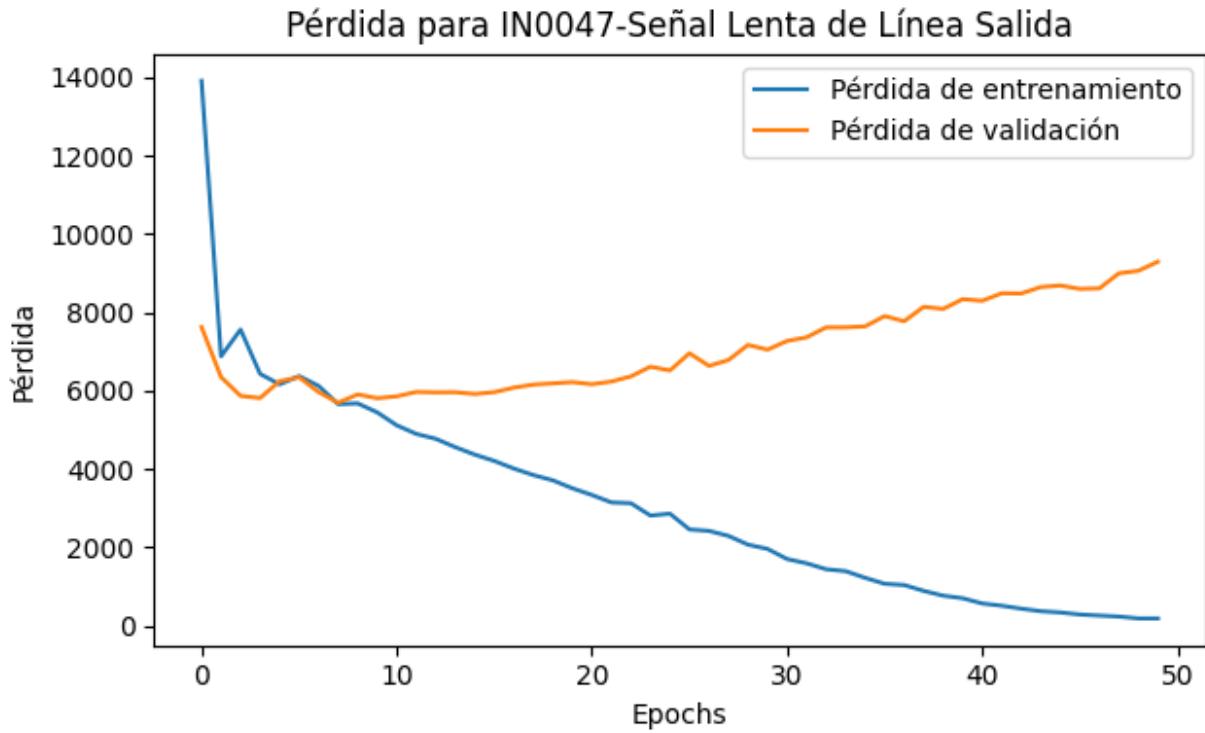


Figure 80: Pérdidas para cada época. Fuente: Elaboración Propia

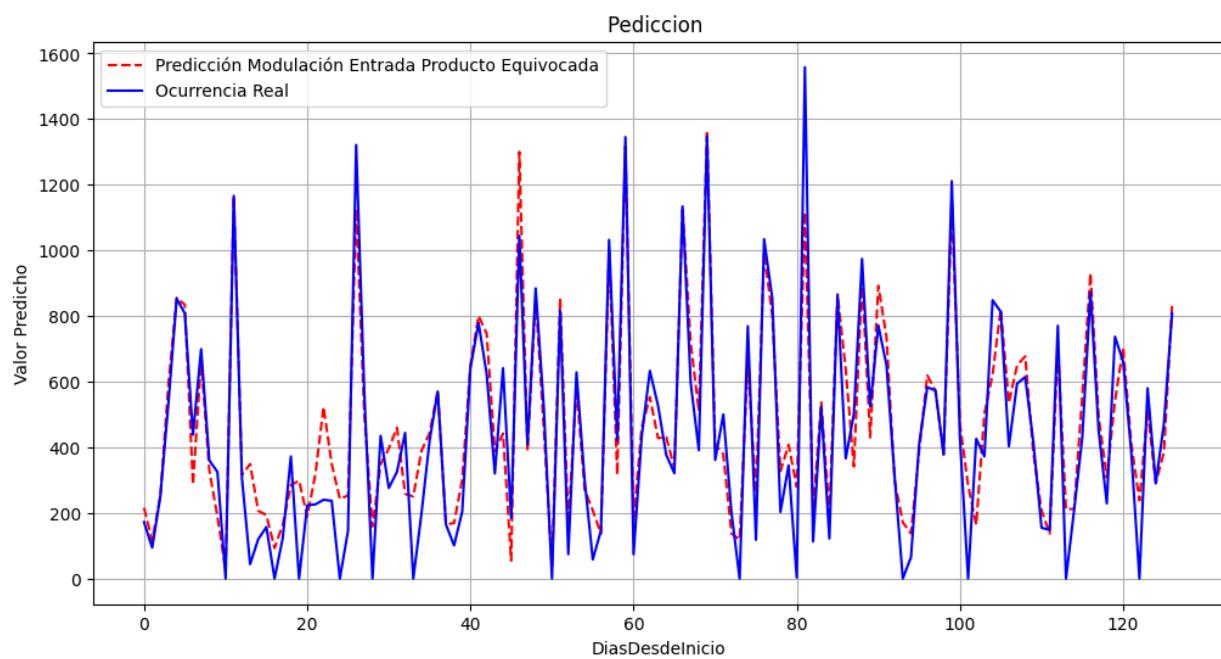


Figure 81: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

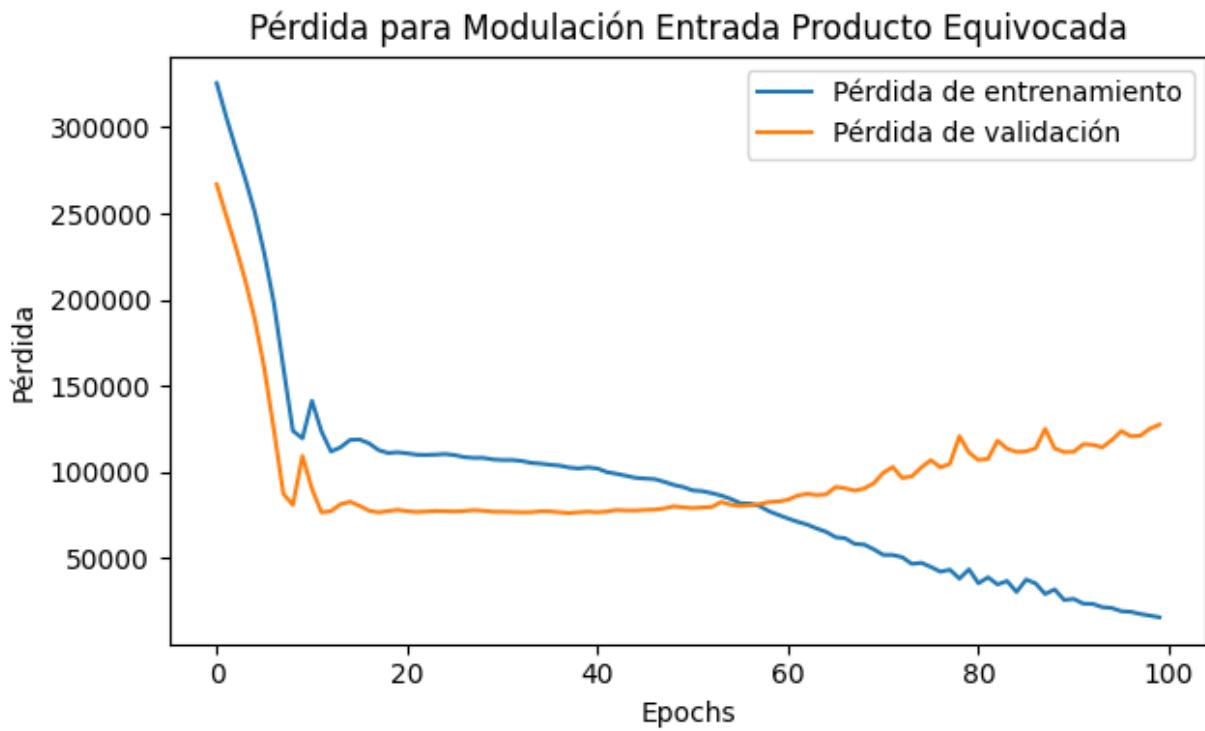


Figure 82: Pérdidas para cada época. Fuente: Elaboración Propia

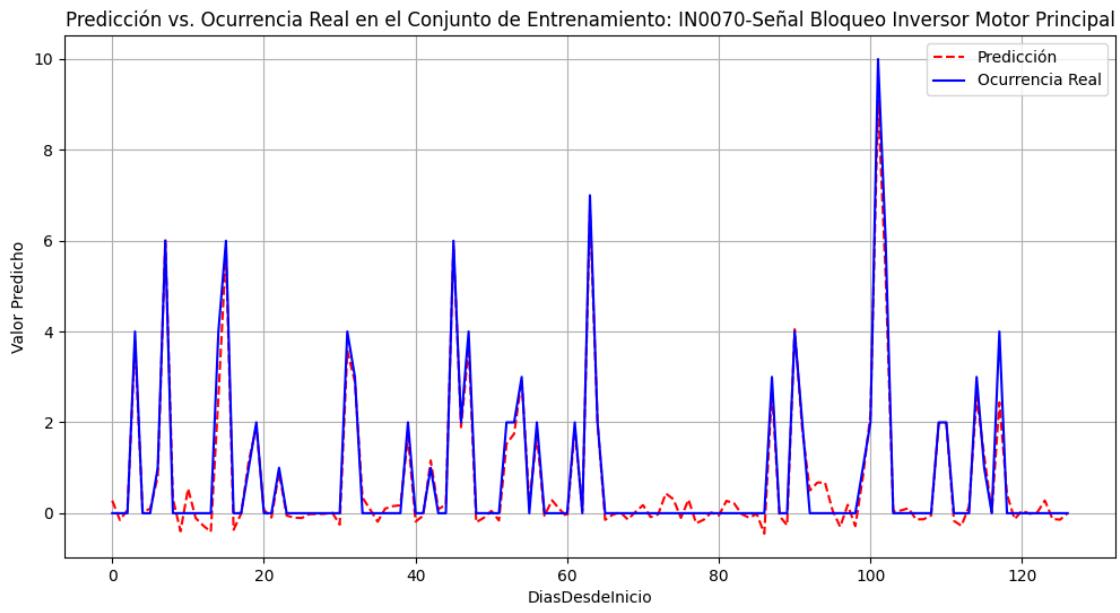


Figure 83: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

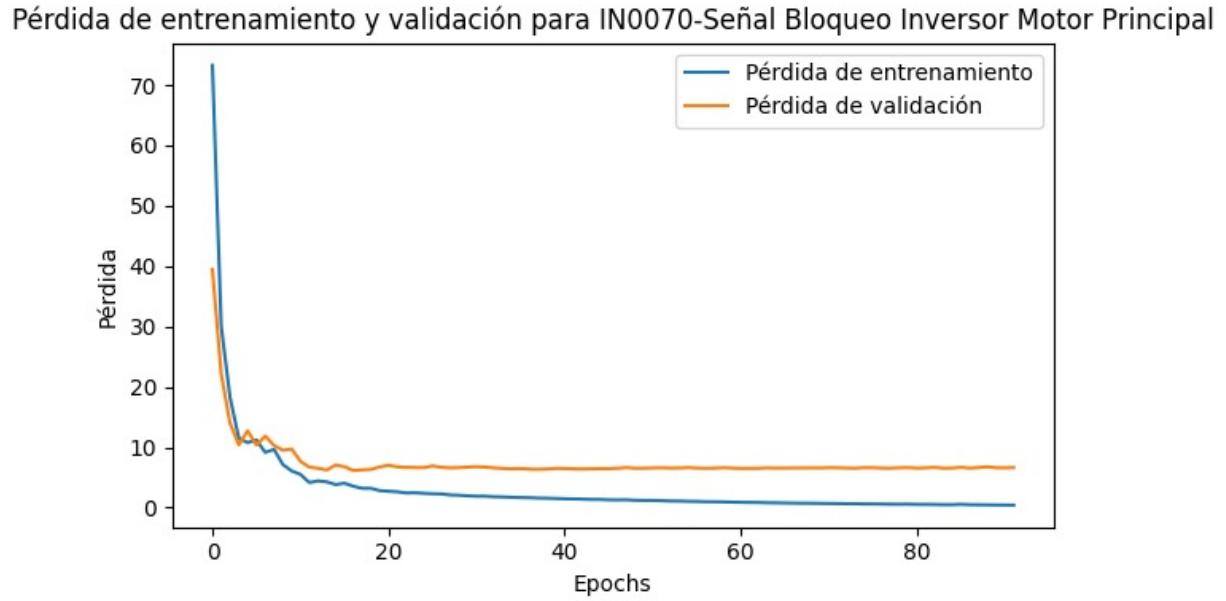


Figure 84: Pérdidas para cada época. Fuente: Elaboración Propia

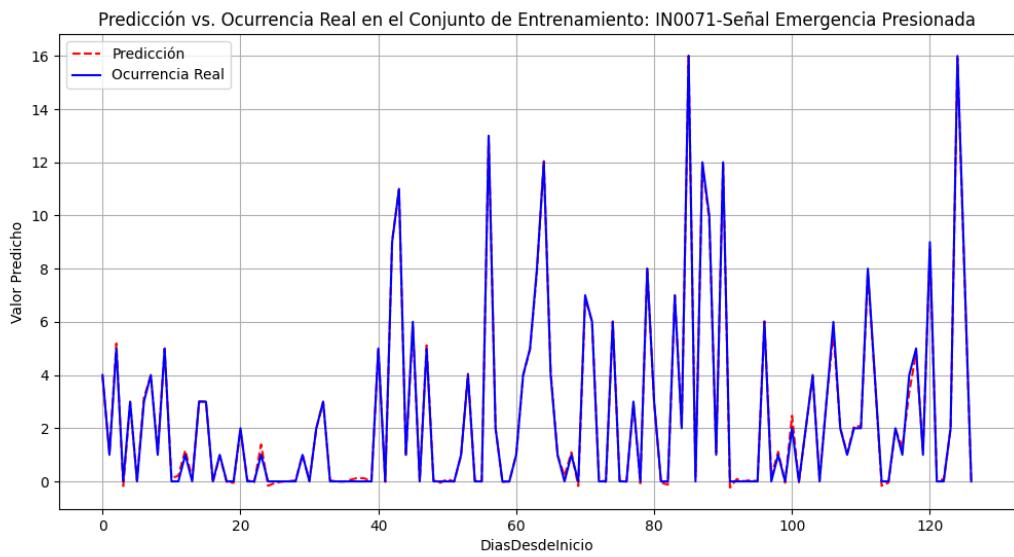


Figure 85: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

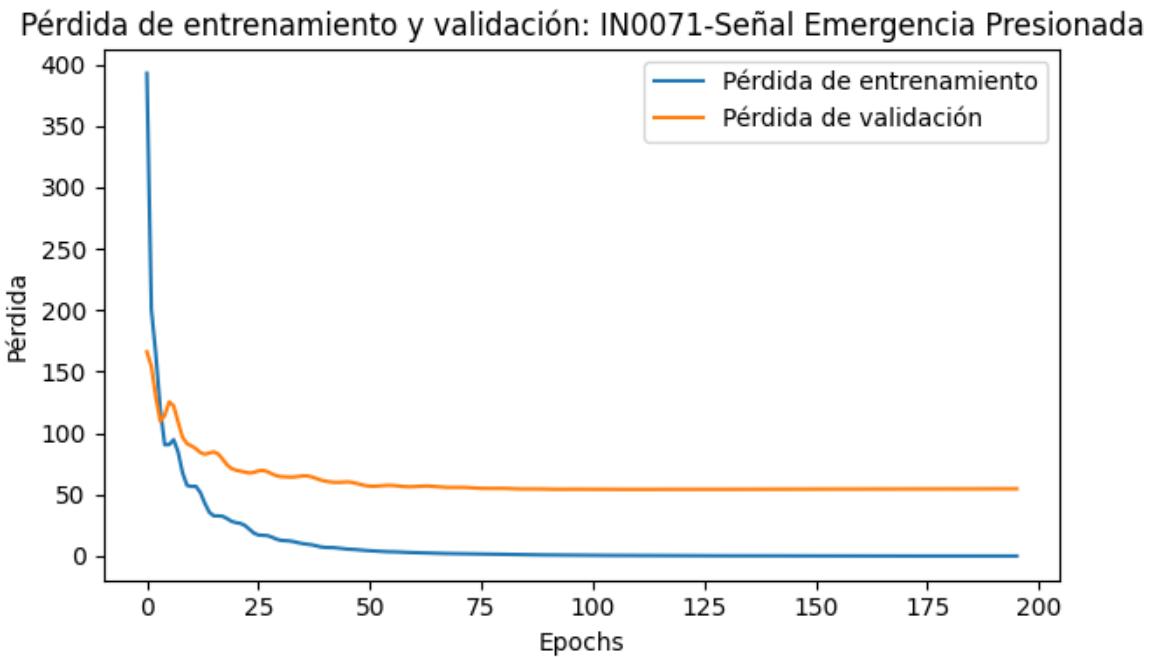


Figure 86: Pérdidas para cada época. Fuente: Elaboración Propia

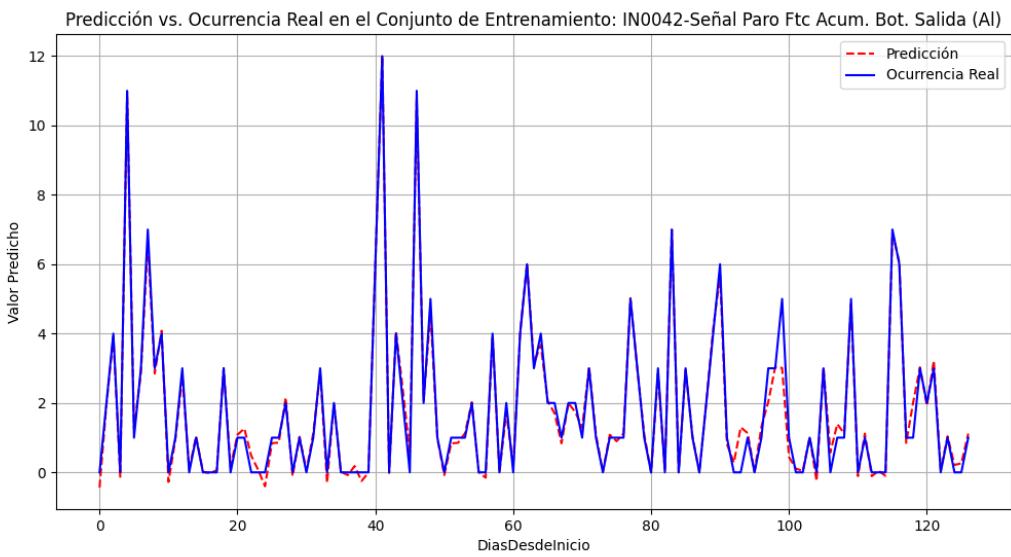


Figure 87: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

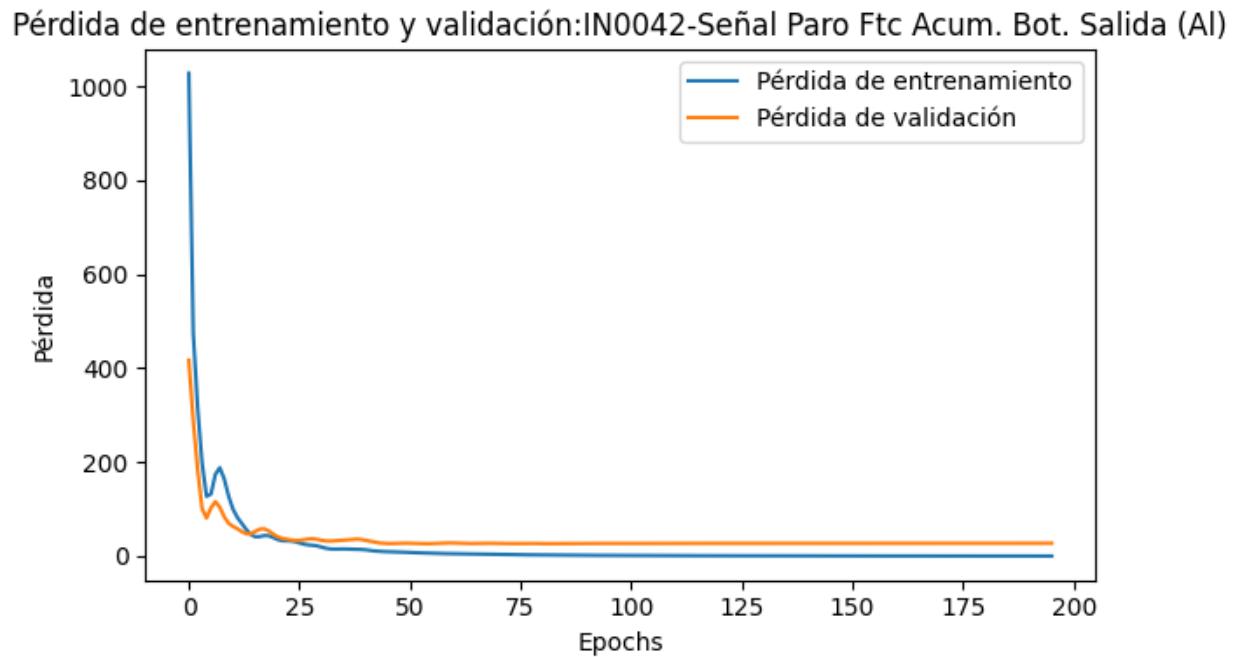


Figure 88: Pérdidas para cada época. Fuente: Elaboración Propia

