

**Universidad Tecnológica de La Habana**

**“José Antonio Echeverría”**

**CUJAE**

**Facultad de Ingeniería Informática**

**Módulo de detección de anomalías en un sistema de control industrial**

**Informe de las prácticas profesionales II**

**Autor: Fabio Ford Campbell (fford@ceis.cuaje.edu.cu)**

**Tutor: Dra.C Nayma Cepero Pérez (no me sé el correo)**

**La Habana,**

**enero de 2025**

**Resumen**

En el contexto de la Industria 4.0, la detección temprana de anomalías en sistemas industriales es crucial para garantizar la continuidad y eficiencia operativa. Este estudio se enfoca en el desarrollo de un módulo de detección de anomalías que supervisa en tiempo real los datos operativos y de red de un sistema industrial experimental. Utilizando algoritmos de minería de datos y aprendizaje automático, el sistema es capaz de identificar desviaciones significativas que podrían indicar fallos operacionales. El objetivo principal es desarrollar un módulo capaz de detectar y alertar sobre anomalías en tiempo real, mejorando la resiliencia y eficiencia de los procesos industriales.

**Palabras Claves:** detección de anomalías, sistemas industriales, minería de datos, aprendizaje automático, Industria 4.0

**Abstract**

In the context of Industry 4.0, early detection of anomalies in industrial systems is crucial to ensuring operational continuity and efficiency. This study focuses on developing an anomaly detection module that monitors operational and network data from an experimental industrial system in real-time. Using data mining and machine learning algorithms, the system can identify significant deviations that may indicate operational failures. The main goal is to develop a module capable of detecting and alerting anomalies in real-time, thereby improving the resilience and efficiency of industrial processes.

**Keywords:** anomaly detection, industrial systems, data mining, machine learning, Industry 4.0

# 

**Índice de contenido**

[Introducción 1](#_Toc15025)

[Capítulo 1: Fundamentos Teóricos 4](#_Toc23793)

[1.1 Detección de Anomalías en Sistemas Industriales 4](#_Toc4766)

[1.1.1 Tipos de Anomalías en Sistemas Industriales 4](#_Toc186)

[1.1.2 Principios de la Detección de Anomalías 5](#_Toc7081)

[1.1.3 Algoritmos para la Detección de Anomalías 5](#_Toc8969)

[1.1.4 Sistemas Industriales y su Monitoreo 6](#_Toc11411)

[1.1.5 Minería de Datos y Aprendizaje Automático en la Detección de Anomalías 7](#_Toc30820)

[1.1.6 Alertas en Tiempo Real 7](#_Toc17750)

[1.1.7 Desafíos en la Detección de Anomalías 8](#_Toc20546)

[1.1.8 Impacto de la Detección de Anomalías en la Industria 4.0 8](#_Toc4450)

[1.2 Fundamentación de la Tecnología a Usar 8](#_Toc17095)

[1.3 Conclusiones Parciales 10](#_Toc16634)

[Referencias Bibliográficas 12](#_Toc32042)

**Índice de figura**

# Introducción

La Industria 4.0 ha transformado el panorama de los sistemas industriales con la integración de tecnologías conectadas y sistemas ciberfísicos. Estos cambios han traído consigo la necesidad de implementar mecanismos de monitoreo avanzados para asegurar el correcto funcionamiento de los procesos industriales. En este contexto, garantizar la fiabilidad de los datos que generan y procesan estos sistemas es esencial para optimizar las operaciones y evitar interrupciones inesperadas. Para ello, es fundamental contar con sistemas capaces de identificar y responder a comportamientos anómalos que puedan comprometer la estabilidad del entorno industrial [1].

Un enfoque ampliamente utilizado para la protección de la continuidad operativa en sistemas industriales es la implementación de módulos de detección de anomalías. Estos módulos supervisan en tiempo real los datos generados por sensores y controladores, buscando desviaciones significativas que puedan ser indicadores de fallos operativos. Al detectar estas anomalías de manera temprana, es posible minimizar el impacto en los sistemas, garantizando su disponibilidad, integridad y eficiencia operativa [1-3].

