



Diagnóstico de Fallas en la Sopladora de Electrolit: Un Enfoque Basado en Análisis Exploratorio

Marissa E. Luna¹, Ximena A. Cantón¹, Nubia G. Barajas¹ and Mariana L. Maldonado¹

¹ Instituto Tecnológico y de Estudios Superiores de Monterrey, Escuela de Ingeniería y Ciencias, Guadalajara, Jalisco

Abstract— El proceso de fabricación de botellas de Electrolit enfrenta fallas prematuras en la sopladora, cuya vida útil se ha reducido de 12 a 2-3 meses. Mediante un Análisis Exploratorio de Datos (EDA), se identificaron patrones anómalos en variables como temperatura, presión y consumo energético. Se utilizaron técnicas estadísticas y algoritmos de detección de anomalías para encontrar posibles causas de las fallas. Los resultados señalaron fluctuaciones térmicas extremas y variaciones abruptas en la presión como factores críticos. Este análisis sentó las bases para optimizar el mantenimiento y operación de la sopladora.

Keywords— Análisis Exploratorio de Datos, Detección de Anomalías, Industria Farmacéutica, Electrolit, Sopladora, Temperatura, Presión, Machine Learning, Calidad de Producción, Análisis de Series Temporales, Algoritmos de Clustering.

I. INTRODUCCIÓN

En el proceso de fabricación de botellas de Electrolit, la sopladora es un componente fundamental, ya que se encarga de inflar adecuadamente las botellas para que cumplan con los estándares de calidad. Sin embargo, actualmente este equipo está presentando una vida útil de entre 2 y 3 meses, en lugar de los 12 meses esperados, lo que está generando pérdidas económicas significativas. Debido a esto, se ha proporcionado un conjunto masivo de datos con información generada por la máquina para analizar posibles causas de esta falla prematura.

El objetivo de este análisis exploratorio de datos (EDA) es comprender el comportamiento de las variables registradas por la sopladora, identificar patrones anómalos y encontrar correlaciones entre distintos factores que podrían estar contribuyendo a la reducción de su vida útil. No se busca en este punto generar un modelo predictivo ni proponer soluciones definitivas, sino obtener hallazgos clave que permitan dirigir futuras investigaciones y esfuerzos de optimización.

II. DESCRIPCIÓN DE LOS DATOS

Para este estudio, se recibieron dos conjuntos de datos en formato Parquet, correspondientes a los meses de noviembre y enero. Dado que cada archivo contenía información de eventos registrados minuto a minuto, el volumen total de datos era considerablemente alto. Después de combinar ambos archivos, se obtuvo una base con 90,798,747 observaciones. Este tamaño de datos implica desafíos en términos de almacenamiento y procesamiento, por lo que fue necesario aplicar filtros y estrategias de reducción antes de realizar un análisis detallado.

Dado que la base de datos contenía cientos de variables registradas por la sopladora, se realizó una selección preliminar de 28 variables, las cuales fueron elegidas de manera aleatoria pero garantizando que representaran distintos as-

pectos del funcionamiento de la máquina, como consumo de energía, temperatura y presión. Sin embargo, durante el proceso de limpieza, se identificó que 8 de estas variables contenían menos del 0.1 % de los registros, lo que las hacía poco representativas y llevaron a su eliminación. Como resultado de este proceso, se trabajó finalmente con 20 variables en el análisis inicial.

Otro aspecto clave del preprocesamiento fue la corrección del timestamp, ya que se solicitó ajustar los datos restando 6 horas para asegurar que los registros estuvieran alineados correctamente con la zona horaria en la que opera la máquina. Esto es fundamental, ya que una incorrecta alineación temporal podría generar inconsistencias al comparar eventos y tendencias.

III. PREPROCESAMIENTO DE DATOS

Debido al gran volumen de datos, se implementó una estrategia basada en división por chunks, en la que los datos fueron divididos en fragmentos de 100,000 registros para facilitar su procesamiento sin sobrecargar la memoria. Posteriormente, se aplicaron distintas funciones para evaluar la relevancia de cada variable en el contexto del problema, entre ellas:

1. Cálculo de Variabilidad: Se identificaron variables con valores prácticamente constantes, ya que aquellas sin variación significativa no aportan información útil para el análisis.
2. Detección de Outliers: Se utilizaron métodos como el rango intercuartil (IQR) para identificar valores atípicos que podrían estar indicando fallos en la máquina.
3. Distribución de Variables: Se evaluó la distribución de cada variable clave para comprender sus patrones normales de comportamiento.
4. Proporción de Observaciones Útiles: Se eliminó

cualquier variable con una cantidad insignificante de datos disponibles.

Tomando en cuenta lo anterior, pudimos descartar de esa manera un total de 12 variables, por ejemplo, hay variables que tienen una desviación estándar muy baja y un IQR de 0.0, lo que significa que prácticamente todos los valores son iguales y no aportan información (gran variedad de columnas de la variable: "CONTIFORM_MMA_CONTIFORM_MMA1_AirWizard-BasicController_0").

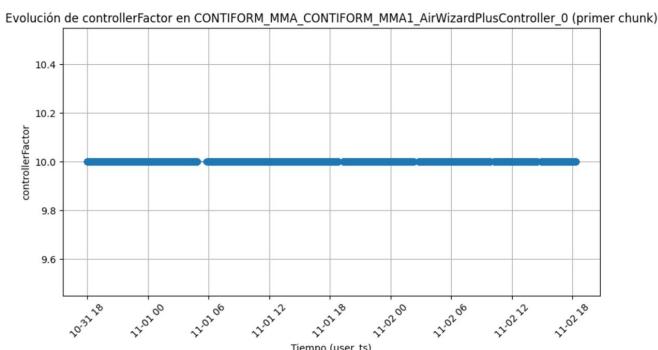


Fig. 1: Serie de tiempo de AirWizard en el primer chunk.

Por otro lado, variables que tienen una proporción de datos de solo 0.0011 (0.1%) del total, lo que significa que aparecen en muy pocas observaciones y podrían no ser representativas. De esa manera definimos el nuevo número de variables relevantes con las que trabajaremos, variables que tienen una desviación estándar razonable, un IQR significativo y suficientes datos para ser útiles:

- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_1
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_3
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_5
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_7
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_9
- CONTIFORM_MMA_CONTIFORM_MMA1_FinalBlowingPressureActualValue_0
- CONTIFORM_MMA_CONTIFORM_MMA1_PressureCompensationChamberPressureActualValue_0
- CONTIFORM_MMA.CONTIFORM_MMA1.WS_Tot_Rej.0

A partir de estas, se generaron distintos archivos CSV con información procesada, permitiendo realizar un análisis detallado de cada una.

IV. ANÁLISIS EXPLORATORIO DE DATOS

a. Análisis Descriptivo Inicial

Antes de analizar relaciones complejas o detectar anomalías, se llevó a cabo un análisis descriptivo inicial para comprender la naturaleza de las variables seleccionadas. Se calcularon medidas estadísticas clave como:

- Media y mediana para observar la tendencia central de los datos.
- Desviación estándar para evaluar la dispersión de las mediciones.
- Valores mínimos y máximos para detectar posibles errores en la recolección de datos.
- Percentiles para analizar la distribución de las variables.

b. Distribución de Datos y Detección de Outliers

Para evaluar la distribución de las variables, se generaron histogramas que permitieron visualizar la frecuencia de distintos valores y detectar sesgos en la información. Además, se utilizaron boxplots para identificar la presencia de outliers, es decir, valores que se encuentran significativamente fuera del rango típico de los datos.