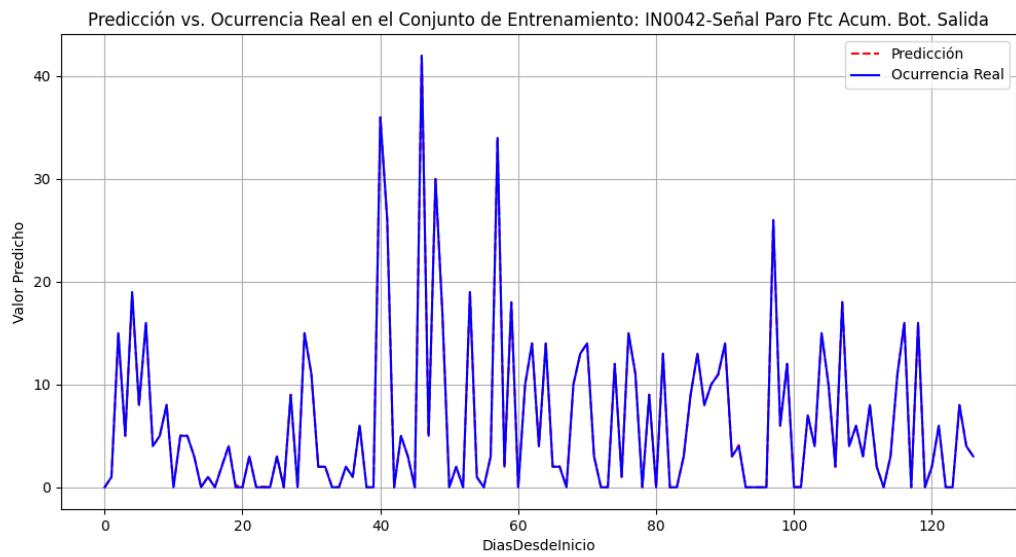


Figure 89: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

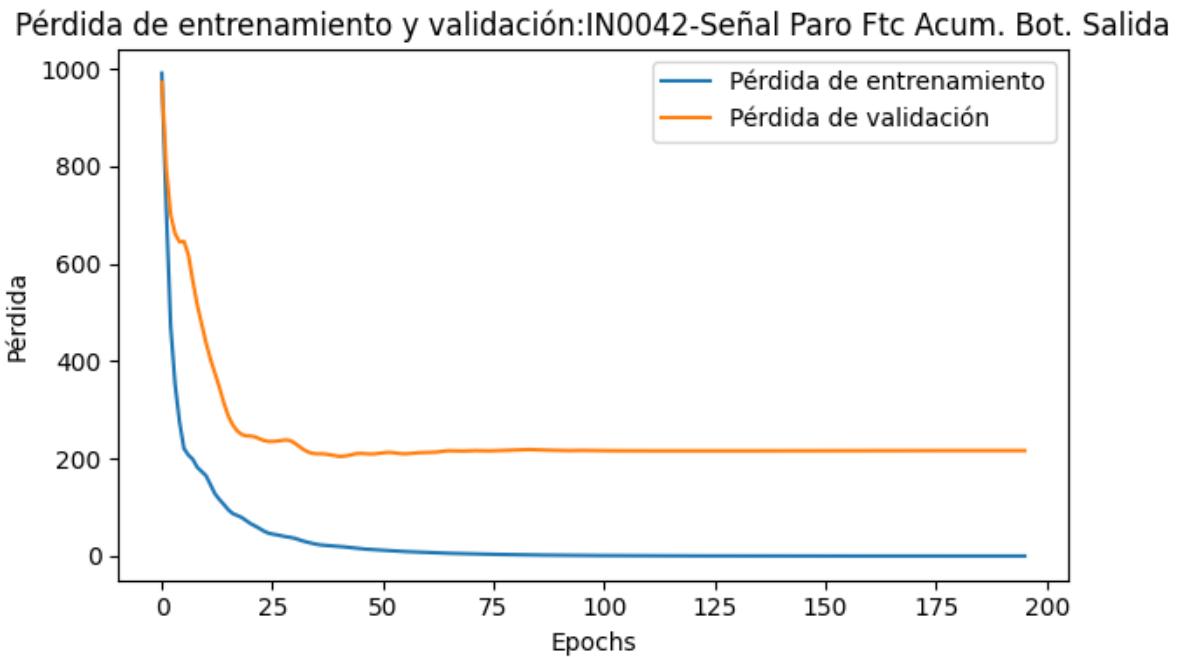


Figure 90: Perdidas para cada época. Fuente: Elaboración Propia

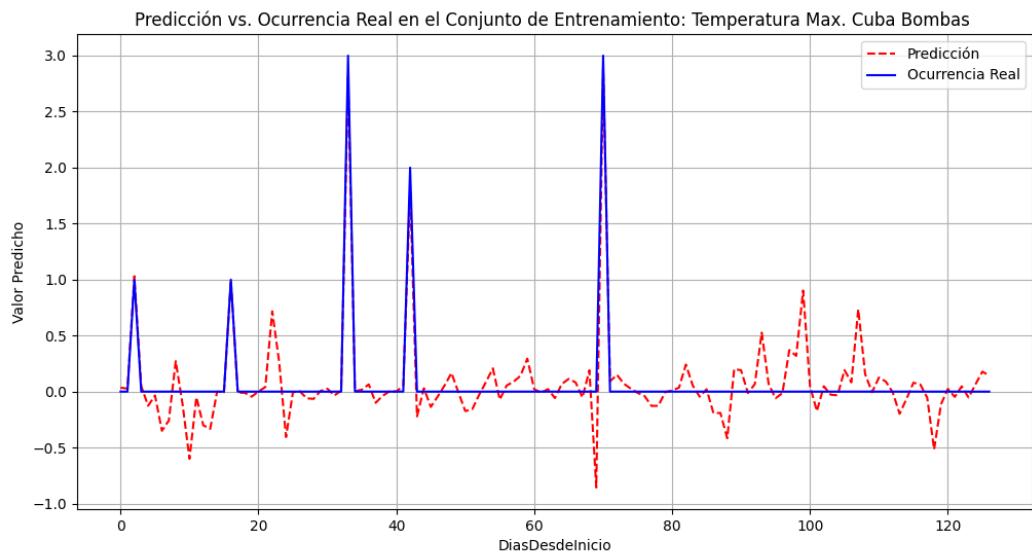


Figure 91: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

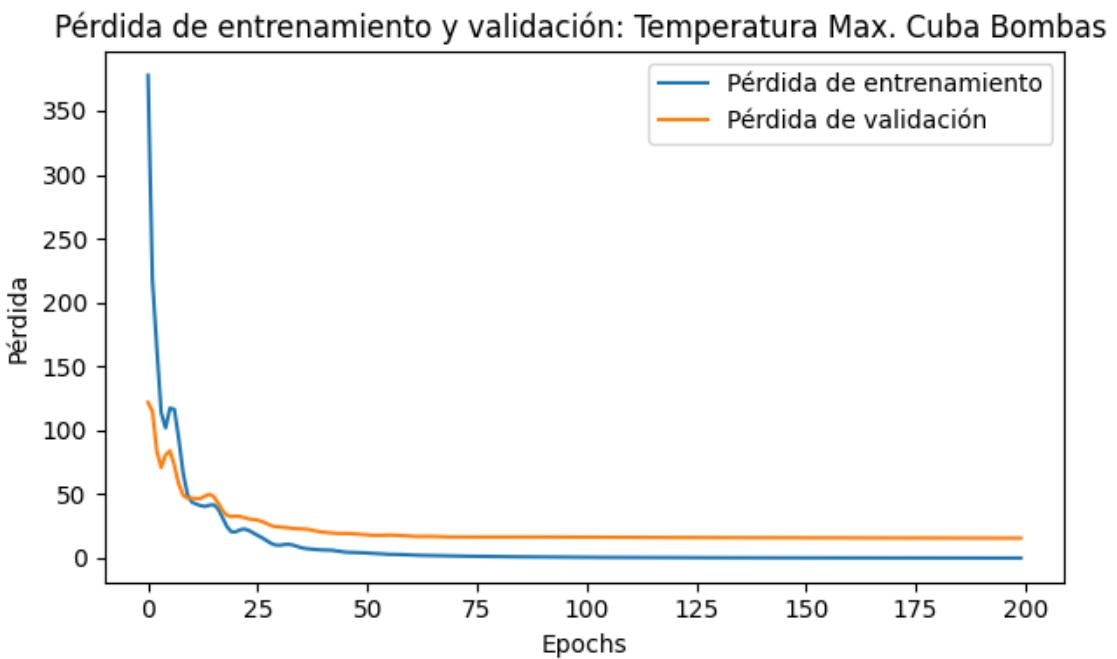


Figure 92: Pérdidas para cada época. Fuente: Elaboración Propia

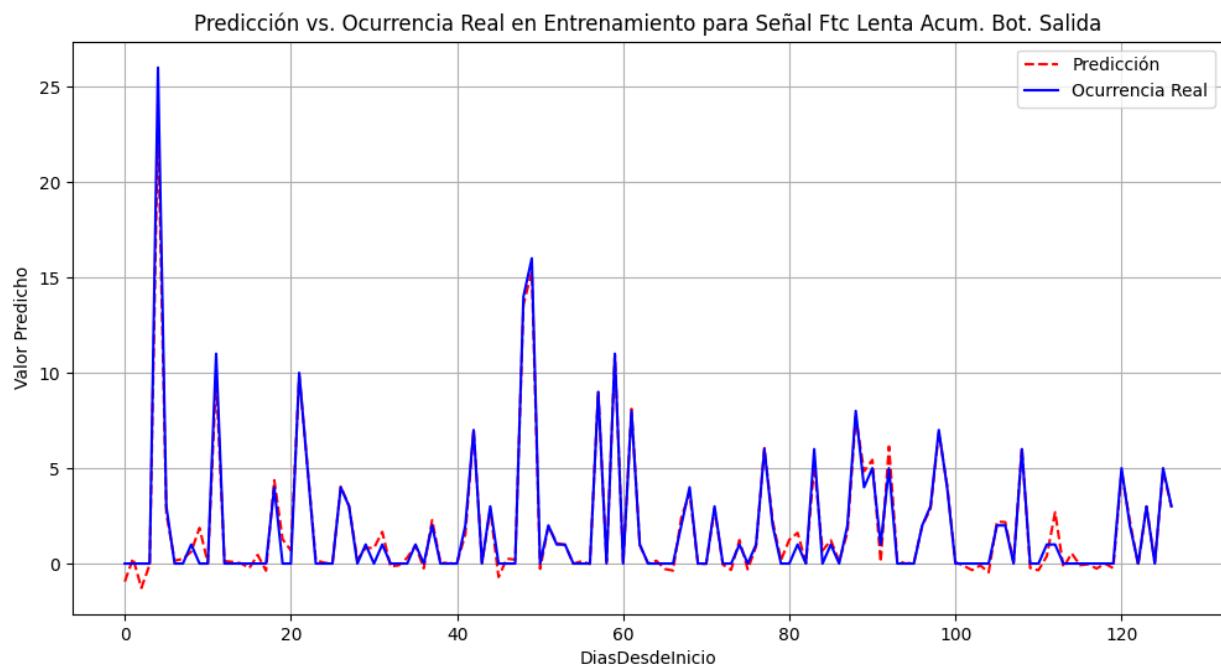


Figure 93: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

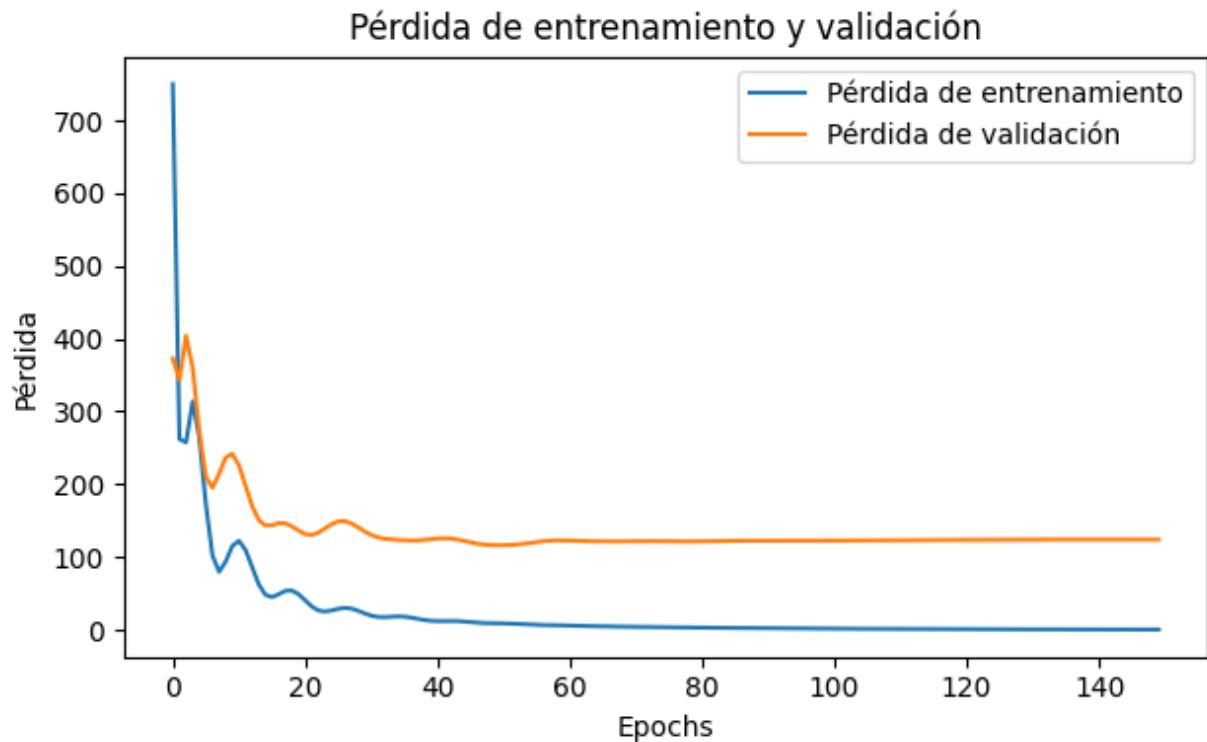


Figure 94: Pérdidas para cada época. Fuente: Elaboración Propia

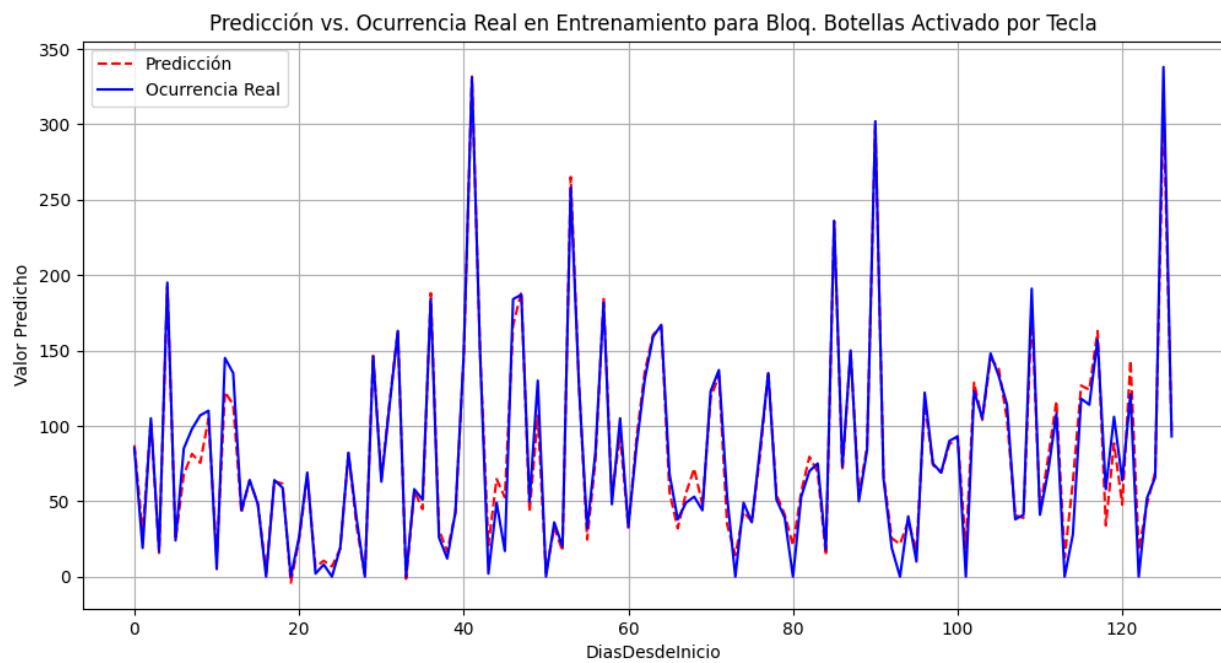


Figure 95: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

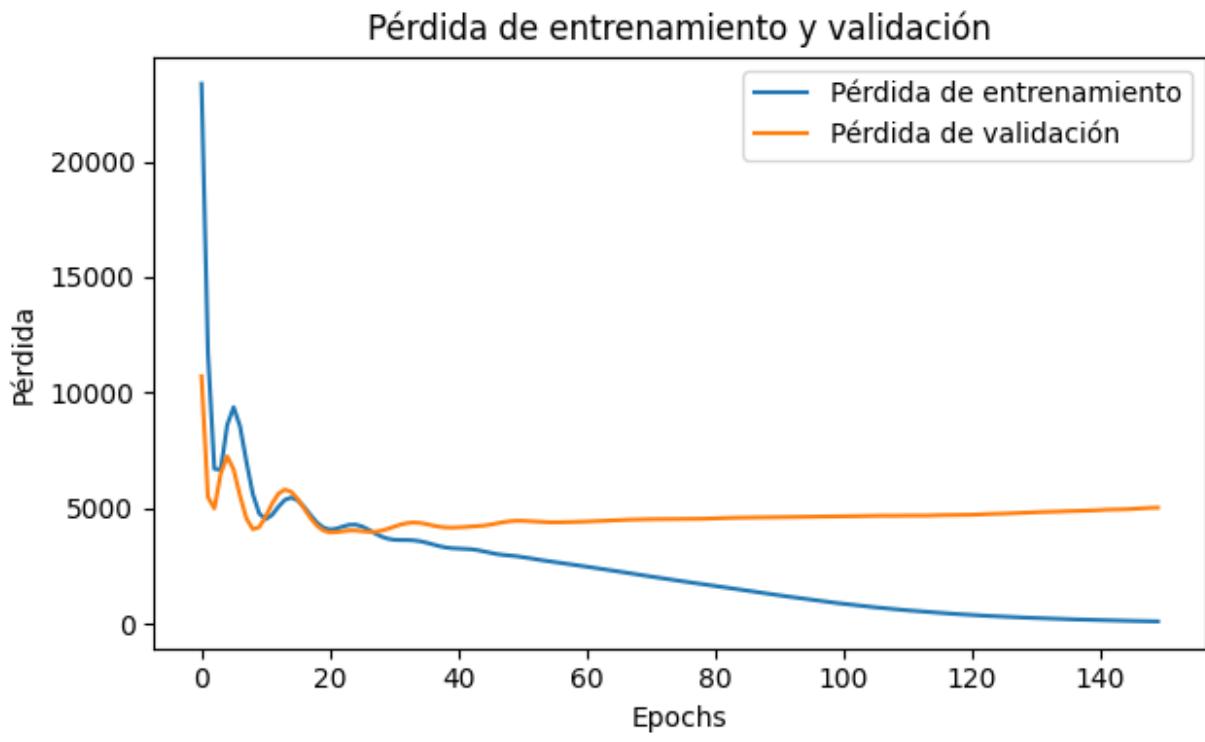


Figure 96: Pérdidas para cada época. Fuente: Elaboración Propia

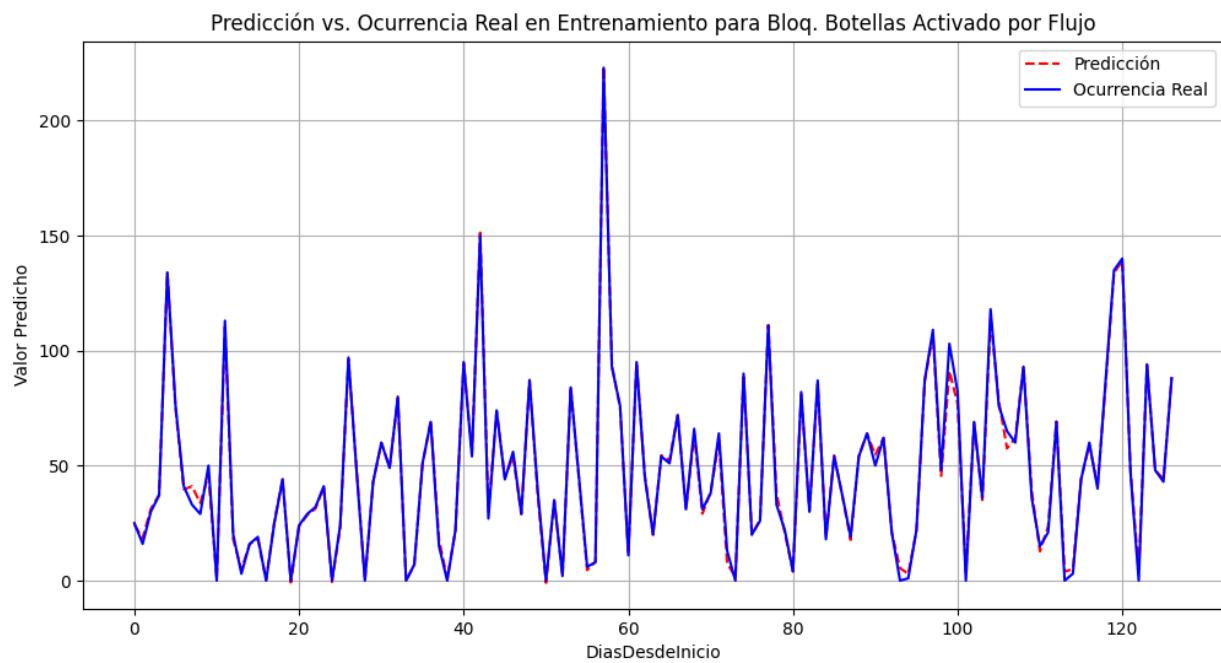