Uno de los mayores desafíos en la detección de anomalías en sistemas industriales es la capacidad de procesar grandes volúmenes de datos provenientes de diversas fuentes en tiempo real. A esto se suma la necesidad de que los algoritmos de detección sean lo suficientemente precisos como para diferenciar entre fluctuaciones normales del sistema y comportamientos que podrían derivar en fallos o problemas operativos, evitando tanto falsos positivos como falsos negativos que puedan generar costos innecesarios o fallos no detectados [4].

El problema de investigación en este campo surge ante la creciente complejidad de los sistemas industriales y la necesidad de contar con mecanismos automatizados que puedan adaptarse a las condiciones cambiantes del entorno operativo [3, 5]. ¿Cómo se puede desarrollar un módulo de detección de anomalías en tiempo real que sea capaz de identificar tanto fallos operacionales como patrones anómalos en un sistema industrial simulado?

El objeto de estudio de esta investigación es un sistema industrial experimental que incluye sensores, microcontroladores STM32 y un PLC flexible basado en la plataforma OpenPLC. Este sistema simula un entorno industrial con capacidad para capturar y procesar datos operativos, ofreciendo un marco adecuado para implementar y probar el módulo de detección de anomalías [1].

Si bien existen tecnologías para el monitoreo de procesos industriales, la complejidad de los entornos y la cantidad de datos que deben analizarse en tiempo real siguen siendo un reto. Los fallos operativos pueden comprometer seriamente la producción, generando pérdidas económicas y disminuyendo la eficiencia de los sistemas. Por lo tanto, contar con un sistema que permita la detección temprana de fallos es crucial para evitar estos contratiempos [1, 4].

El uso de módulos de detección de anomalías en este contexto es fundamental, ya que permiten identificar patrones inusuales de comportamiento en los sistemas antes de que el daño sea irreversible. El campo de acción de esta investigación se centra en la implementación de algoritmos de detección de anomalías que puedan supervisar tanto las señales operativas de los sensores como el comportamiento general del sistema industrial experimental [4, 5].

El objetivo general de esta investigación es desarrollar un módulo de detección de anomalías que permita identificar comportamientos anómalos en tiempo real en un sistema industrial simulado, utilizando técnicas de minería de datos y aprendizaje automático aplicadas a los datos operativos recolectados por el sistema [2].

Para alcanzar este objetivo, se realizará la implementación de algoritmos de detección que se integrarán con los controladores STM32 y el PLC del sistema industrial. Además, se llevarán a cabo pruebas exhaustivas para validar la efectividad del módulo en la detección de anomalías, garantizando su capacidad de respuesta ante posibles fallos operacionales [2, 3].

El alcance de esta investigación incluye la detección de anomalías en señales operativas y la implementación de un sistema de alertas en tiempo real para que los operadores puedan tomar decisiones informadas sobre el mantenimiento y ajuste del sistema. Este enfoque proporcionará un marco claro para el desarrollo de la investigación y asegurará que se aborden los aspectos más críticos de la continuidad operativa en sistemas industriales [3, 4].

# Capítulo 1: Fundamentos Teóricos

Este capítulo establece los fundamentos teóricos necesarios para comprender y alcanzar los objetivos propuestos en este trabajo. Se discuten conceptos clave relacionados con la detección de anomalías, los sistemas industriales, los algoritmos de minería de datos y aprendizaje automático, y las tecnologías asociadas.

## 1.1 Detección de Anomalías en Sistemas Industriales

La detección de anomalías es una técnica fundamental utilizada para identificar eventos inusuales o patrones de comportamiento que se desvían de la operación normal de un sistema. En un sistema industrial, la identificación temprana de anomalías es crucial para prevenir fallos operacionales o deterioros en los equipos, lo que podría causar tiempos de inactividad no planificados o interrupciones en la producción. La detección de anomalías permite que los operadores puedan intervenir de manera preventiva y corregir posibles problemas antes de que estos escalen [6].