Para cuantificar los outliers de manera precisa, se emplearon dos enfoques estadísticos:

1. Z-score, que mide cuántas desviaciones estándar se encuentra un dato respecto a la media.
2. IQR (Interquartile Range), que permite identificar valores atípicos con base en los percentiles 25 y 75 de la distribución.

La presencia de outliers podría estar indicando fallos operativos en la sopladora, por lo que su identificación es clave para entender por qué la máquina está fallando antes del tiempo esperado.

c. Análisis Temporal y Tendencias

Dado que la sopladora opera de manera continua, es fundamental analizar cómo evolucionan sus mediciones a lo largo del tiempo. Para ello, se generaron series de tiempo, en las que se observaron las variaciones de temperatura, presión y consumo energético durante el periodo de estudio. Además, se aplicaron técnicas de rolling mean y rolling standard deviation para identificar anomalías en la variabilidad de las mediciones.

Para detectar posibles patrones repetitivos en los datos, se utilizaron funciones de autocorrelación (ACF y PACF), que permiten identificar si los valores actuales de una variable están influenciados por eventos ocurridos en el pasado. Esto es particularmente útil para entender si las fallas en la sopladora siguen un patrón cíclico o aleatorio.

d. Correlaciones y Relaciones entre Variables

Con el fin de evaluar cómo se relacionan las distintas mediciones de la máquina, se generó una matriz de correlación



(heatmap), en la que se visualizaron las dependencias entre temperatura, presión y fallos en la sopladora. Se complementó este análisis con gráficos de dispersión (scatter plots) y análisis de pares (pairplots) para detectar relaciones no lineales.

Este análisis es crucial para determinar si existe una relación entre las variables registradas y la reducción de la vida útil de la sopladora. Por ejemplo, si se encuentra una alta correlación entre temperatura y fallos, se podría hipotetizar que un sobrecalentamiento está contribuyendo al desgaste prematuro del equipo.

e. Detección de Anomalías Basada en Modelos

Para complementar el análisis exploratorio, se emplearon algoritmos de detección de anomalías, como:

- DBSCAN, útil para detectar outliers en datos de alta dimensión. [1]
- Isolation Forest, que identifica valores atípicos basándose en la estructura del conjunto de datos.
- Local Outlier Factor (LOF), que permite encontrar anomalías en series temporales.

Estos métodos proporcionan una visión más avanzada de los patrones anormales en la operación de la sopladora. [2]

V. RESULTADOS DE EDA

Calidad de datos:

La columna “message” es completamente nula para todas las layers de temperatura, por lo que puede eliminarse sin afectar el análisis.

Layer 1:

- Total de registros: 884,229
- Valores nulos eliminados: 1 (en la columna value)
- Valores duplicados: 10,512

Layer 3:

- Valores duplicados: 5,118

Layer 5:

- Valores duplicados: 7,039

Layer 7:

- Total de registros: 880,395
- Valores nulos eliminados: Ninguno
- Valores duplicados: 4,512

Layer 9:

- Total de registros: 2,016,295
- Valores nulos eliminados: Ninguno
- Valores duplicados: 10,512

Presión final de soplado:

- Total de registros: 1,512,753
- Valores nulos eliminados: 3 (en la columna value)
- Valores duplicados: 8,341

Presión de la cámara de compensación:

- Total de registros: 1,233,448
- Valores nulos eliminados: 3 (en la columna value)
- Valores duplicados: 4,012

Botellas rechazadas:

- Valores nulos eliminados: 3 (en la columna value)
- Valores duplicados: 727
- Conversión a formato de datetime para análisis temporal.

Estadística descriptiva:

Layer 1:

Media	112.6°C
Mediana (Q2)	123.0°C
Desviación Estándar	27.58°C
Mínimo	0.0°C
Máximo	168.0°C
Percentil 25 (Q1)	122.0°C
Percentil 75 (Q3)	124.0°C

La mayoría de los valores están entre 120°C y 125°C (posible rango óptimo). Hay temperaturas inusualmente bajas (0°C) y altas (>130°C), lo que indica falta de control térmico en algunos momentos.

Layer 3:

Media	118.33°C
Mediana (Q2)	127.0°C
Desviación Estándar	17.33°C
Mínimo	0.0°C (valor sospechoso)
Máximo	449.0°C

Valores bajos de temperatura (cerca de 0) pueden indicar fallas en la medición o momentos de inactividad.

Layer 5:

Media	116.89°C
Mediana (Q2)	124°C
Desviación Estándar	25.46°C
Mínimo	16.0°C
Máximo	449.0°C

Menos variabilidad en los datos comparado con Layer 3, pero aún presenta valores extremos.

Layer 7:

Media	105.55°C
Mediana (Q2)	116.0°C
Desviación Estándar	29.63°C
Mínimo	14.0°C
Máximo	449.0°C
Percentil 25 (Q1)	114.0°C
Percentil 75 (Q3)	118.0°C

Layer 9:

Media	96.97°C
Mediana (Q2)	102.0°C
Desviación Estándar	17.33°C
Mínimo	14.0°C
Máximo	449.0°C
Percentil 25 (Q1)	97.0°C
Percentil 75 (Q3)	104.0°C

Valores extremadamente altos (hasta 449°C): Posible error en el sensor o sobrecalentamiento crítico.

Presión final de soplado:

Media	11.96245
Mediana (Q2)	12.0
Desviación Estándar	0.6386692
Mínimo	0.0
Máximo	15.3
Percentil 25 (Q1)	11.9
Percentil 75 (Q3)	12.0

Presión de la cámara de compensación:

Media	16.11
Mediana (Q2)	16.0
Desviación Estándar	0.688
Mínimo	0.0
Máximo	17.2
Percentil 25 (Q1)	15.8
Percentil 75 (Q3)	16.7

Botellas rechazadas:

Media	3000
Mediana (Q2)	2435
Desviación Estándar	Alta variabilidad
Valores extremos:	18,170

Distribuciones**Layer 1:**

Se observan caídas drásticas a 50°C y valores altos >130°C.

Impacto en la Sopladora: Caídas térmicas pueden impedir que la preforma se expanda correctamente, generando botellas defectuosas y más rechazos.

Layer 3 y Layer 5:

En ambas capas, la mayoría de los valores están altamente concentrados entre 100 y 130. Sin embargo, hay valores extremos en los rangos 0 y 400+, lo cual puede ser anómalo.

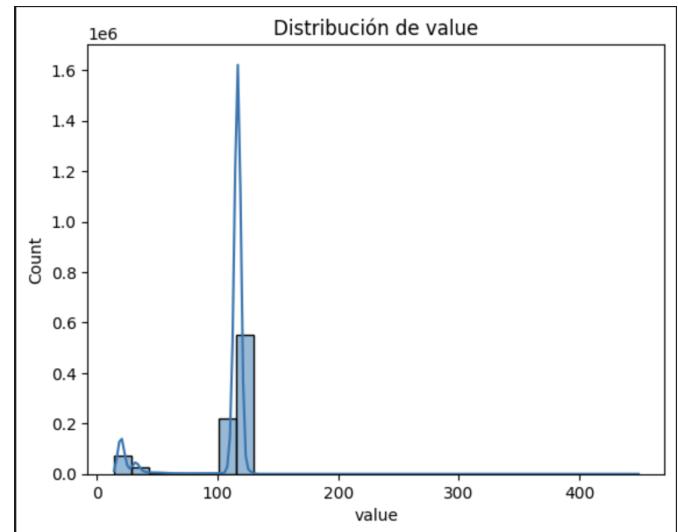
Layer 7:

Fig. 2: Distribución de Temperatura Layer 9.