Figure 97: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

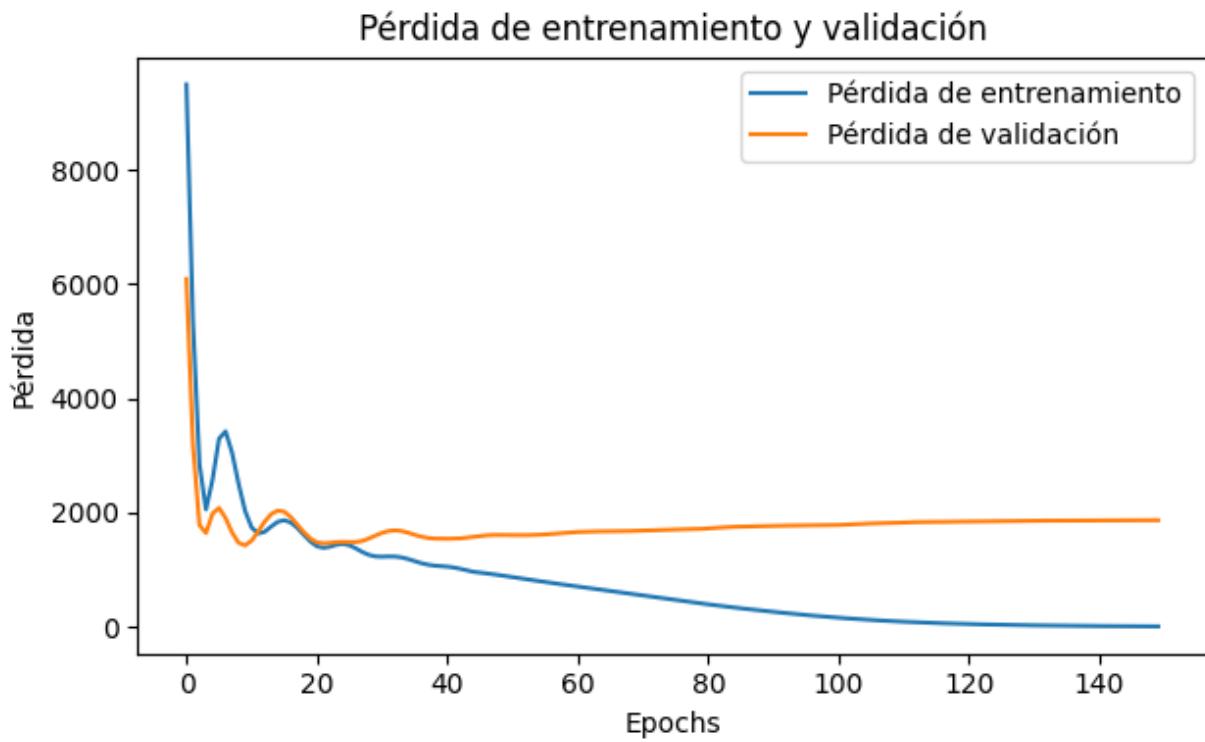


Figure 98: Pérdidas para cada época. Fuente: Elaboración Propia

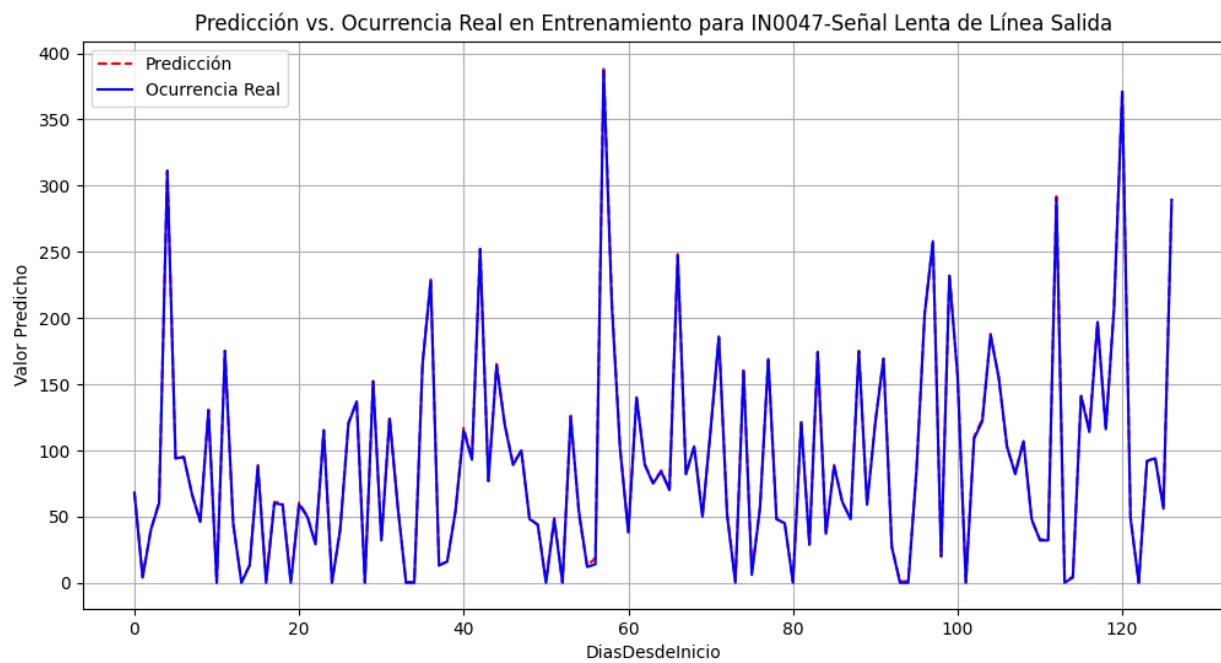


Figure 99: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

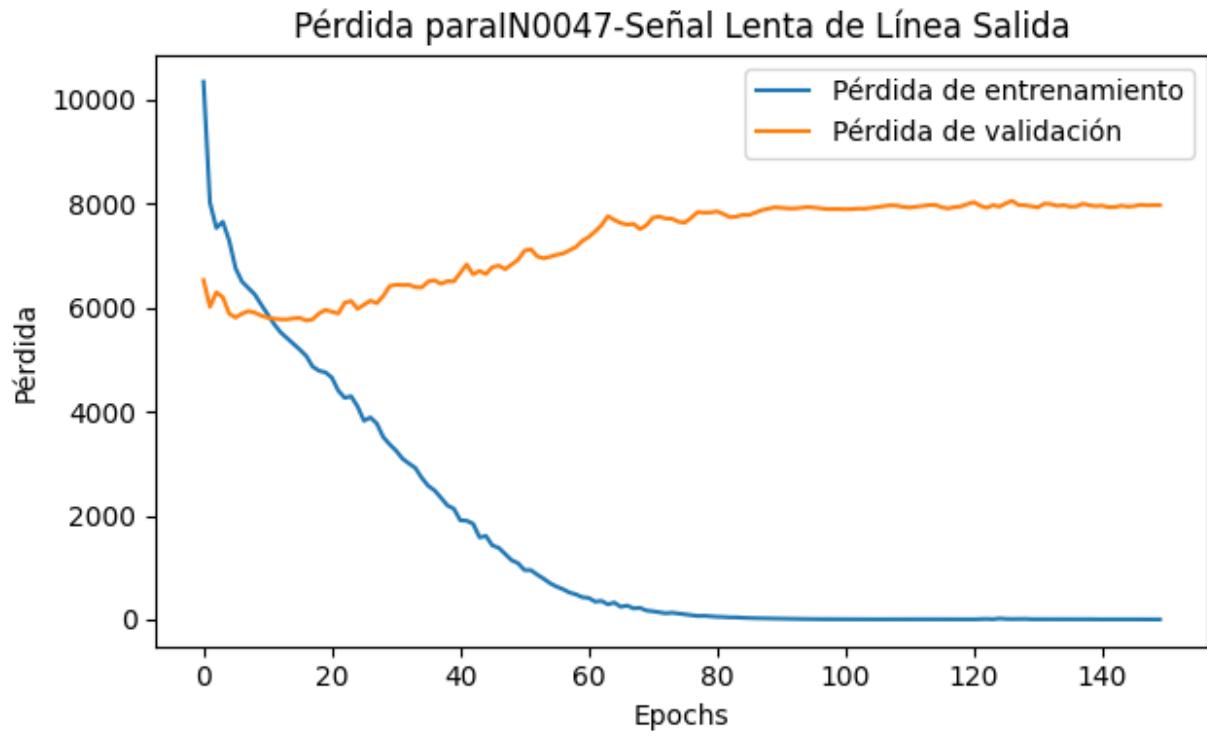


Figure 100: Pérdidas para cada época. Fuente: Elaboración Propia

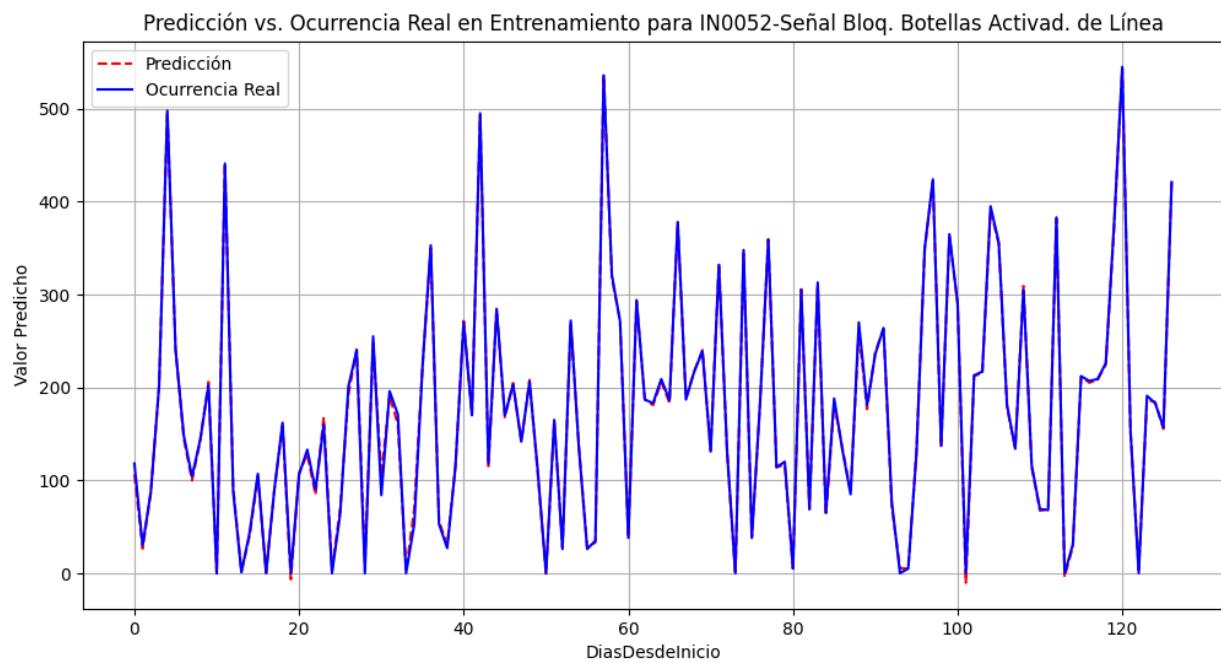


Figure 101: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

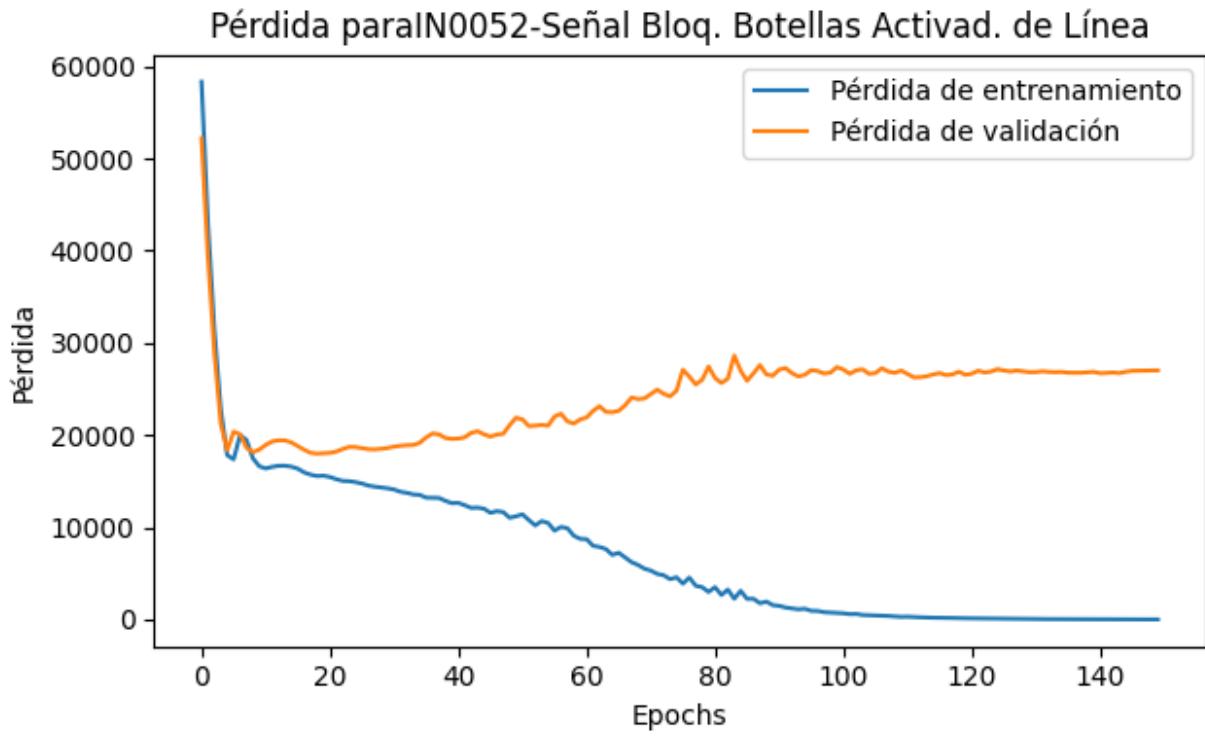


Figure 102: Pérdidas para cada época. Fuente: Elaboración Propia

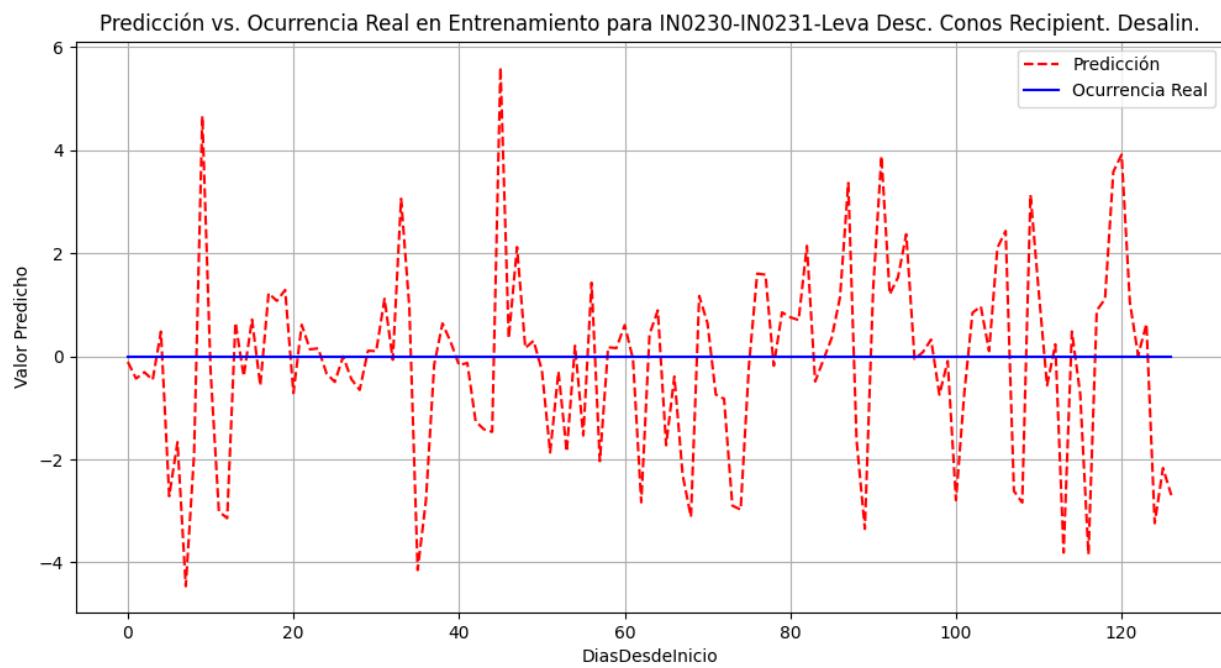


Figure 103: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

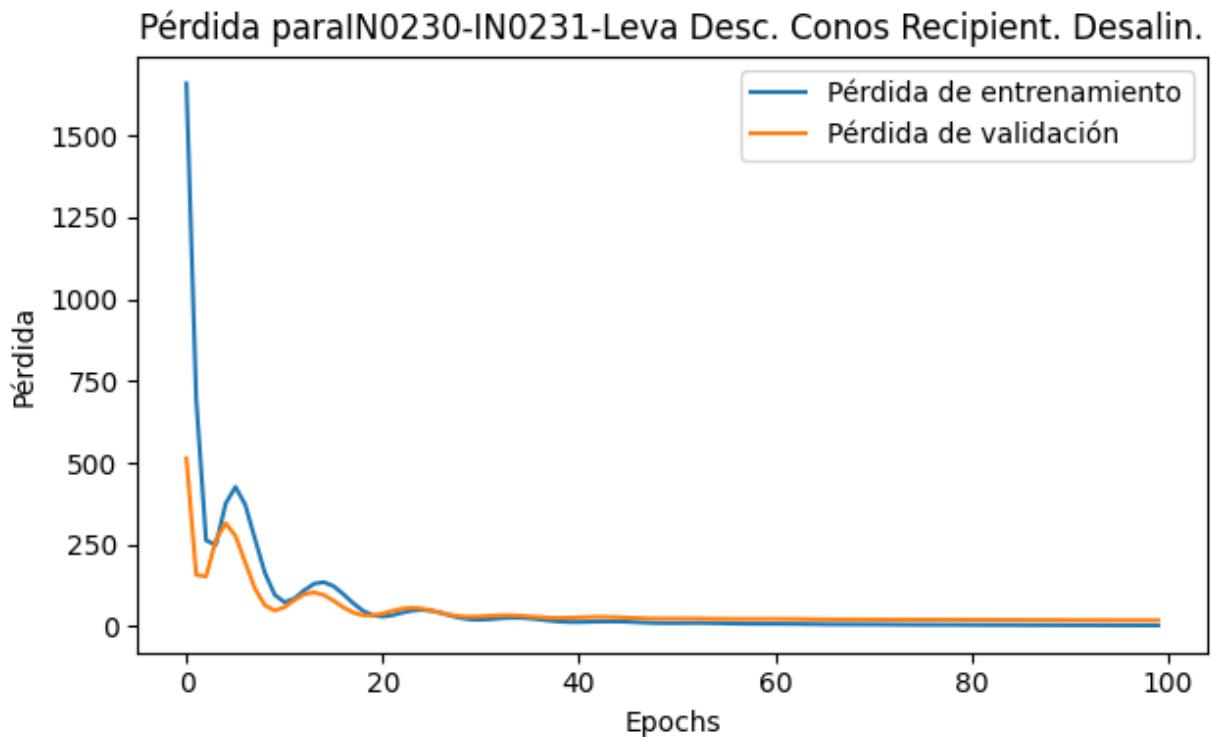


Figure 104: Pérdidas para cada época. Fuente: Elaboración Propia

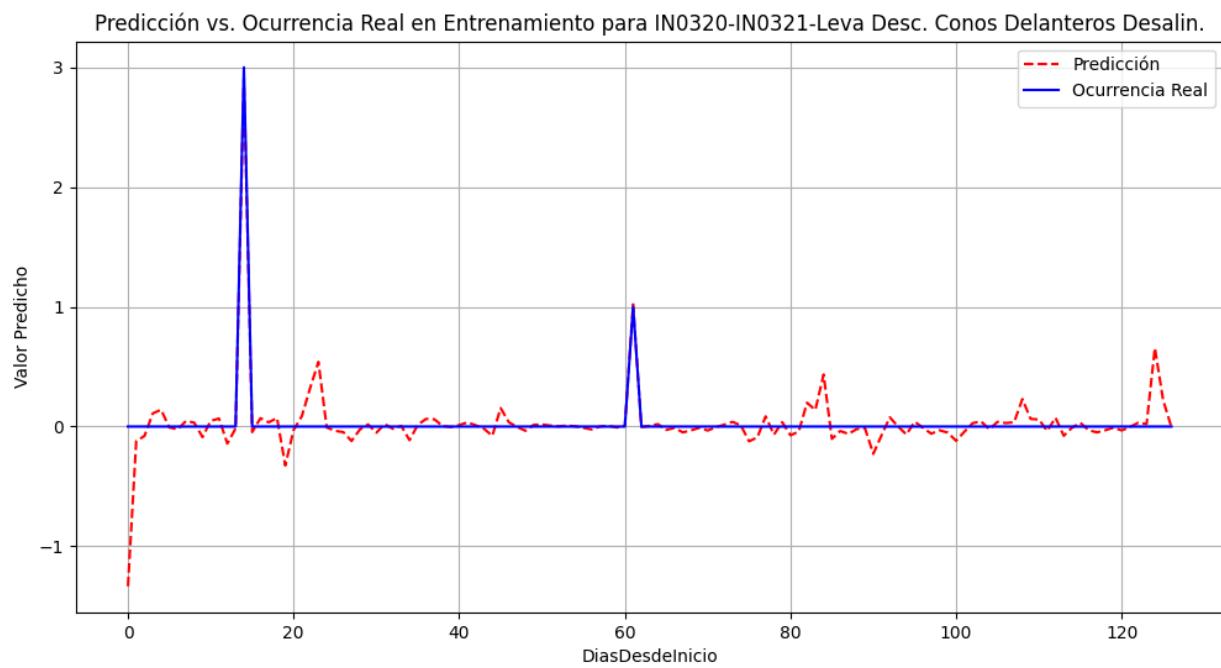


Figure 105: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

