

Análisis de la Supervisión Industrial mediante Algoritmos de Detección de Anomalías con Técnicas de Machine Learning

Enrique Campos Rodríguez¹, Reyna Esmeralda García Guaderrama²

¹Tecnológico Nacional de México, Instituto Tecnológico de Cd. Jiménez. Departamento de Sistemas y Computación.

²Tecnológico Nacional de México, Instituto Tecnológico de Cd. Cuauhtémoc. Departamento de Sistemas y Computación.

Resumen

Los resultados muestran que determinados periodos temporales y líneas de producción concentran entre un 18 % y 25 % de eventos clasificados como anomalías colectivas, permitiendo focalizar acciones correctivas específicas. Aunque el enfoque es exploratorio y no supervisado, los hallazgos evidencian la utilidad del enfoque para apoyar la supervisión industrial basada en datos.

Cabe destacar que los porcentajes reportados corresponden a la concentración y distribución de eventos clasificados como anomalías colectivas, y no a una evaluación cuantitativa del desempeño del algoritmo. Al tratarse de un enfoque exploratorio y no supervisado, y ante la ausencia de datos etiquetados en el entorno industrial analizado, no se calcularon métricas tradicionales de validación como precisión, recall o exactitud. El objetivo del modelo fue identificar patrones atípicos recurrentes que apoyaran el análisis operativo, más que medir su capacidad predictiva formal.

Palabras clave: Machine Learning; tiempos muertos; Isolation Forest; supervisión industrial; anomalías colectivas.

Abstract

The results show that specific temporal periods and production lines concentrate between 18% and 25% of events classified as collective anomalies, enabling the targeting of specific corrective actions. Although the

approach is exploratory and unsupervised, the findings highlight the usefulness of this methodology in supporting data-driven industrial supervision.

It should be noted that the reported percentages correspond to the concentration and distribution of events classified as collective anomalies, and not to a quantitative evaluation of the algorithm's performance. Given the exploratory and unsupervised nature of the approach, and the absence of labeled data in the analyzed industrial environment, traditional validation metrics such as precision, recall, or accuracy were not computed. The objective of the model was to identify recurrent atypical patterns that support operational analysis, rather than to formally measure predictive performance.

Keywords: Machine Learning; downtime; Isolation Forest; industrial supervision; collective anomalies.

Introducción

En el entorno industrial actual, caracterizado por la alta competitividad y la necesidad de eficiencia operativa, los tiempos muertos representan uno de los principales desafíos para la productividad. Estos periodos de inactividad no planificada afectan directamente la capacidad de respuesta, los costos operativos y la rentabilidad de las empresas manufactureras (Espacio ERP, s.f.).

Diversas causas como fallas técnicas, falta de materiales, problemas de calidad, errores humanos y deficiencias organizativas contribuyen a la aparición

de estos tiempos improproductivos (Espacio ERP, s.f.). Identificar y reducir estos lapsos es esencial para mejorar el rendimiento global de los procesos productivos.

En este contexto, el uso de técnicas de Machine Learning ha emergido como una solución innovadora para el análisis de datos operativos y la detección de patrones que permiten anticipar y mitigar los tiempos muertos. En particular, el algoritmo Isolation Forest, diseñado para la detección de anomalías en conjuntos de datos no etiquetados, ha demostrado ser eficaz en la identificación de eventos atípicos que preceden a interrupciones en la producción (Amat Rodrigo, 2020). La presente investigación fue desarrollada en colaboración con Leviton de México, sucursal Jiménez, dedicada a la manufactura de componentes eléctricos. Esta organización proporcionó el entorno industrial real y los datos operativos necesarios para el análisis, permitiendo aplicar técnicas de Machine Learning en la detección de anomalías colectivas relacionadas con tiempos muertos. La participación de Leviton fue fundamental para validar la aplicabilidad del modelo en condiciones productivas reales, fortaleciendo el vínculo entre la investigación académica y la mejora continua en la industria.

Pregunta de investigación:

¿Puede el algoritmo Isolation Forest identificar de forma efectiva patrones de anomalías colectivas asociados a tiempos muertos en un entorno industrial real sin datos etiquetados?