- Se observan dos o más picos claros en la distribución, lo que sugiere que la variable no sigue una distribución normal.
- Hay algunos valores que llegan a más de 400°C, lo que se encuentra lejos de la mayoría de las observaciones.
- El pico más alto se encuentra alrededor de los 100°C, lo que indica que la mayoría de los valores de temperatura están concentrados en ese rango.

Layer 9:

Pico central en 100°C, pero con una cola larga hacia valores más altos. Valores muy altos (>400°C) pueden ser anomalías graves.

Impacto en la Producción: Valores anómalos en 400°C podrían deteriorar la calidad de la preforma. Si hay valores más bajos que 50°C, la botella no se inflará correctamente.

Presión final de soplado:

La mayoría de los valores están concentrados en 12. Hay una distribución altamente sesgada, con algunos valores más allá de 14 y menores a 10. La densidad de datos muestra picos muy marcados, lo que podría indicar mediciones discretas o sensores con límites fijos.

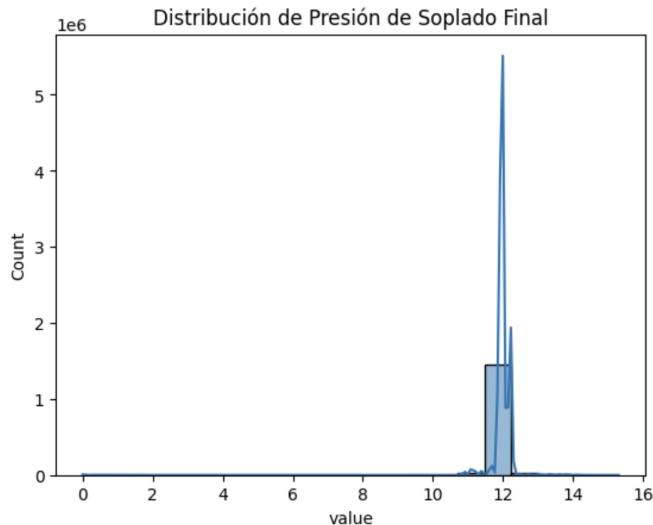


Fig. 3: Muestra la distribución asimétrica y la presencia de valores altos poco frecuentes.

Presión de la cámara de compensación:

Hay una alta cantidad de outliers por debajo de 12 y hasta valores cercanos a 0. La línea central (mediana) está en 12, confirmando que la mayoría de los datos están en ese rango. Los outliers pueden indicar mediciones anómalas o fallas en la presión

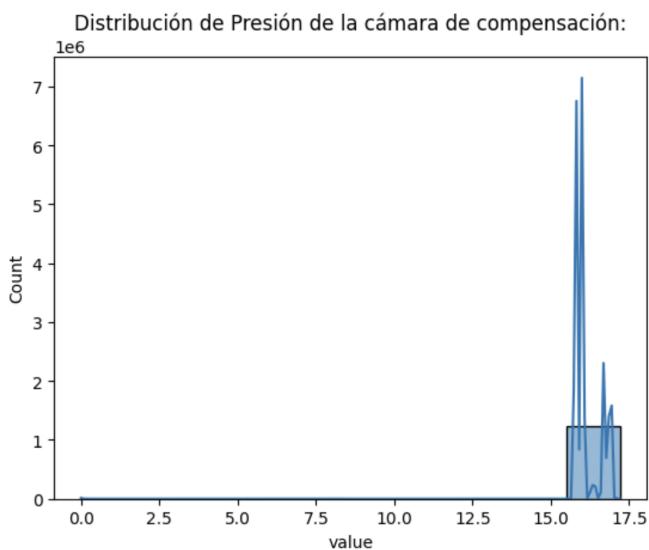


Fig. 4: Muestra la distribución asimétrica y la presencia de valores altos poco frecuentes.

Botellas rechazadas:

Se generó un histograma para analizar la distribución de botellas rechazadas. Se observó que

- Una distribución asimétrica a la derecha.
- Una mayor concentración de valores bajos y una cola larga hacia valores altos.

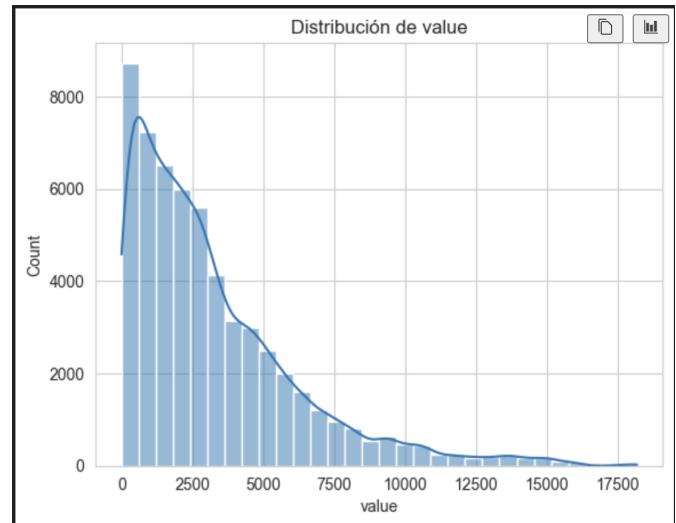


Fig. 5: Distribución asimétrica sesgada a la derecha.

Análisis de Outliers

Layer 1:

Dos tipos de outliers:

- Temperaturas bajas ($<50^{\circ}\text{C}$) → Problemas en el sistema de calentamiento.
- Temperaturas altas ($>130^{\circ}\text{C}$) → Riesgo de degradación del plástico.

Layer 3 y 5:

Dos tipos de outliers: En ambas capas, la mayoría de los valores están altamente concentrados entre 100°C y 130°C . Sin embargo, hay valores extremos en los rangos 0°C y $+400^{\circ}\text{C}$, lo cual puede ser anómalo.

Boxplots

- Se observan valores muy altos (>400) que deben ser investigados.
- Los valores de temperatura en 0 también parecen ser outlier
- Si los valores atípicos coinciden con períodos de falla de la máquina, pueden estar relacionados con su vida útil reducida.

Z-score (Outliers detectados)

- Layer 3: 70,318 valores atípicos
- Layer 5: 83,599 valores atípicos.

Cantidad considerable de valores atípicos, lo que sugiere instabilidad en la temperatura.

Layer 7:

- Valores atípicos se encuentran en el extremo superior de la distribución, alcanzando hasta 400°C o más.

Normalmente, la temperatura óptima para el proceso de soplado de botellas no debería alcanzar estos valores, lo que indica una posible anomalía en la regulación térmica de la máquina.

Se ha observado que los valores más altos de temperatura coinciden con un mayor número de botellas rechazadas, lo que sugiere que este fenómeno afecta la calidad del producto final.

Layer 9:

Pico central en 100°C, pero con una cola larga hacia valores más altos. Valores muy altos ($>400^\circ\text{C}$) pueden ser anomalías graves.

Impacto en la Producción: Valores anómalos en 400°C podrían deteriorar la calidad de la preforma. Si hay valores más bajos que 50°C, la botella no se inflará correctamente.