## 1.1.1 Tipos de Anomalías en Sistemas Industriales

En los sistemas industriales, las anomalías se pueden clasificar en tres tipos principales:

Anomalías puntuales: son eventos aislados que se desvían del comportamiento esperado. Por ejemplo, un sensor de temperatura que registra un valor anormal durante un breve período.

Anomalías contextuales: ocurren cuando un evento es anormal en un contexto específico. Por ejemplo, una temperatura elevada podría ser normal en un proceso, pero anómala en otro contexto operativo.

Anomalías colectivas: se refieren a un grupo de puntos de datos que, colectivamente, representan un comportamiento anómalo. Esto podría ser, por ejemplo, una tendencia inusual en los niveles de vibración de una máquina durante un período de tiempo.

## 1.1.2 Principios de la Detección de Anomalías

La detección de anomalías en sistemas industriales se basa en los siguientes principios fundamentales:

Monitorización en tiempo real: los datos operativos de sensores y controladores se analizan constantemente para detectar comportamientos anómalos.

Modelado del comportamiento normal: el sistema aprende a reconocer lo que se considera una operación normal del sistema industrial a través de la recopilación y análisis de datos históricos.

Respuesta a tiempo: una vez detectada una anomalía, el sistema genera alertas en tiempo real que permiten a los operadores tomar medidas correctivas antes de que ocurran daños graves.

## 1.1.3 Algoritmos para la Detección de Anomalías

Existen diferentes tipos de algoritmos utilizados para la detección de anomalías en sistemas industriales, entre ellos:

Métodos basados en reglas: se configuran reglas específicas que definen los umbrales operativos normales de un sistema. Cuando un dato excede estos umbrales, se detecta una anomalía.

Métodos estadísticos: se basan en modelos probabilísticos que evalúan la probabilidad de que un dato sea anómalo con respecto al comportamiento histórico del sistema.

Algoritmos de minería de datos y aprendizaje automático: estos métodos permiten el análisis de grandes volúmenes de datos operativos. Los modelos de aprendizaje automático pueden entrenarse para identificar patrones complejos que indican anomalías. Algoritmos como el k-means clustering, SVM (Support Vector Machines) y redes neuronales se utilizan comúnmente para este fin [7] .

## 1.1.4 Sistemas Industriales y su Monitoreo

Los sistemas industriales son entornos altamente interconectados donde diferentes dispositivos, como sensores, controladores lógicos programables (PLC), y sistemas SCADA (Supervisory Control and Data Acquisition), supervisan y controlan procesos físicos. La cantidad de datos que estos sistemas generan es considerable, y la detección de anomalías en estos sistemas requiere el procesamiento continuo de datos en tiempo real.

Sensores y controladores: monitorean variables clave como temperatura, presión, velocidad de motores, entre otros. Estos datos son fundamentales para la toma de decisiones en tiempo real.

PLC y SCADA: permiten la automatización de procesos y la recopilación de datos operativos. Estos datos proporcionan la base para aplicar algoritmos de detección de anomalías y generar alertas en caso de que ocurra un comportamiento fuera de lo esperado.

## 1.1.5 Minería de Datos y Aprendizaje Automático en la Detección de Anomalías

El uso de técnicas de minería de datos y aprendizaje automático ha revolucionado la manera en que se detectan las anomalías en sistemas industriales. Estas técnicas permiten analizar grandes cantidades de datos y detectar patrones sutiles que podrían no ser evidentes mediante métodos convencionales.

Minería de datos: implica la extracción de información útil a partir de grandes volúmenes de datos. En el contexto de los sistemas industriales, se aplica para identificar tendencias y patrones anómalos.

Aprendizaje automático: a través de técnicas supervisadas y no supervisadas, los algoritmos de aprendizaje automático pueden aprender a diferenciar entre el comportamiento normal y anómalo. Modelos como las redes neuronales profundas y los modelos de regresión son ampliamente utilizados en estos contextos [8, 9].