Objetivo del estudio:

Analizar la aplicación del algoritmo Isolation Forest en la detección de anomalías colectivas relacionadas con tiempos muertos en datos históricos de producción industrial.

Hipótesis exploratoria:

El uso de Isolation Forest permite identificar patrones anómalos colectivos que no son evidentes mediante análisis descriptivo tradicional, proporcionando información útil para la supervisión operativa.

Fundamentos teóricos

1. Tiempos Muertos en la Producción

Los tiempos muertos se definen como los periodos en los que los recursos productivos (máquinas, personal, materiales) están disponibles, pero no se utilizan de manera efectiva (Espacio ERP, s.f.). Estos pueden clasificarse en:

- Fallas técnicas (mecánicas, eléctricas, software)
- Falta de materiales o insumos
- Cambios de línea o ajustes de maquinaria
- Problemas de calidad
- Desorganización operativa
- Falta de personal o capacitación

Estudios como el de Hernández et al. (2024) en la industria metal-mecánica demuestran que la aplicación de herramientas como el diagrama de Pareto y metodologías Lean Manufacturing permite identificar y reducir significativamente estos tiempos (Hernández Ledezma et al., 2024).

2. Machine Learning en la Optimización de Procesos

El Machine Learning permite analizar grandes volúmenes de datos operativos para identificar patrones ocultos, predecir eventos y tomar decisiones automatizadas (Improvitz, 2024). En manufactura, se aplica en:

- **Mantenimiento predictivo**
- **Control de calidad**
- **Optimización de la cadena de suministro**
- **Detección de anomalías**

La combinación de técnicas supervisadas y no supervisadas permite abordar problemas complejos como los tiempos muertos desde múltiples perspectivas.

3. Isolation Forest para Detección de Anomalías

Isolation Forest es un algoritmo no supervisado que detecta anomalías mediante el aislamiento de observaciones en árboles de decisión aleatorios (Amat Rodrigo, 2020). Las observaciones que se aíslan rápidamente (con pocas divisiones) son consideradas anomalías.

Este método es especialmente útil en entornos industriales donde los datos no están etiquetados y se requiere identificar eventos atípicos que podrían estar relacionados con interrupciones en la producción.

Garcés y Castrillón (2017), aplicaron técnicas inteligentes como árboles de decisión (J48 en WEKA) para identificar causas de tiempos muertos, logrando mejoras significativas en la eficiencia global.

Materiales y métodos

Software y herramientas:

Para el desarrollo de esta investigación se utilizó el lenguaje de programación Python, junto con las librerías especializadas en detección de anomalías.

Datos utilizados:

Se analizaron datos históricos de producción correspondientes a un periodo de aproximadamente seis meses, con el objetivo de detectar anomalías colectivas. Las anomalías colectivas involucran un grupo de puntos de datos que muestran comportamiento anómalo a pesar de que los puntos de datos individuales pueden verse normales (Elastic, s.f.).

Métodos de detección:

Se consideraron tres algoritmos populares para la detección de anomalías:

- **Isolation Forest**
- **One-Class SVM**
- **DBSCAN**

En comparativa de los tres métodos, Isolation Forest mostró un desempeño favorable (Lu, 2025), en la identificación de anomalías colectivas, por lo que se seleccionó como método principal para el análisis.

Variables del conjunto de datos:

El conjunto de datos incluye las siguientes variables:

- **ID:** Identificador único del evento
- **HoraInicio / HoraFin:** Tiempos de inicio y fin del evento, agrupados por bloques de una hora
- **Job:** Número de trabajo asociado
- **Línea:** Línea de producción en la que pasa el evento
- **Departamento:** Departamento al que pertenece la línea de producción
- **Razón:** Causa del tiempo muerto
- **Duración:** Tiempo total del evento
- **Turno:** Turno en el que pasa el evento
- **Semana:** Segmentación temporal del evento
- **Usuario / Operadores:** Información del personal involucrado

Preparación de los datos:

Para evitar errores en el procesamiento de los datos se hace normalización en los campos como Razón, Departamento y Turno para evitar duplicados por errores de escritura. Los valores nulos o vacíos en la columna “Razón” fueron reemplazados por el término “Desconocido”.