Presión final de soplado:

Hay una alta cantidad de outliers por debajo de 12 y hasta valores cercanos a 0. La línea central (mediana) está en 12, confirmando que la mayoría de los datos están en ese rango. Los outliers pueden indicar mediciones anómalas o fallas en la presión

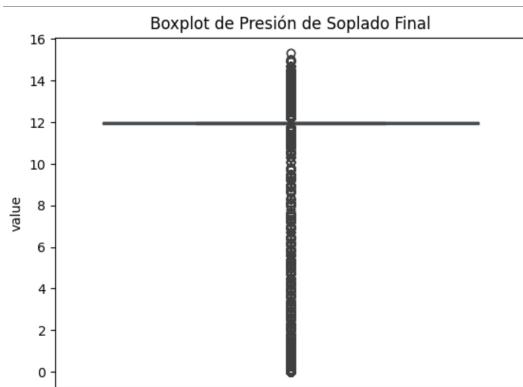


Fig. 6: Muestra la distribución asimétrica y la presencia de valores altos poco frecuentes.

Presión de la cámara de compensación:

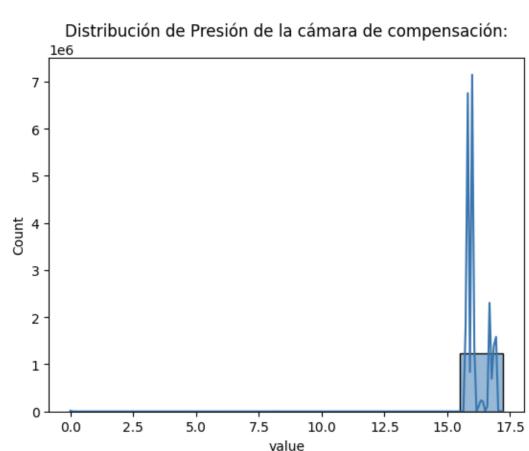


Fig. 7: Muestra la distribución asimétrica y la presencia de valores altos poco frecuentes.

- Se observa una presión estable en torno a 15-17 bar, pero con una gran cantidad de valores atípicos por debajo del rango intercuartil

- Esto indica que hay momentos en los que la presión cae abruptamente, lo que puede estar generando problemas en la estabilidad del proceso.

Botellas rechazadas:

Se aplicaron dos métodos para detectar valores atípicos:

- **Z-score (>3):** Se identificaron 1,148 valores atípicos.
- **IQR (rango intercuartílico):** Se detectaron 2,487 valores fuera del rango esperado.
- **Boxplot:** Confirmó la presencia de múltiples valores atípicos en la parte superior de la distribución.

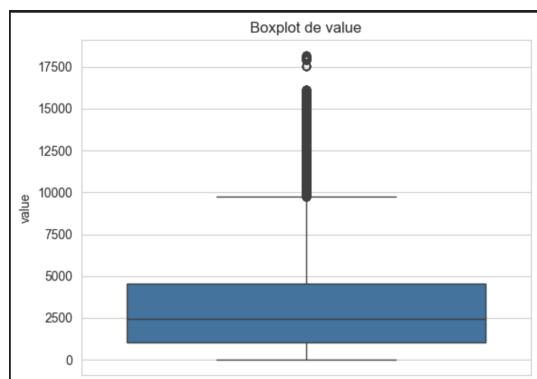


Fig. 8: Permite visualizar de manera clara la presencia de valores atípicos y su magnitud.

Análisis de series de tiempo

Layer 1:

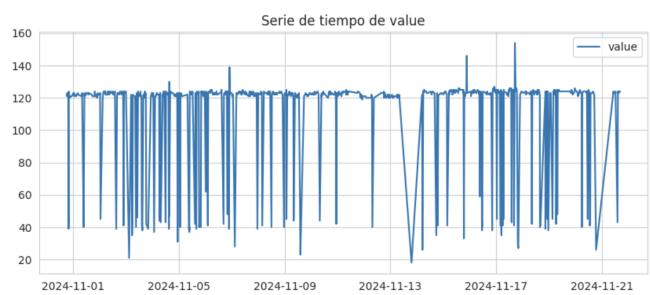


Fig. 9: Serie de tiempo Temperatura Layer 1.

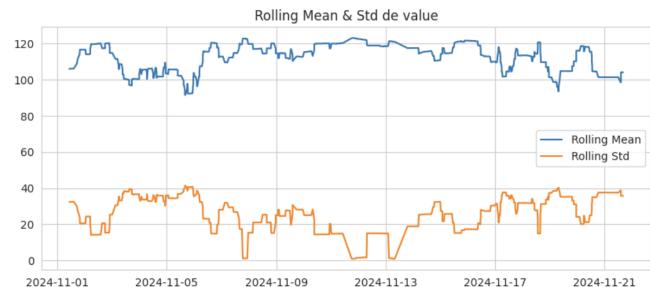


Fig. 10: Comportamiento de Media móvil y desviación estándar.



- Se observan fluctuaciones térmicas constantes.
- Hay descensos bruscos en la temperatura en varios períodos.
- En momentos específicos, la temperatura se estabiliza en su rango óptimo, pero con variaciones.

Impacto en la Producción: Las fluctuaciones térmicas podrían correlacionarse con un aumento en los rechazos. El sistema de control térmico de la preforma no es estable, lo que afecta la calidad del inflado de la botella.

Layer 3:

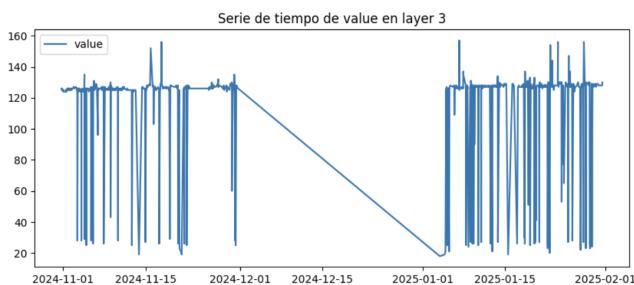


Fig. 11: Serie de tiempo Temperatura Layer 3.

Layer 5:

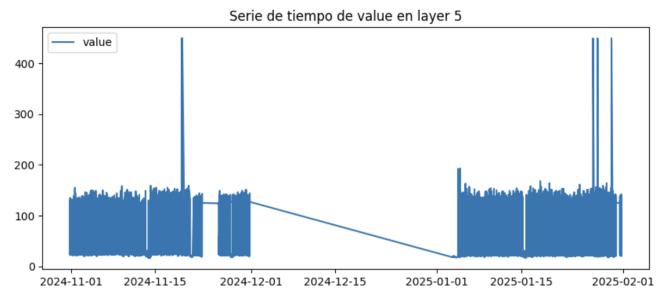


Fig. 13: Serie de tiempo Temperatura Layer 5.

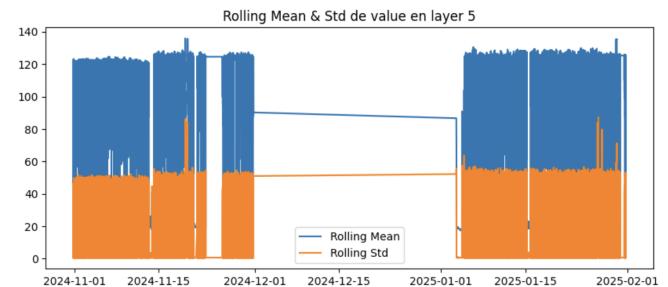


Fig. 14: Comportamiento de Media móvil y desviación estándar.