## 1.1.6 Alertas en Tiempo Real

El objetivo de la detección de anomalías no es solo identificar comportamientos inusuales, sino también generar alertas en tiempo real que permitan una intervención inmediata. Estas alertas pueden enviarse a los operadores a través de diferentes medios, como pantallas HMI (Interfaz Hombre-Máquina), dispositivos móviles o sistemas de monitoreo centralizado. La respuesta oportuna a las alertas es esencial para evitar fallos mayores en los sistemas.

## 1.1.7 Desafíos en la Detección de Anomalías

Algunos de los principales desafíos en la detección de anomalías en sistemas industriales incluyen:

Falsos positivos y negativos: uno de los mayores retos es garantizar que el sistema sea lo suficientemente preciso para minimizar las falsas alarmas (falsos positivos) y, al mismo tiempo, no pase por alto comportamientos anómalos reales (falsos negativos).

Procesamiento en tiempo real: debido a la gran cantidad de datos generados, los sistemas deben ser capaces de procesar la información en tiempo real sin comprometer el rendimiento de la operación industrial.

## 1.1.8 Impacto de la Detección de Anomalías en la Industria 4.0

La implementación de módulos de detección de anomalías contribuye significativamente a la Industria 4.0, proporcionando un nivel adicional de seguridad y optimización operativa. Estos módulos permiten que los sistemas industriales se vuelvan más resilientes, mejorando tanto la eficiencia como la seguridad operativa, y ayudando a prevenir interrupciones en la producción.

## 1.2 Fundamentación de la Tecnología a Usar

En el desarrollo de soluciones para la detección de anomalías en sistemas industriales, la elección de las tecnologías adecuadas es un paso crucial que puede influir significativamente en la eficacia y eficiencia de la solución final. En este contexto, hemos seleccionado varias tecnologías clave para emplear en el desarrollo de nuestra solución: el lenguaje de programación Python y sus bibliotecas asociadas, SQLite3 para la base de datos, y PyCharm como entorno de desarrollo integrado (IDE).

A continuación, presentamos una justificación detallada para la selección de cada una de estas tecnologías, destacando cómo su uso contribuirá a la eficacia de nuestra solución y cómo se alinean con los objetivos de nuestra investigación. Nuestra meta es proporcionar una base sólida para el desarrollo de una solución eficiente y eficaz que permita la identificación de comportamientos anómalos en sistemas industriales.

1. Python

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general que ha ganado popularidad en el ámbito de la ciencia de datos y la inteligencia artificial. Su sintaxis sencilla y clara permite un desarrollo rápido y eficiente. Python cuenta con una extensa comunidad de desarrolladores y una rica colección de bibliotecas que facilitan la implementación de algoritmos de detección de anomalías y procesamiento de datos. La flexibilidad de Python lo convierte en una excelente elección para prototipos y desarrollo ágil, permitiendo una rápida iteración y prueba de ideas.

2. Pandas y NumPy

Pandas: Es una biblioteca fundamental para la manipulación y análisis de datos en Python. Proporciona estructuras de datos flexibles como DataFrames, que facilitan la carga, manipulación y análisis de grandes volúmenes de datos operativos generados por sistemas industriales. Con Pandas, es posible realizar operaciones complejas de filtrado y transformación de datos de manera eficiente, lo que es crucial para preparar los datos antes de aplicar algoritmos de detección de anomalías.

NumPy: Esta biblioteca proporciona soporte para arreglos multidimensionales y operaciones matemáticas avanzadas, lo que es especialmente útil para cálculos numéricos y análisis estadístico. Al trabajar con datos de sensores en tiempo real, NumPy permite realizar cálculos eficientes que son esenciales para implementar algoritmos de aprendizaje automático en la detección de anomalías.