Configuración del modelo Isolation Forest

El algoritmo Isolation Forest fue implementado utilizando la librería *scikit-learn*. El modelo se configuró con 100 árboles ($n_estimators = 100$), una proporción de anomalías esperada del 0.1 ($contamination = 0.1$) y muestreo automático ($max_samples = auto$).

Dado que el conjunto de datos carece de etiquetado previo, el estudio adopta una perspectiva exploratoria, orientada a la detección de patrones anómalos colectivos. En consecuencia, no se aplicaron esquemas de validación supervisada ni métricas de desempeño estándar, y la evaluación del modelo se centró en la consistencia temporal, la coherencia de los patrones detectados y su interpretabilidad desde el punto de vista operativo.

Resultados y discusión

Los resultados presentados corresponden al análisis descriptivo de los patrones detectados por el modelo y no constituyen una evaluación de su desempeño predictivo.

El propósito de encontrar anomalías colectivas es para analizar la ocurrencia de tiempos muertos en varios aspectos:

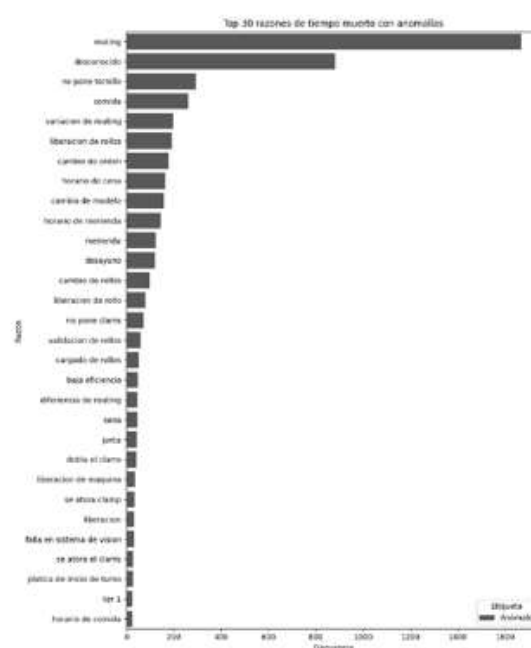
- Anomalías colectivas detectadas en total de los datos históricos
- Anomalías colectivas agrupados por semana
- Anomalías colectivas por día
- Duración de anomalías colectivas, tanto por día como por semana
- Ocurrencia de anomalías colectivas por línea de producción o por departamento

Tras la limpieza de datos, se identificaron las causas más frecuentes de tiempos muertos:

Tabla 1. Tiempos muertos agrupados por razón.

Elaboración propia

Razón	Frecuencia
Routing	1666
Desconocido	881
No pone tornillo	293
Comida	262
Variación de routing	197
Liberación de rollos	192
Cambio de orden	179
Horario de cena	165
Cambio de modelo	157
Horario de merienda	145

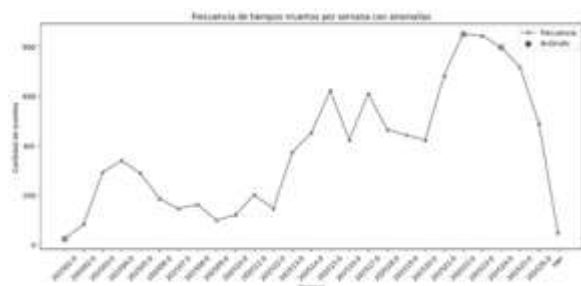


Gráfica 1. Top 30 razones de tiempo muerto con anomalía. Esta gráfica presenta las causas más frecuentes de tiempos muertos detectados como anomalías colectivas por el modelo Isolation Forest, permitiendo identificar focos de atención operativa.

Análisis semanal de anomalías colectivas

El modelo detectó anomalías colectivas en tres semanas específicas: la semana 1 (por baja frecuencia

de eventos), y las **semanas 22 y 24** (por alta frecuencia).