- El patrón es similar al de Layer 3, pero con una variabilidad menos extrema.
- Se observan picos muy elevados ($>400^{\circ}\text{C}$) que podrían estar relacionados con sobrecalentamiento de la sopladora
- El Z-score confirma múltiples valores atípicos, sobre todo en los picos altos
- La media móvil (Rolling Mean) se mantiene relativamente estable, pero se observan caídas abruptas en momentos específicos.
- La desviación estándar (Rolling Std) es alta en varios períodos, indicando que hay momentos de fluctuación térmica.

Fig. 12: Comportamiento de Media móvil y desviación estándar.

- Se observa un patrón constante en la temperatura, con períodos de estabilidad seguidos de caídas abruptas y picos inusuales.
- Hay descensos bruscos en la temperatura en varios períodos.
- Existe un período donde la temperatura cae repentinamente a valores muy bajos (cerca de 0), lo cual podría indicar fallas en la sopladora o mediciones erróneas.
- El Z-score confirma múltiples valores atípicos, sobre todo en los picos altos.
- La media móvil (Rolling Mean) se mantiene relativamente estable, pero se observan caídas abruptas en momentos específicos..
- La desviación estándar (Rolling Std) es alta en varios períodos, indicando que hay momentos de fluctuación térmica.

Layer 7:

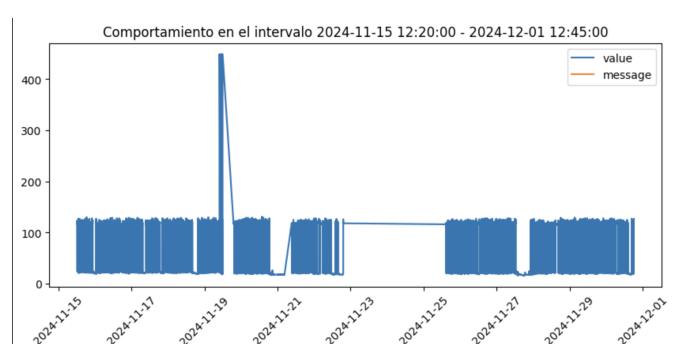


Fig. 15: Serie de tiempo de Temperatura Layer 7.

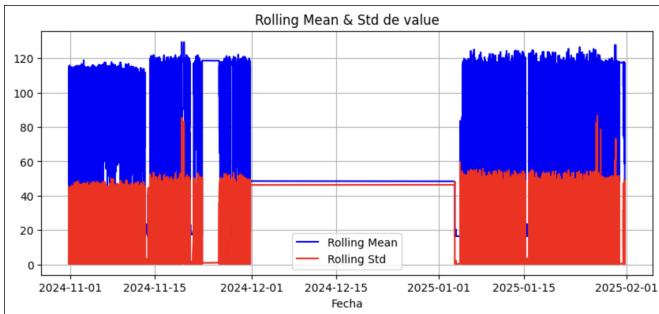


Fig. 16: Comportamiento de Media móvil y desviación estándar.

- Se observa un pico pronunciado el 19 de Noviembre que supera los 400°C, lo que representa un valor muy por encima de los niveles regulares de operación.
- Antes del pico, los valores parecen oscilar entre 100°C y 150°C, lo que podría representar el rango operativo normal de la máquina.
- Del 23-25 de Noviembre la temperatura permanece constante y baja, lo que sugiere que la máquina pudo haber estado apagada o en modo inactivo.

Layer 9:

Se detecta una caída abrupta en diciembre.

Después de la caída, la temperatura cambia drásticamente, lo que podría afectar la producción.

Impacto en la Producción Si la caída de temperatura coincide con aumentos en el rechazo, la hipótesis se fortalece.

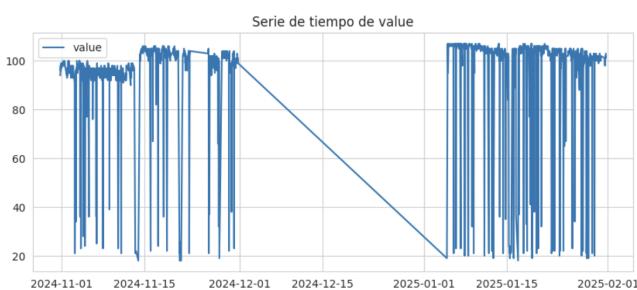


Fig. 17: Serie de Tiempo de Temperatura Layer 9.

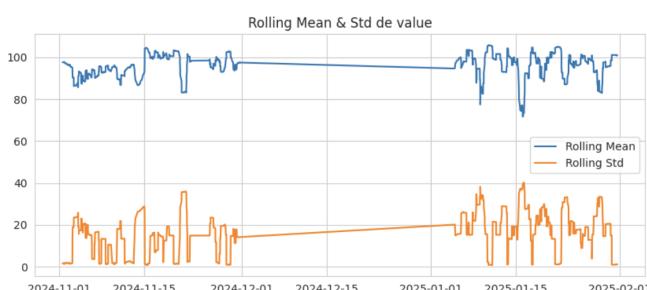


Fig. 18: Comportamiento de Media móvil y desviación estándar.

Presión en la cámara de compensado:

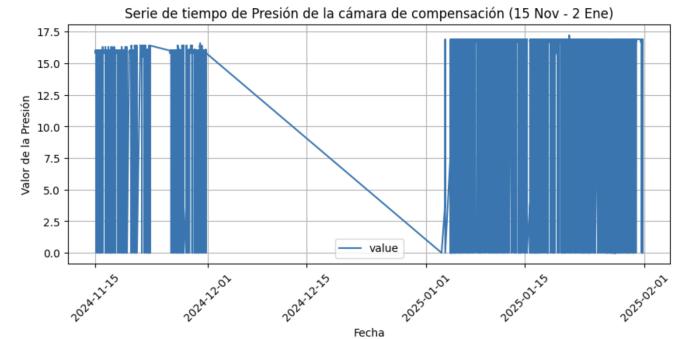


Fig. 19: Muestra la distribución asimétrica y la presencia de valores altos poco frecuentes.

- Antes del periodo vacacional, la presión se mantiene relativamente estable, con algunas fluctuaciones.
- Durante el periodo vacacional, hay una caída drástica y luego una recuperación con variabilidad más pronunciada

Presión final de soplado:

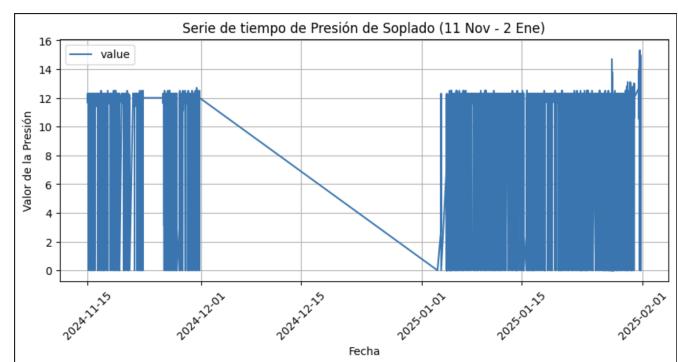


Fig. 20: Muestra la distribución asimétrica y la presencia de valores altos poco frecuentes.

- La tendencia es similar a la de la cámara de compensación: estabilidad antes del periodo vacacional, caída significativa durante la pausa y luego una recuperación con mayor variabilidad.

- Las fluctuaciones después del reinicio podrían estar afectando la calidad de las botellas y la tasa de rechazo.