3. SQLite3

SQLite es un sistema de gestión de bases de datos relacional ligero y de código abierto. Su integración con Python a través del módulo sqlite3 facilita el almacenamiento y gestión de datos sin la necesidad de configurar un servidor de base de datos. Esto es ventajoso para proyectos de menor escala o durante las etapas de desarrollo y prueba, ya que permite un acceso rápido y eficiente a los datos sin complicaciones adicionales. SQLite es ideal para almacenar los registros de datos operativos y las alertas generadas por el sistema de detección de anomalías.

4. PyCharm

PyCharm es un entorno de desarrollo integrado (IDE) muy popular para Python, conocido por su robustez y su amplia gama de características que facilitan el desarrollo de software. Ofrece herramientas como autocompletado de código, depuración avanzada y control de versiones, lo que mejora la productividad del desarrollador y reduce la posibilidad de errores. Además, PyCharm proporciona integración con sistemas de control de versiones y soporte para bibliotecas de ciencia de datos, lo que lo convierte en una opción ideal para el desarrollo de soluciones complejas en detección de anomalías.

## 1.3 Conclusiones Parciales

A lo largo de este capítulo, se ha logrado establecer una base sólida para el desarrollo de un módulo de detección de anomalías en sistemas industriales. Se ha evidenciado la importancia de la detección de anomalías como un mecanismo crucial para garantizar la seguridad y continuidad operativa en entornos industriales. Este enfoque permite identificar de manera temprana comportamientos inusuales, lo que contribuye a prevenir fallos operativos y minimizar tiempos de inactividad.

Se ha definido y analizado un marco conceptual que abarca los distintos tipos de anomalías y los principios fundamentales que rigen la detección de estas en sistemas industriales. La identificación de métodos efectivos de análisis, particularmente aquellos basados en aprendizaje automático y minería de datos, ha demostrado ser esencial para abordar los desafíos de detección, como la reducción de falsos positivos y negativos.

La selección de Python como lenguaje de programación y el uso de bibliotecas como Pandas y NumPy han permitido plantear una estrategia efectiva para el manejo y análisis de grandes volúmenes de datos operativos. Estas tecnologías no solo optimizan la manipulación de datos, sino que también facilitan la implementación de algoritmos de detección de anomalías.

Además, el uso de SQLite3 para el almacenamiento de datos garantiza una gestión eficiente de la información recolectada, mientras que la elección de PyCharm como entorno de desarrollo proporciona las herramientas necesarias para una programación ágil y eficiente. Estas decisiones tecnológicas aseguran que el desarrollo del módulo se realice en un entorno propicio para la innovación y la efectividad.

En conclusión, el capítulo ha logrado establecer un marco teórico y práctico robusto que guiará el desarrollo del módulo de detección de anomalías, alineándose con los objetivos de la investigación y posicionando la solución propuesta como una herramienta efectiva para mejorar la seguridad y eficiencia en los sistemas industriales.

# Referencias Bibliográficas

1. Lasi, H., et al., *Industry 4.0.* Business & Information Systems Engineering, 2014. **6**: p. 239-242.

2. Chandola, V., A. Banerjee, and V. Kumar, *Anomaly Detection: A Survey.* ACM Comput. Surv., 2009. **41**.

3. Ahmed, M., A. Mahmood, and J. Hu, *A Survey of Network Anomaly Detection Techniques.* Journal of Network and Computer Applications, 2015. **60**: p. 19-31.

4. Qin, J., *Survey on data-driven industrial process monitoring and diagnosis.* Annual Reviews in Control, 2012. **36**: p. 220–234.

5. Hastie, T., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2009.

6. Grunova, D., et al., *Machine Learning for Anomaly Detection in Industrial Environments*. 2024. 25.

7. Liu, J., et al., *Anomaly Detection Method for Industrial Control System Operation Data Based on Time–Frequency Fusion Feature Attention Encoding.* Sensors, 2024. **24**: p. 6131.

8. Kim, B., et al., *A Comparative Study of Time Series Anomaly Detection Models for Industrial Control Systems.* Sensors, 2023. **23**: p. 21.

9. Mokhtari, S., et al., *A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data.* Electronics, 2021. **10**: p. 407.