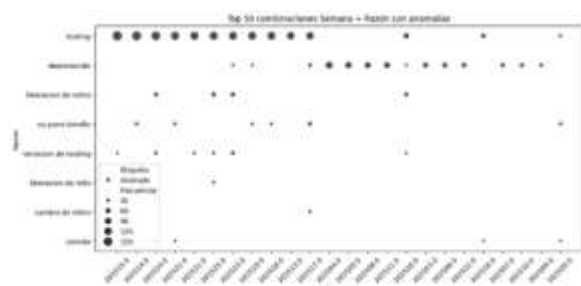


Gráfica 2. Anomalías colectivas por semana

En la gráfica 2 se observa la ocurrencia de anomalías colectivas a lo largo del periodo analizado.

Combinaciones frecuentes razón-semana:

Se identificaron las 50 combinaciones más frecuentes entre razón y semana, lo que permitió visualizar patrones recurrentes y posibles focos de atención operativa.

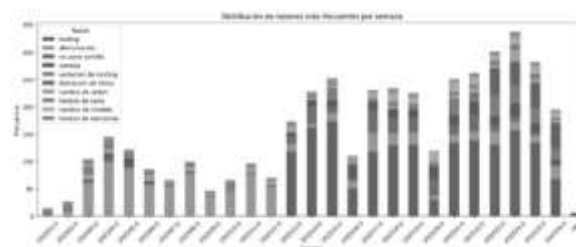


Gráfica 3. Top 50 anomalías colectivas por semana, así como su razón

Esta visualización permite analizar simultáneamente la causa y la semana de ocurrencia de las anomalías colectivas.

Visualización apilada por semana

Se utilizó una gráfica de barras apiladas para mostrar la frecuencia de anomalías por semana, facilitando la comparación entre causas.

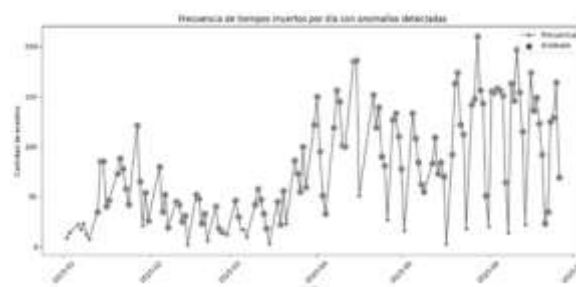


Gráfica 4. Frecuencia apilada de anomalías por semana

Esto permite observar la distribución de causas de anomalías colectivas en cada semana.

Frecuencia diaria con anomalías.

Se analizó la ocurrencia diaria de eventos, destacando aquellos días en los que al menos una anomalía colectiva fue detectada.



Gráfica 5. Frecuencia de tiempos muertos con existencia de anomalías colectivas

Esta gráfica indica la cantidad total de eventos diarios y si al menos uno de ellos fue clasificado como anomalía colectiva.

Resumen de anomalías frecuentes:

Tabla 2. Resumen de anomalías frecuentes.

Elaboración propia

Razón	Días detectados
Routing	60
Desconocido	55
No pone tornillo	28
Variación de routing	22
Comida	21
Liberación de rollos	19
Desayuno	9
Cambio de modelo	8
Cambio de orden	7

Cambio de rollos	6
------------------	---

Ahora, un dato relevante de las anomalías colectivas es también la duración de las mismas.

A continuación, la tabla muestra las anomalías colectivas por día con mayor duración:

Tabla 3. Duración de anomalías colectivas por día. Elaboración Propia.

Fecha	Razón	Frecuencia	Duración Total (minutos)
2025-03-20	desconocido	35	1280
2025-01-21	desconocido	32	1260
2025-01-30	desconocido	32	1170
2025-01-14	desconocido	34	1035
2025-04-23	desconocido	23	1015
2025-04-07	routing	52	965
2025-01-23	desconocido	28	920
2025-06-25	desconocido	20	917
2025-03-27	routing	74	856
2025-03-31	routing	61	831

Cabe mencionar que estos datos son anomalías colectivas diarias, pero no significa que hayan sucedido en una sola línea de producción.

La siguiente tabla da una muestra de anomalías colectivas diarias, y en qué línea de producción sucedieron:

Tabla 4. Desglose de anomalías colectivas diarias por línea de producción.