Botellas rechazadas:

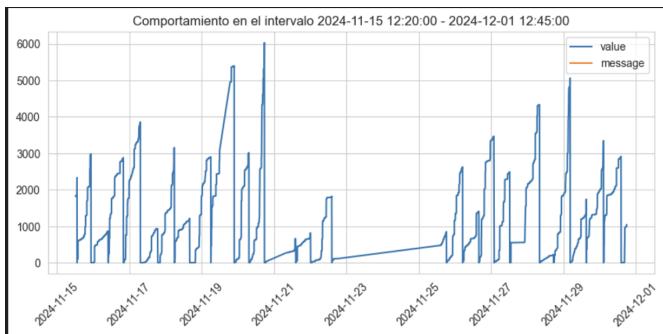


Fig. 21: Evidencia la variabilidad en los datos de Bottles Rejected y los periodos en los que la máquina pudo haber estado inactiva.

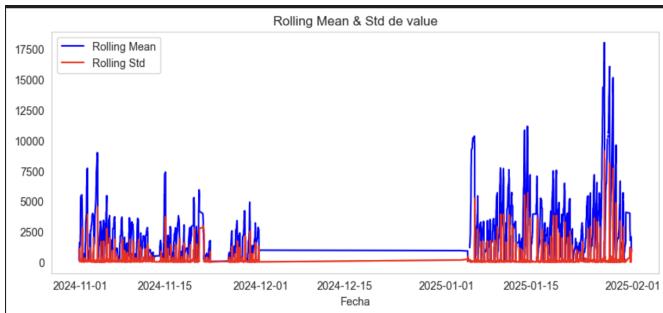


Fig. 22: Destaca cómo cambia la tendencia y variabilidad a lo largo del tiempo.

- Se detectaron períodos donde los valores son constantes, lo que sugiere que la máquina pudo haber estado apagada o en modo inactivo.
- Se observaron fluctuaciones altas en la variabilidad, con períodos de estabilidad intercalados con picos abruptos.
- Se aplicó un análisis de media móvil y desviación móvil para visualizar mejor la variabilidad en el tiempo.

Análisis de detección de anomalías

Layer 1:

- **Isolation Forest:** 38 valores anómalos detectados.
- **DBSCAN:** Clústeres distintos, lo que indica cambios constantes en la temperatura de la preforma.
- **LOF:** 34 valores anómalos, coincidiendo con los métodos anteriores.

Las anomalías térmicas detectadas podrían coincidir con picos en el rechazo de botellas.

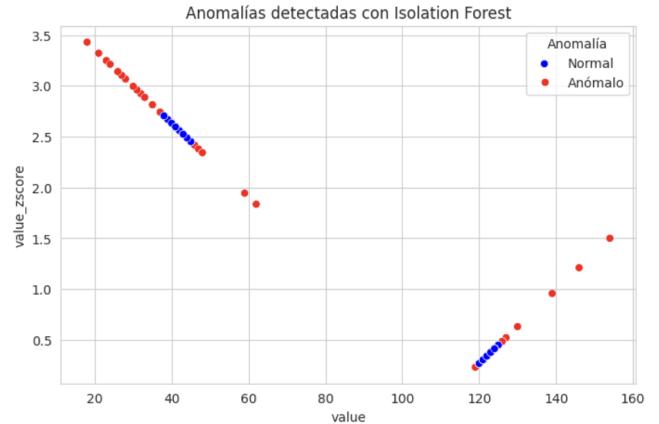


Fig. 23: Descripción de la imagen.

Layer 3

Resultados con Isolation Forest:

- 858 valores normales (azul) y 44 valores anómalos (rojo).
- Las anomalías detectadas se concentran en los valores extremos de temperatura:
 - Bajas temperaturas ($<50^{\circ}\text{C}$).
 - Altas temperaturas ($>140^{\circ}\text{C}$).
- Los valores bajos pueden estar relacionados con paros inesperados o lecturas defectuosas.
- Los valores altos pueden indicar sobrecalentamiento de la máquina.

Resultados con DBSCAN:

- DBSCAN detectó 59 anomalías principales (-1), lo que confirma la presencia de grupos de valores atípicos.
- Se observa una dispersión de puntos atípicos en diferentes clústeres, lo que sugiere que hay más de un tipo de anomalía en la distribución de temperatura.

Resultado con Local Outlier Factor (LOF):

- LOF identificó 67 anomalías, lo que refuerza la detección previa de valores extremos.

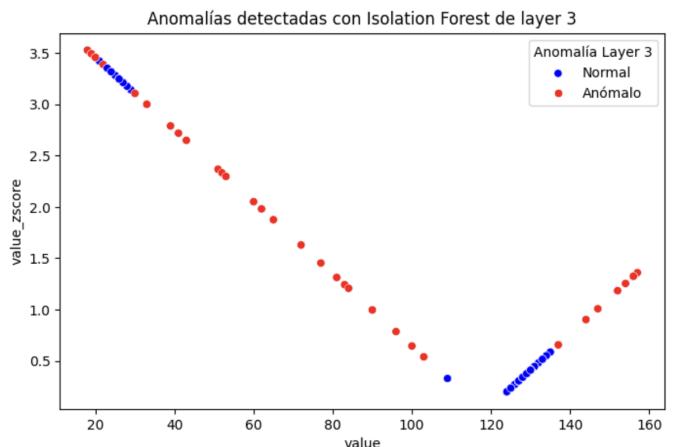


Fig. 24: Descripción de la imagen.

1. Layer 5

Resultados con Isolation Forest:

- 47,599 valores normales, 2,401 valores anómalos.
- Patrón de anomalías similar a Layer 3, con valores anómalos en temperaturas bajas ($<50^{\circ}\text{C}$) y altas ($>140^{\circ}\text{C}$).
- Se confirma que los valores extremos de temperatura afectan ambas capas.
- La mayor cantidad de anomalías en Layer 5 puede significar que este sensor es más sensible a las variaciones térmicas.

Resultados con MiniBatch K-Means (Clusters):

- Se identificaron 5 grupos de temperatura, con el grupo 4 como el más grande (16,933 registros).
- Los valores fuera de los clusters principales podrían indicar diferentes estados de la máquina (por ejemplo, en funcionamiento normal, en enfriamiento, en sobrecarga, etc.).

Resultados con Local Outlier Factor (LOF):

- Se detectaron 49 anomalías, confirmando que hay valores fuera de los patrones esperados.

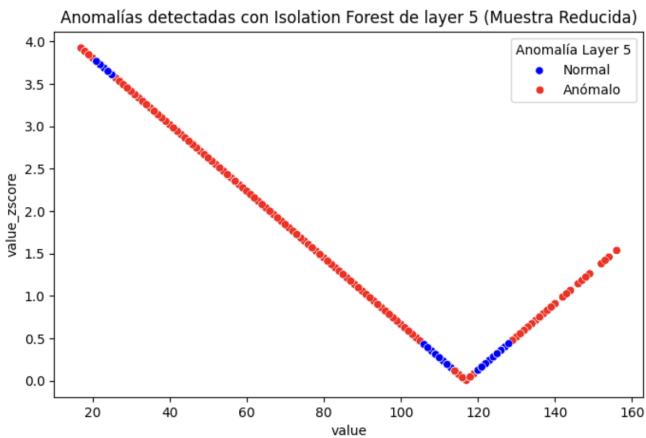


Fig. 25: Descripción de la imagen.

Layer 7

Se identificaron valores atípicos en la temperatura de la preforma utilizando el método del rango intercuartílico (IQR).

Hay dos períodos principales con altas concentraciones de anomalías:

- 1 de noviembre de 2024: Variaciones térmicas entre 25°C y 43°C .
- 31 de enero de 2025: Valores extremadamente bajos, llegando hasta 18°C .