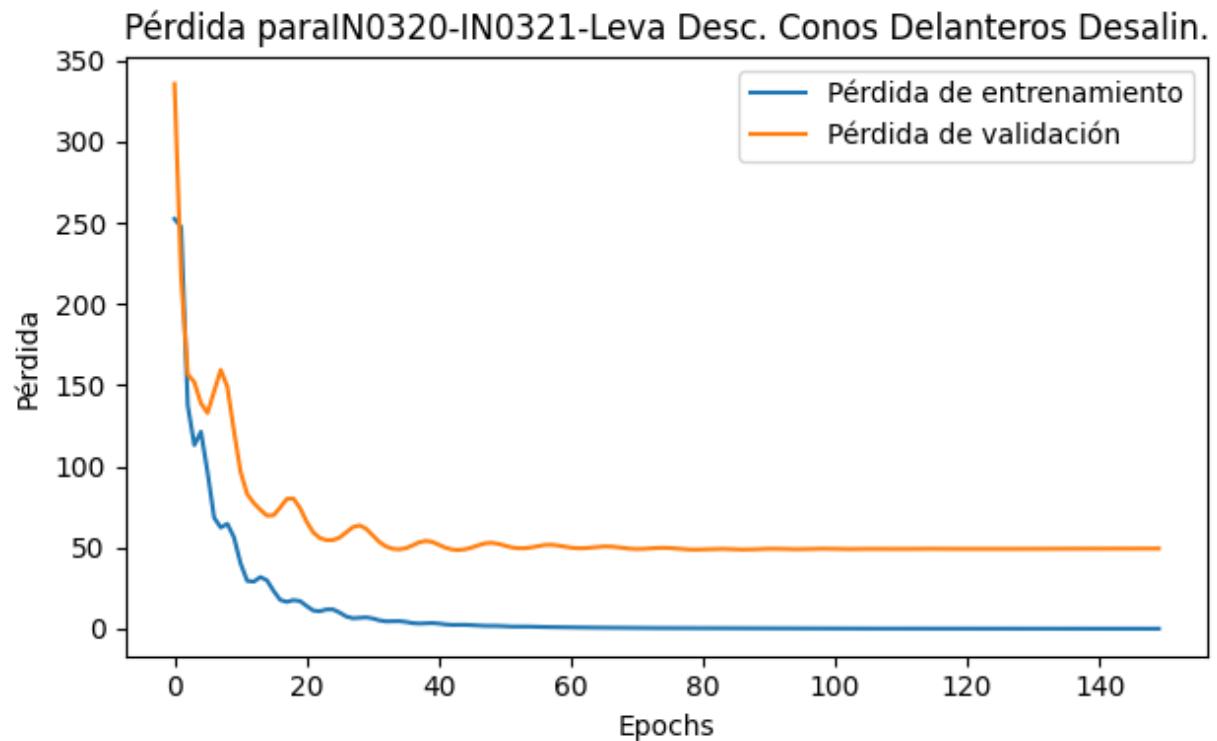


Figure 106: Pérdidas para cada época. Fuente: Elaboración Propia

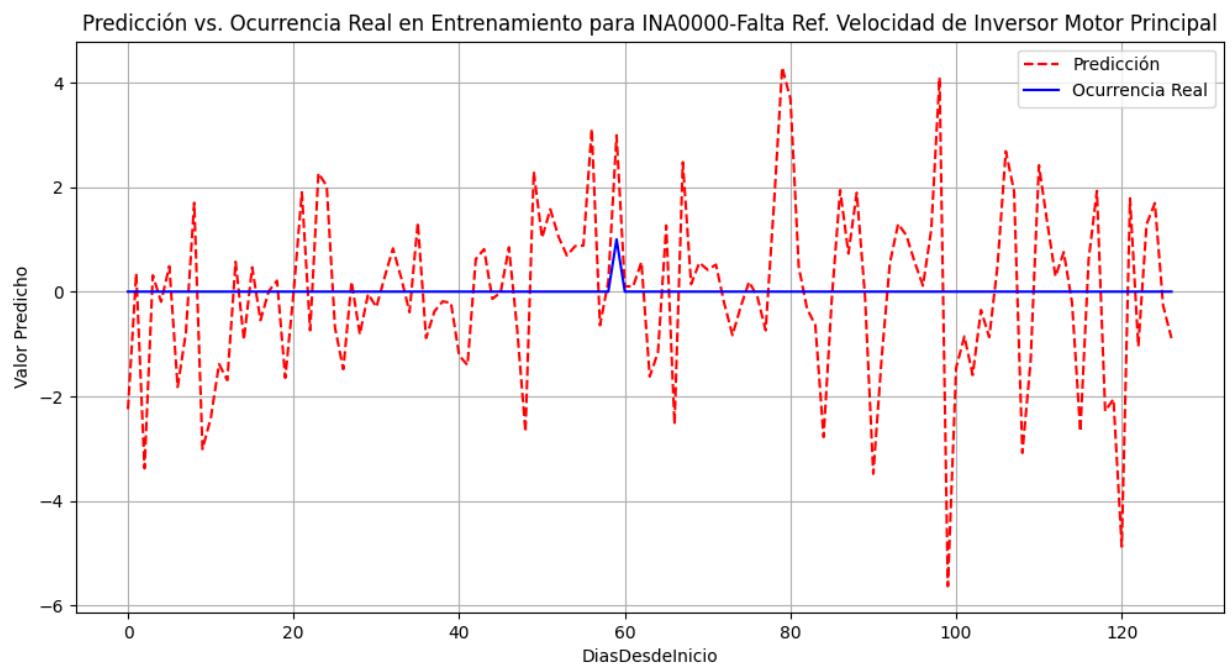


Figure 107: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

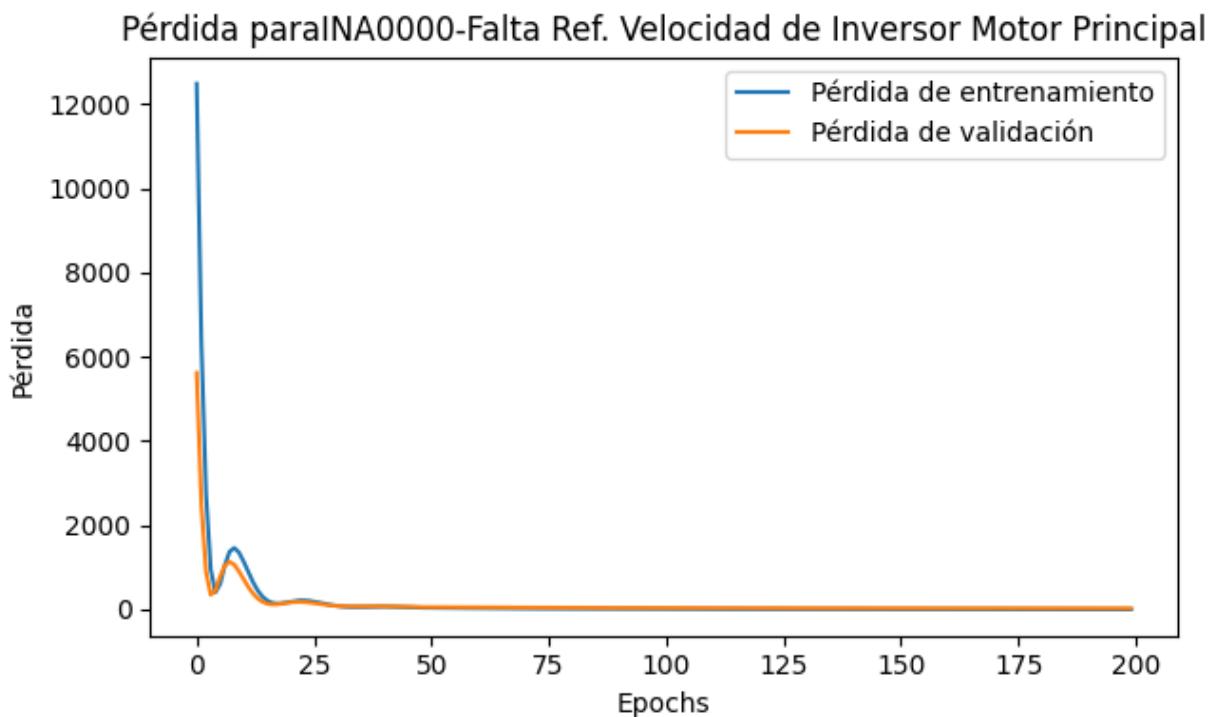


Figure 108: Pérdidas para cada época. Fuente: Elaboración Propia

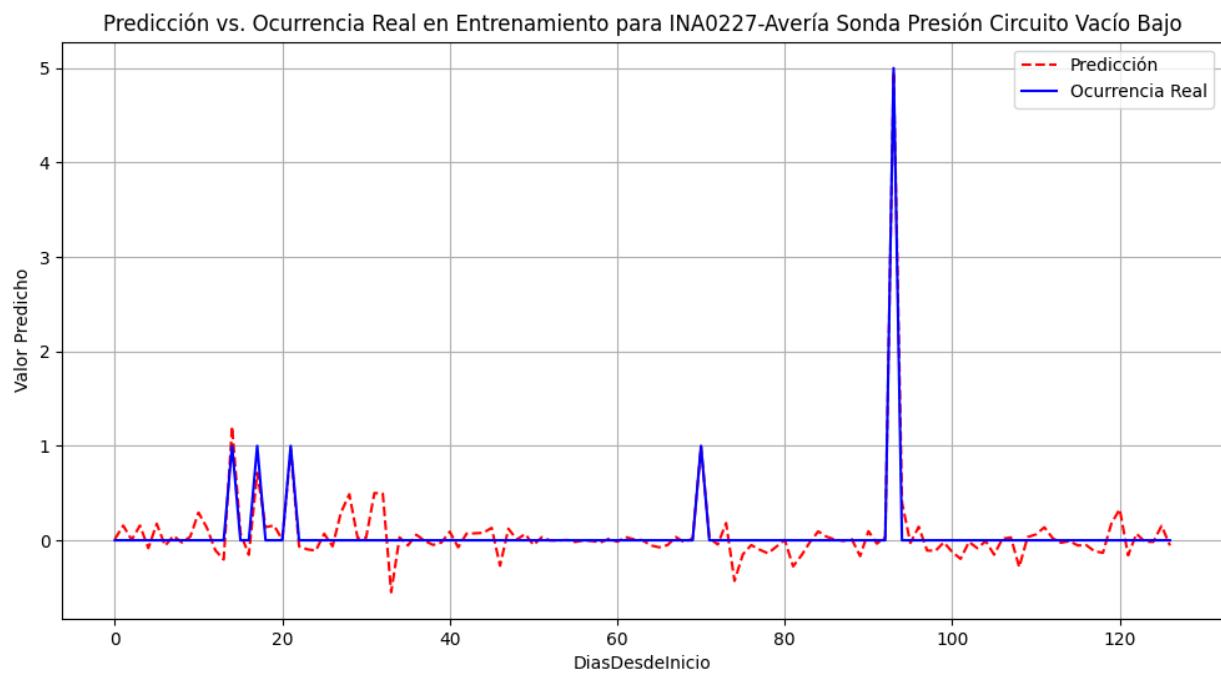


Figure 109: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

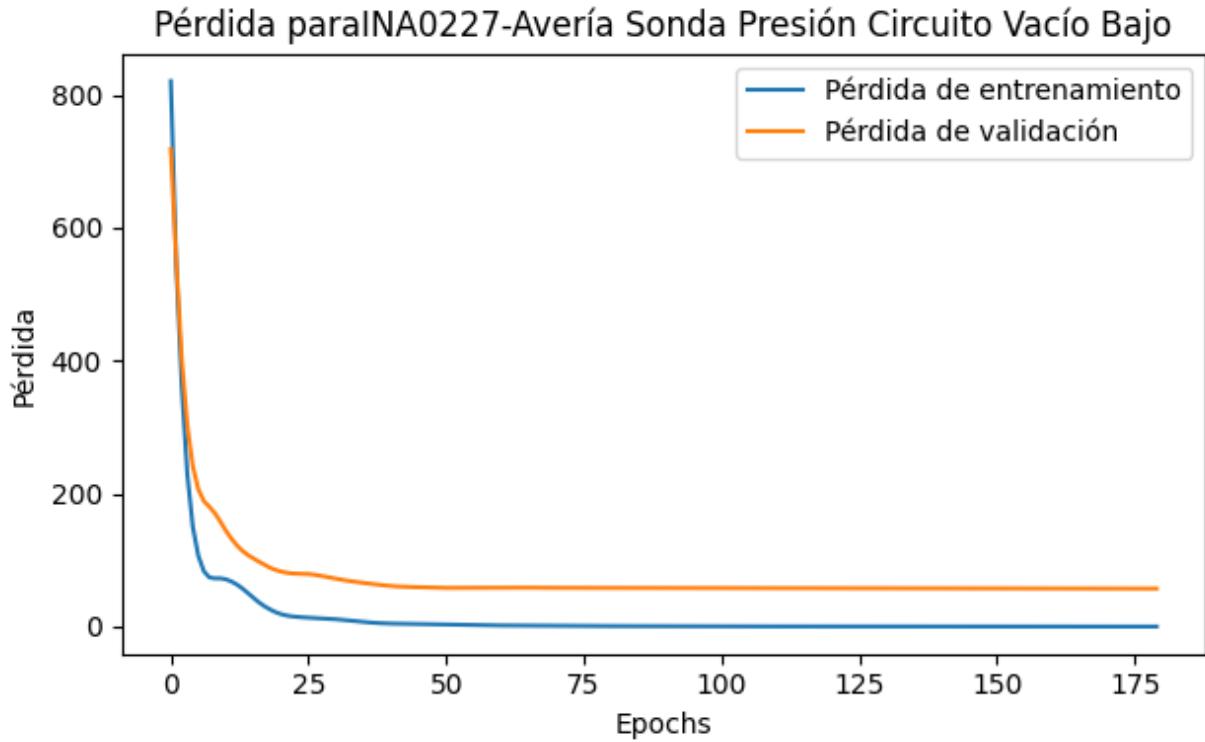


Figure 110: Pérdidas para cada época. Fuente: Elaboración Propia

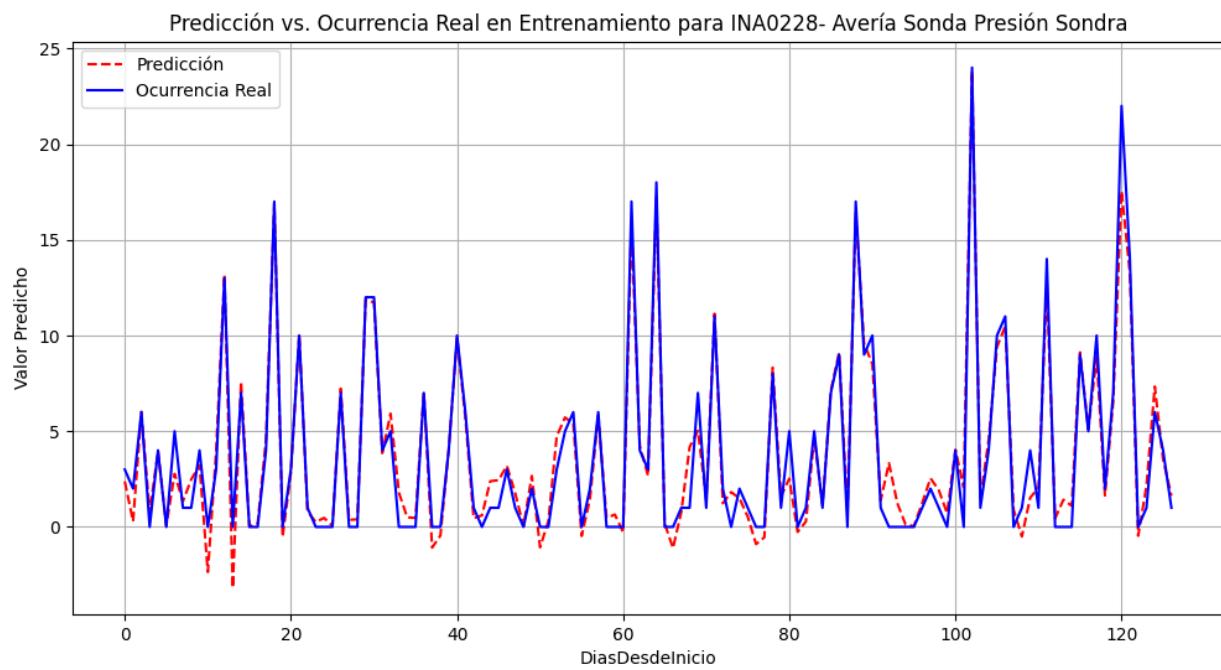


Figure 111: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

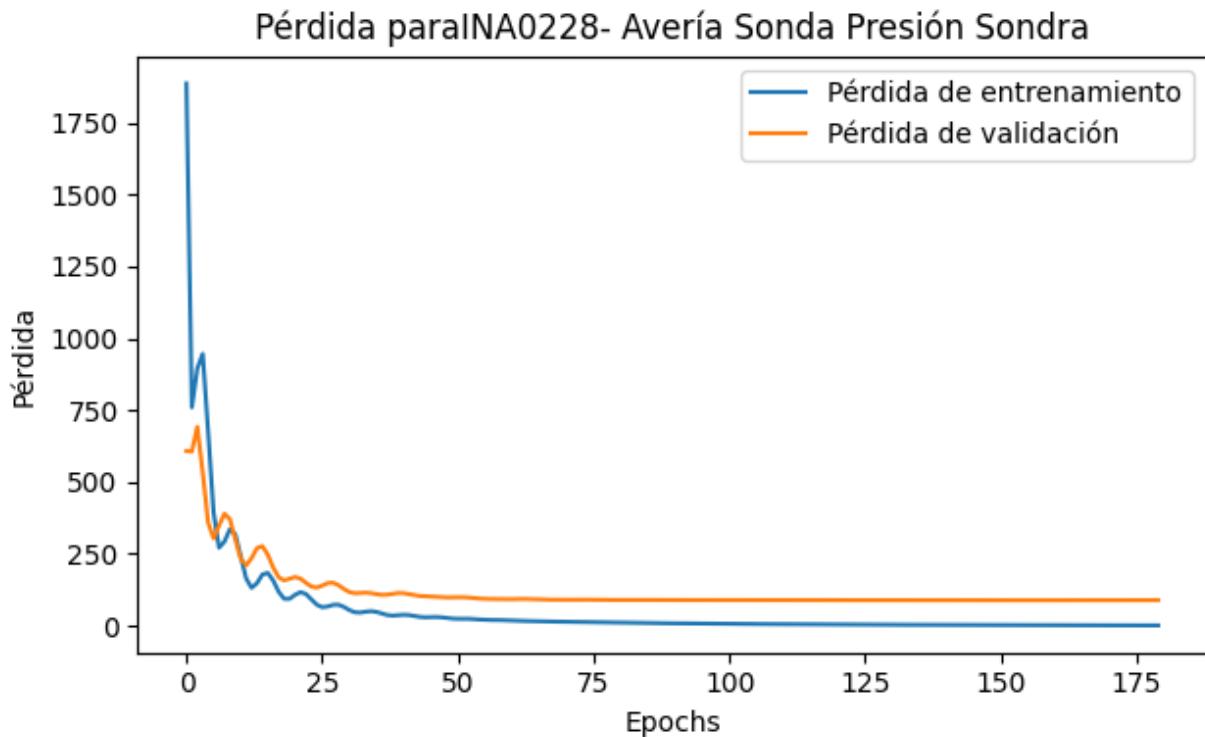


Figure 112: Pérdidas para cada época. Fuente: Elaboración Propia

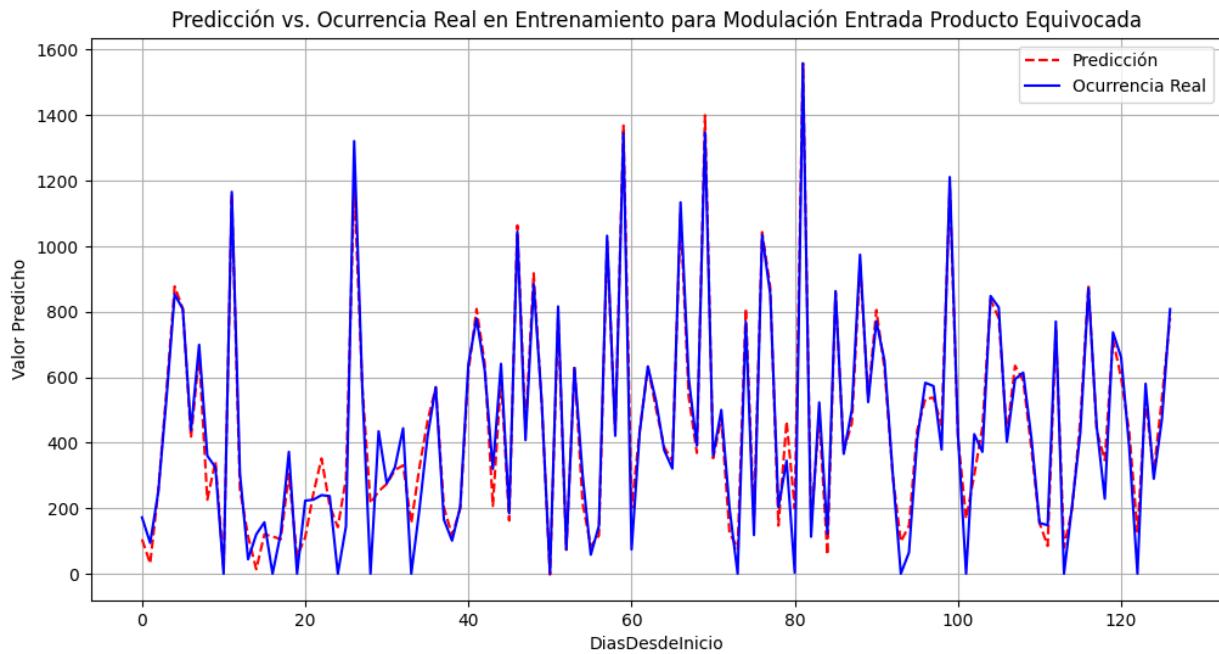


Figure 113: Predicción del modelo vs ocurrencia. Fuente: Elaboración Propia

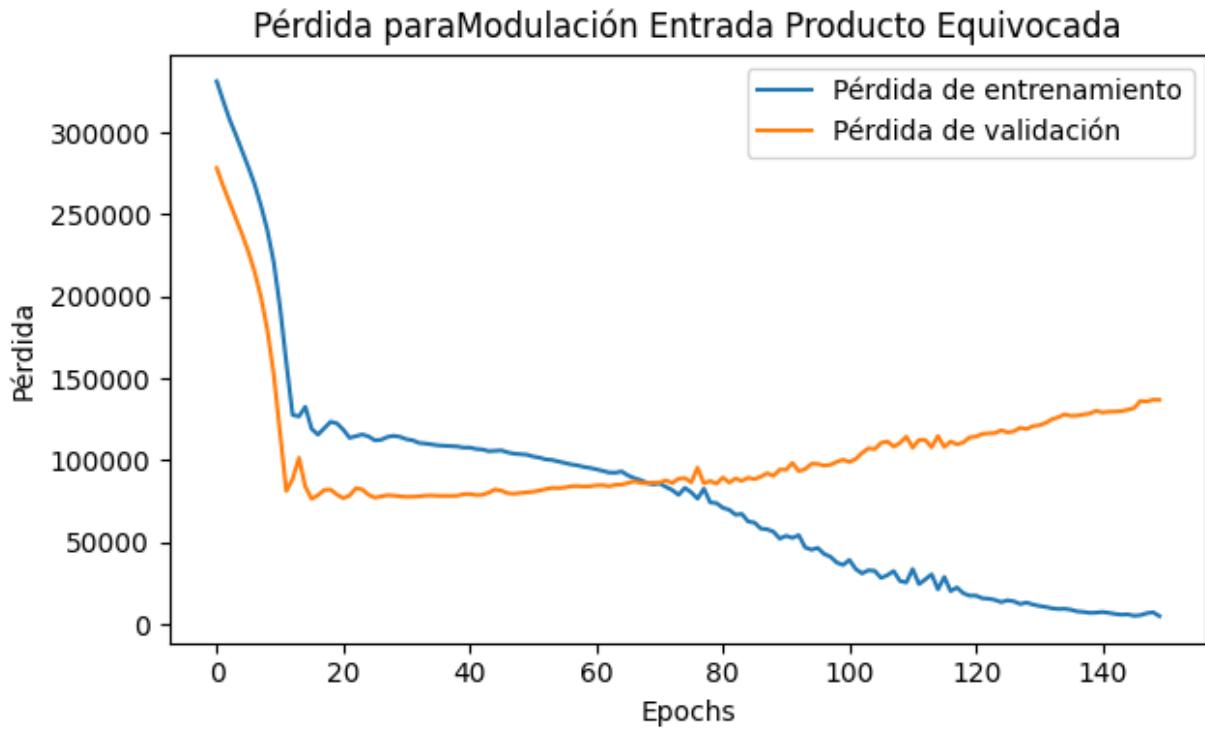