Elaboración propia

Fecha	Razón	Línea	Eventos	Duración Total
2025-01-13	irobot	Decora	10	250
2025-01-14	desconocido	BDN16	6	360
2025-01-14	desconocido	BDN06	5	140
2025-01-14	desconocido	MB252	5	140
2025-01-14	desconocido	BDN50 50	4	110
2025-01-14	desconocido	RTF01	3	100
2025-01-14	desconocido	TAM45	4	55
2025-01-14	desconocido	MB693	3	50
2025-01-14	desconocido	MB117 4	2	40
2025-01-14	desconocido	BDN04	1	20

Los resultados obtenidos son consistentes con estudios previos que reportan la utilidad del algoritmo Isolation Forest en contextos industriales con datos no etiquetados. Al igual que lo documentado por Lu (2025), el modelo permitió identificar anomalías colectivas asociadas a concentraciones temporales de eventos, lo cual resulta de valor para la supervisión operativa, aunque sin establecer comparaciones directas de desempeño con métodos supervisados.

Interpretación de resultados

El modelo de Isolation Forest demostró ser funcional como herramienta exploratoria para identificar patrones de anomalías colectivas en los tiempos muertos, con lo que se obtiene una visualización detallada por razón, semana, día y línea de producción.

Esta información puede ser utilizada para optimizar procesos operativos, reducir tiempos improductivos y priorizar acciones correctivas. Además, el modelo es ajustable a las necesidades específicas de la empresa, permitiendo “percibir” una cantidad diferente de anomalías y en conjunto con los criterios expertos de la empresa puedan mejorar la sensibilidad del análisis. Mejoras futuras:

Una de las recomendaciones sería la posibilidad de integración del algoritmo en sistemas de monitoreo continuo o en tiempo real, con el propósito de automatizar la detección temprana de anomalías colectivas y contribuir a la mejora continua de los procesos productivos.

Conclusiones

El presente estudio demuestra que el algoritmo Isolation Forest es una herramienta adecuada para la detección de anomalías colectivas en datos operativos de tiempos muertos dentro de procesos de producción en la industria manufacturera. A diferencia de otros métodos tradicionales, Isolation Forest permite identificar patrones anómalos que no son evidentes a nivel individual, pero que emergen al analizar grupos de eventos que comparten características temporales causales o estructurales.

El presente trabajo no tiene como finalidad evaluar comparativamente el rendimiento del algoritmo Isolation Forest, sino demostrar su utilidad como herramienta exploratoria para el análisis de tiempos muertos en contextos industriales reales sin datos etiquetados. Los resultados obtenidos sientan las bases para investigaciones futuras que incorporen validación cuantitativa, datos etiquetados o enfoques supervisados.

Referencias bibliográficas

- Espacio ERP. (s.f.). Clasificación de causas de tiempos muertos en producción. <https://espacioerp.com/tiempos-muertos-produccion/>
- Hernández Ledezma, I. L., Castillo Treviño, Z., Zapata Reboloso, A., & Alcalá Salinas, C. A. (2024). Identificación y reducción de tiempos muertos como estrategia de productividad en la industria metalmeccánica. *Ciencia Latina Revista Científica Multidisciplinaria*, 8(5), 14726. https://doi.org/10.37811/cl_rcm.v8i5.14726
- Garcés, D. A., & Castrillón, O. D. (2017). Diseño de una técnica inteligente para identificar y reducir los tiempos muertos en un sistema de producción. *Información Tecnológica*, 28(3), 17–28. <https://doi.org/10.4067/S071807642017000300017>
- Amat Rodrigo, J. (2020). Detección de anomalías con Isolation Forest. *Ciencia de Datos*. https://cienciadedatos.net/documentos/66_deteccion_anomalias_isolationforest
- Improvitz. (2024). Machine learning y la optimización de procesos: una guía completa para mejorar la eficiencia operativa. <https://improvitz.com/machine-learning-y-la-optimizacion-de-procesos-una-guia-completa-para-mejorar-la-eficiencia-operativa/>
- Elastic. (s.f.). ¿Qué es la detección de anomalías? <https://www.elastic.co/es/what-is/anomaly-detection>
- Lu, H. (2025). Evaluating the performance of SVM, Isolation Forest, and DBSCAN for anomaly detection. *Semantic Scholar*. <https://pdfs.semanticscholar.org/a290/623a4584cb8cbd081dc4b592ae0b8ca79612.pdf>