Layer 9

- **Isolation Forest:** 812 valores anómalos detectados.
- **DBSCAN:** Se detectaron múltiples clústeres.

- **LOF:** 816 valores anómalos.

La alta cantidad de anomalías térmicas podría estar afectando la calidad de la preforma y la tasa de rechazo.

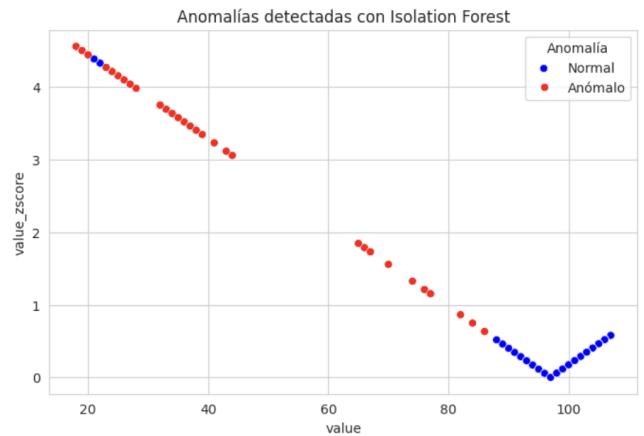


Fig. 26: Descripción de la imagen.

Bottles Rejected

- Valores superiores a 9399.6: Se identificaron 2812 registros que superan este umbral, lo que indica una cantidad de botellas rechazadas inusualmente alta.
- 19 de noviembre de 2024: Este fue el día con la mayor cantidad de botellas rechazadas, coincidiendo con un pico extremo de temperatura que superó los 400°C .

Impacto en la Producción Estos eventos representan una pérdida significativa de material y recursos.

VI. HIPÓTESIS DEL ANÁLISIS EXPLORATORIO DE DATOS

Hipótesis 1: "El aumento en la temperatura de la preforma está asociado con un mayor número de botellas rechazadas."

VARIABLES CLAVE:

- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_1
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_3
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_5
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_7
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_9
- CONTIFORM_MMA.CONTIFORM_MMA1.WS_Tot_Rej.0

A partir de la combinación de las series de tiempo y box-plots, encontramos hallazgos clave que ayudan a evaluar esta hipótesis. En este caso, buscamos una relación en el comportamiento de las variables de temperatura con respecto a la



cantidad total de botellas rechazadas en el proceso. Después de representar la cantidad total de datos en una misma dimensión para asegurar un análisis preciso, se graficaron series de tiempo que representarán precisamente el total de rechazos con la temperatura de la preforma, para ver cómo es que estas variables se relacionan, se muestra a continuación.

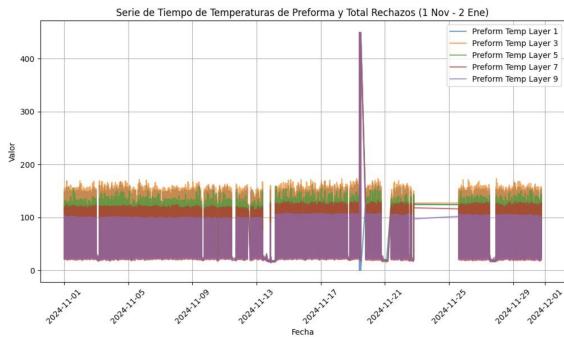


Fig. 27: Comportamiento de Temperaturas de Preforma a lo largo del tiempo

De manera visual percibimos especialmente dos aspectos: Primero, hay un pico muy grande alrededor del día 19 de noviembre de 2024, un pico que representa una elevación en el nivel de la temperatura muy por encima del promedio usual, y es interesante porque precisamente ese pico se empata con el día en que más botellas rechazadas se registraron.

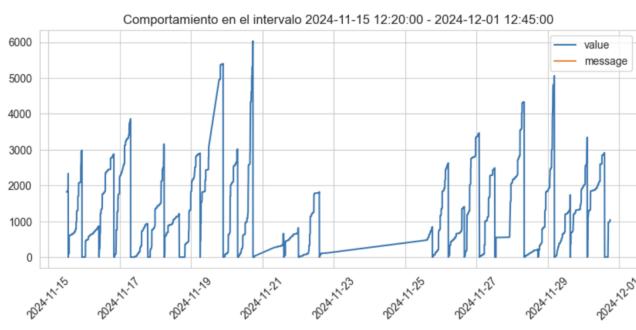


Fig. 28: Serie de tiempo de Botellas rechazadas en periodo de fallo de máquina

Por otro lado, podemos ver que hay un espacio vacío entre el día 23 de noviembre y 25 de noviembre, un espacio que definitivamente representa una falla después de lo acontecido, un espacio en el que los datos fueron constantes, donde probablemente se apagó la máquina, confirmando así una evidente falla en la máquina, representando una pérdida en muchos otros aspectos. A su vez, para confirmar lo anterior, realizamos boxplots que comparan la distribución de la temperatura en diferentes capas de la preforma entre lotes con alto y bajo rechazo.

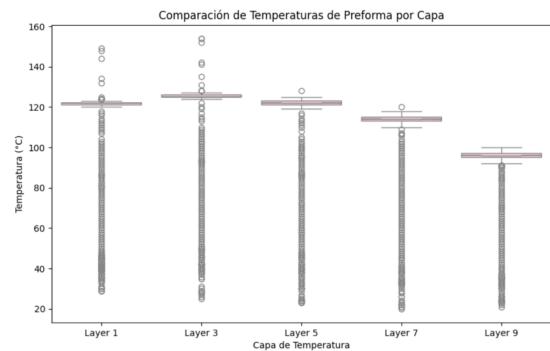


Fig. 29: Boxplots para Temperaturas de Preforma

En los boxplots de las diferentes capas de temperatura, se observa que los lotes con alto rechazo tienen mayor dispersión térmica, en todas las capas de la preforma, los lotes con alto rechazo presentan más valores extremos (outliers) hacia temperaturas bajas y altas. Los outliers a temperaturas extremadamente bajas ($20-50^{\circ}\text{C}$) aparecen en los lotes con alto rechazo, lo que sugiere que la falta de calor en ciertos momentos podría estar generando defectos en la preforma y, en consecuencia, más botellas rechazadas. Los boxplots muestran que la variabilidad térmica es mayor en los lotes con alto rechazo, lo que sugiere que las fluctuaciones térmicas podrían estar contribuyendo a los defectos en la producción.

Hipótesis 2: "El desgaste de la sopladora se acelera cuando las temperaturas de la preforma están fuera de cierto rango óptimo."

Variables clave:

- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_1
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_3
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_5
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_7
- CONTIFORM_MMA_CONTIFORM_MMA1_PreformTemperatureLayer_9
- CONTIFORM_MMA.CONTIFORM_MMA1.WS_Tot_Rej.0

Estadísticas descriptivas

- Mínimos: Algunas capas tienen valores de 0°C , 16°C o 18°C , lo cual no es normal para una sopladora.
- Máximos: En todas las capas, hay temperaturas de hasta 449°C , lo que podría estar indicando sobrecalentamientos.
- Los valores medios están entre 70°C y 103°C , lo que indica que hay variabilidad en la temperatura dependiendo de la capa.

- Layer 9 tiene la menor temperatura promedio (70°C), mientras que Layer 3 tiene la más alta (103°C).

Si la diferencia entre capas es muy grande, el proceso de soplado podría ser inconsistente y afectar la vida útil de la máquina.