Figure 114: Pérdidas para cada época. Fuente: Elaboración Propia

## 15.13 Anexo 12 Análisis FFT

### Fast Fourier Transform

Un paso que se realizó luego de ejecutar el modelo nombrado anteriormente, fue la utilización del algoritmo Fast Fourier Transform (FFT). Este algoritmo se utiliza para encontrar frecuencias de las fallas a lo largo del

tiempo. Aplicar FFT a las fallas de la máquina, entregaría información en el dominio de la frecuencia. La interpretación de los resultados de la FFT puede proporcionar insights valiosos sobre las características de las señales de falla en la máquina. El aplicar este algoritmo puede ayudar a encontrar frecuencias dominantes en los datos. Incluir ambos algoritmos en nuestro modelo, RNN y FFT, puede permitir aprovechar la información tanto en el dominio del tiempo como en el de la frecuencia, puede mejorar la capacidad predictiva y esto podría ayudar a anticipar fallas potenciales en la maquinaria al detectar señales a partir de avisos o patrones ocultos en los datos.

## Resultados FFT

Luego de aplicar el algoritmo FFT sobre las fallas, se obtuvieron una serie de graficos, los que se analizaron para detectar aquellos que pueden ser relevantes para este problema. La siguiente serie de gráficos nos muestra la descomposición de la serie temporal sobre las fallas de “IN0574-Control Protecciones Delanteras Abiertas” (en azul) y “Maquina Desfasada” (en rojo). El primer gráfico de **Observado** muestra la serie temporal original, en cantidad de ocurrencias diarias de cada una de las fallas a lo largo del tiempo. En segundo lugar, el gráfico de **Tendencia**, que representa la dirección general de la serie de tiempo a largo plazo, es decir, la evolución de los datos sin considerar fluctuaciones de corto plazo, podría considerarse como un resumen menos detallado de los datos, para tener un panorama general de sus movimientos. Luego esta la **Estacionalidad**, que