Boxplots de Temperaturas

Se utilizaron diagramas de caja y bigotes (boxplots) para visualizar la dispersión de las temperaturas y la presencia de outliers. Se observaron pocos outliers porque se alinearon los timestamps de las 5 capas, lo que puede haber suavizado algunas diferencias extremas en la visualización.

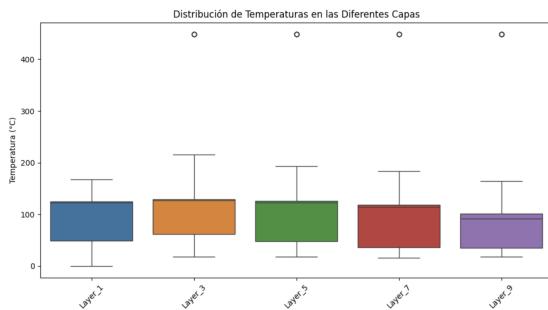


Fig. 30: Boxplots para Temperaturas de Preforma

Resultados

Los datos sugieren que las temperaturas de la proforma no son estables y tienen valores extremos en ciertos momentos, lo cual podría estar causando un mayor desgaste en la sopladora; El siguiente paso para validar que las temperaturas fuera de rango afectan el desgaste de la sopladora sería cruzar estos datos con registros de mantenimiento, analizar la variabilidad termina en el tiempo y comparar estos datos con la cantidad de botellas rechazadas.

Hipótesis 3: "Variaciones abruptas en la presión de la cámara de compensación aceleran el desgaste de la sopladora y el proceso de rechazo de botellas."

Variables clave:

- CONTIFORM_MMA_CONTIFORM_MMA1_Pressure CompensationChamberPressureActualValue_0
- CONTIFORM_MMA.CONTIFORM_MMA1.WS_Tot_Rej.0

Hipótesis 3: Si la media y la desviación estándar de la presión en la cámara de compensación aumentan, entonces el número de botellas rechazadas también se incrementará.

Esta hipótesis surge de la idea de que una mayor variabilidad en la presión podría generar inestabilidad en el proceso de soplado, afectando la calidad de las botellas. La presión juega un papel crucial en la formación de la botella, y fluctuaciones significativas pueden derivar en deformaciones, variaciones en el grosor del material o defectos estructurales.

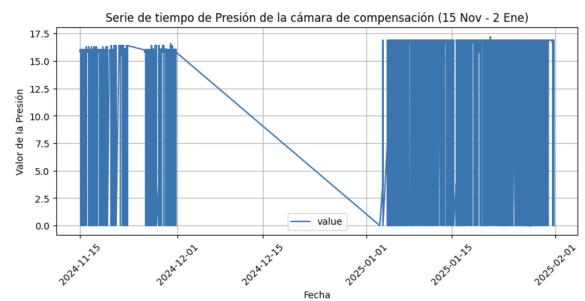


Fig. 31: Serie temporal de la presión en la cámara de compensación. Se observa un cambio en la tendencia antes y después del periodo vacacional.

Conclusión

Los resultados obtenidos en el análisis respaldan la hipótesis planteada.

Al comparar los datos antes y después del periodo vacacional, observamos que:

- La **media de la presión** aumentó tras el periodo vacacional, indicando que el sistema operaba a un nivel de presión más alto en comparación con el periodo previo.
- La **desviación estándar de la presión** también se incrementó, lo que sugiere que hubo más fluctuaciones en la presión a lo largo del tiempo.
- Simultáneamente, se observó un **aumento en la cantidad de botellas rechazadas**, lo que sugiere que esta variabilidad en la presión podría estar afectando la estabilidad del proceso y, en consecuencia, la calidad de las botellas.



Fig. 32: Evolución del número de botellas rechazadas antes y después del periodo vacacional. Se observa un incremento en los rechazos cuando la variabilidad de la presión es mayor.

Para visualizar mejor este comportamiento, también analizamos la media móvil y la desviación estándar de la presión:

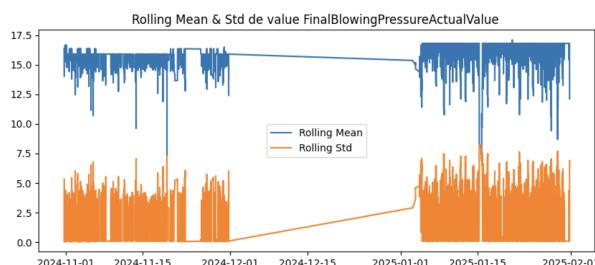


Fig. 33: Media móvil y desviación estándar de la presión en la cámara de compensación. Se observa un aumento en la variabilidad después del periodo vacacional.

Estos resultados refuerzan la importancia de mantener condiciones de presión estables para minimizar defectos en la producción. Se recomienda profundizar en este análisis evaluando si existen factores adicionales que puedan estar contribuyendo a esta variabilidad y explorar estrategias para regular la presión dentro de parámetros óptimos.

VII. CONCLUSIONES

Finalmente, después de haber realizado el análisis total, tomando en cuenta cada hipótesis, podemos concluir que diversas revelaciones por parte del presente análisis. Primeramente, mencionar que el análisis revela que la variabilidad térmica en la preforma está relacionada con un mayor número de botellas rechazadas, ya que las fluctuaciones extremas y los valores fuera del rango óptimo afectan la calidad del inflado. Además, las temperaturas inestables y los valores extremos podrían estar acelerando el desgaste de la sopladora, lo que hace necesario comparar estos datos con registros de mantenimiento. Asimismo, las variaciones abruptas en la presión de la cámara de compensación podrían estar contribuyendo a fallos operativos, como se observó en la interrupción de datos del 23 de noviembre y la desconexión en diciembre. Finalmente, tomando en cuenta los hallazgos, podemos concluir que para mitigar estos efectos, es clave determinar el rango óptimo de temperatura y presión, validar su impacto en la vida útil de la máquina y mejorar la regulación térmica y la estabilidad del sistema, optimizando así la eficiencia del proceso y reduciendo pérdidas.[3]

A. ANEXOS

El código fuente y los datos utilizados en este trabajo están disponibles en el siguiente repositorio:

Repositorio en GitHub

[4] [5] [6] [7] [8]

REFERENCIAS

- [1] DataScientest, “Machine learning & clustering: el algoritmo dbscan,” recuperado de. [Online]. Available: <https://datascientest.com/es/machine-learning-clustering-dbscan>
- [2] —, “Isolation forest: ¿cómo detectar anomalías en un conjunto de datos?” DataScientest, enero 2024. [Online]. Available: <https://datascientest.com/es/isolation-forest>
- [3] OpenAI, “Chatgpt (versión gpt-4) [modelo de lenguaje de inteligencia artificial],” 2025. [Online]. Available: <https://chat.openai.com/>

- [4] S.-L. Developers, “Dbscan — scikit-learn 1.3.0 documentation,” recuperado de. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>
- [5] —, “Isolationforest — scikit-learn 1.3.0 documentation,” recuperado de. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>
- [6] —, “Lof outlier detection example,” recuperado de. [Online]. Available: https://scikit-learn.org/stable/auto_examples/neighbors/plot_lof_outlier_detection.html
- [7] —, “Minibatchkmeans — scikit-learn 1.3.0 documentation,” recuperado de. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MiniBatchKMeans.html>
- [8] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, and S. Contributors, *SciPy v1.10.0 Reference Guide – Statistics module*, 2025. [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/stats.html>