muestra los datos como un patrón repetitivo de frecuencia fija, lo que permite identificar patrones regulares y estacionales en los datos. Finalmente se obtuvo el gráfico de **Residuos**, que corresponden a aquellos datos que se pueden considerar como aleatorios, ya que quedan fuera de la tendencia y estacionalidad, se consideran como impredecibles.

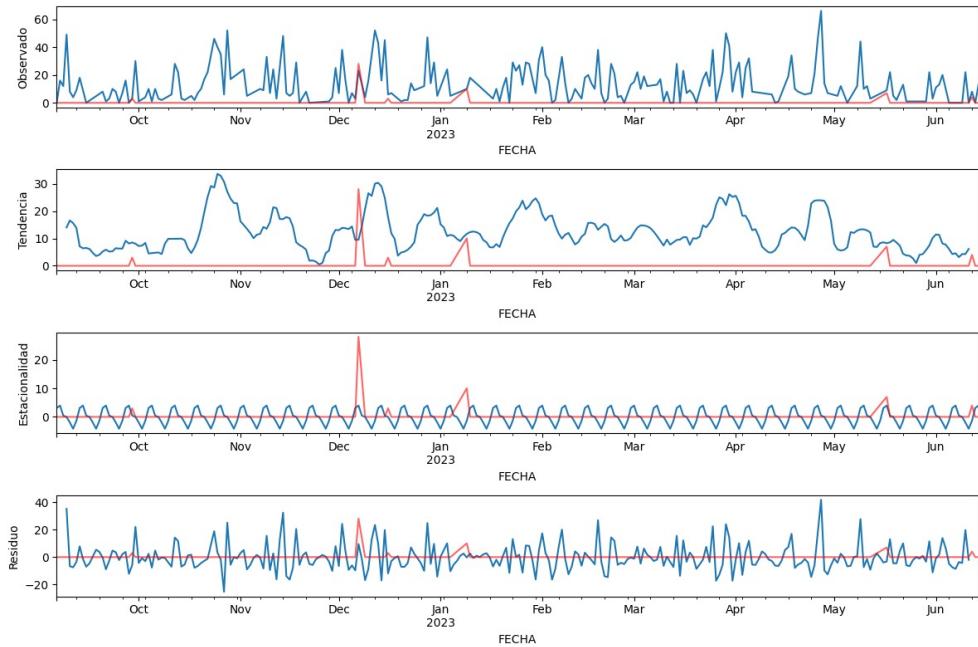


Figure 115: FFT para IN0574-Control Protecciones Delanteras Abiertas y Maquina Desfasada. Fuente: Elaboración Propia

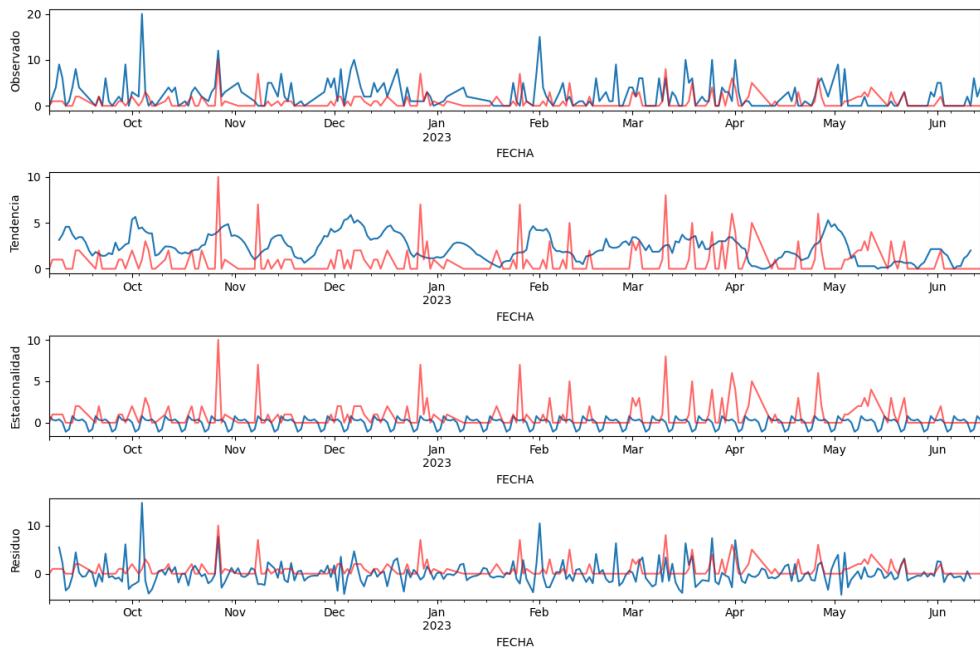


Figure 116: FFT para IN0054-Señal Ftc Presencia Botella Caída y Falta Comunicación PLC Depósito.  
Fuente: Elaboración Propia

En ambas figuras, se observa que el residuo obtenido muestra una considerable variabilidad o 'movimiento'. Esto sugiere que hay una gran cantidad de datos que parecen ser impredecibles o no están vinculados a patrones específicos de frecuencia identificables mediante la transformada rápida de Fourier (FFT). En otras palabras, aunque se realizó un análisis utilizando la FFT, no se encontraron componentes de frecuencia significativos que pudieran explicar la variabilidad en estos datos.

Este tipo de residuo nos indica la presencia de componentes aleatorios o estocásticos en los datos para efectos de este algoritmo. A pesar de haber descompuesto la serie temporal y haber intentado identificar patrones frecuentes con la FFT, parece que estos residuos no muestran una estructura periódica discernible.

Es importante resaltar que la similitud entre el residuo y los datos observados sugiere que estos datos impredecibles son una parte importante de la serie temporal. Esta falta de regularidad en el residuo podría significar que hay factores externos, ruido o variaciones aleatorias en los datos que no se pueden modelar o predecir utilizando métodos de análisis de frecuencia como FFT.