

UNIVERSIDAD TECNOLÓGICA DE CHIHUAHUA

Ingeniería en Desarrollo y Gestión de Software



Extracción de Conocimiento en Bases de Datos

I.2. Reporte de solución de caso de estudio en el que indique la forma en que aplica la inteligencia artificial, machine learning, data mining y big data; indicando los beneficios, lenguajes y herramientas aplicables para el caso de estudio.

IDGS91N

PRESENTA:

Luis Eduardo Aguilar Sarabia
Abraham Camacho Ríos
Giselle Cantú Chávez Karla
Alejandra de la Cruz Zea
Ricardo Hernández Martínez

NOMBRE DEL DOCENTE:

Ing. Luis Enrique Mascote Cano

Chihuahua, Chih., 23 de septiembre de 2025

ÍNDICE

| | |
|---|----|
| INTRODUCCIÓN | 3 |
| JUSTIFICACIÓN DEL DOMINIO..... | 5 |
| BENEFICIOS Y KPIS | 7 |
| Mejora de la Precisión en la Detección de Anomalías | 7 |
| Reducción del Tiempo de Respuesta en la Gestión de Incidentes..... | 7 |
| Optimización de Costos Operativos en el Mantenimiento de la Red | 7 |
| HERRAMIENTAS Y LENGUAJES | 9 |
| ARQUITECTURA DE SOLUCIÓN (DIAGRAMA) | 13 |
| CONCLUSIONES Y PRÓXIMOS PASOS..... | 16 |
| REFERENCIAS Y FUENTES CONSULTADAS | 18 |

ÍNDICE DE FIGURAS

| | |
|----------------------------|-----------|
| Ilustración 1 | 9 |
| Ilustración 2 | 10 |
| Ilustración 3 | 11 |
| Ilustración 4 | 13 |

INTRODUCCIÓN

El presente caso de estudio aborda la detección de anomalías en redes de sensores IoT para smart cities, un desafío que ha cobrado relevancia ante la creciente digitalización de los entornos urbanos. La iniciativa CitySense busca monitorear variables como calidad del aire, ruido y temperatura mediante una red de sensores distribuidos en distintas zonas de la ciudad. No obstante, el uso de estos dispositivos conlleva la generación masiva de series temporales y el riesgo de presentar datos atípicos que pueden indicar fallas técnicas o eventos críticos. Para dar respuesta a esta necesidad, se requiere un sistema de análisis en tiempo real que identifique outliers y activar mecanismos de alerta, integrando técnicas de inteligencia artificial y machine learning como DBSCAN o Isolation Forest, ya aplicadas en proyectos similares de ciudades inteligentes (Romero et al., 2018; Hidalgo et al., 2025).

La importancia de desarrollar esta propuesta se relaciona con alcanzar la confiabilidad de los datos que sustentan la toma de decisiones en entornos urbanos conectados. Detectar anomalías oportunamente contribuye a evitar interrupciones en el servicio y también mejora la capacidad de respuesta ante contingencias ambientales o técnicas. Este tipo de investigaciones influyen positivamente en la sostenibilidad y la seguridad en el marco de las smart cities, ya que maneja de modo correcto los recursos, reduce riesgos y fortalece la resiliencia urbana frente a amenazas que van surgiendo.

Como equipo de trabajo, consideramos que la elaboración de este caso de estudio aporta directamente a nuestra formación profesional. El diseño de soluciones basadas en inteligencia artificial y minería de datos aumenta nuestras competencias en el uso de herramientas analíticas, programación y gestión de datos en tiempo real. Nos brinda, de igual forma, la oportunidad de aplicar de manera práctica los conocimientos adquiridos en asignaturas previas relacionadas con bases de datos, seguridad informática y análisis de sistemas, mientras nos acerca a problemáticas reales de gran impacto social como la optimización de ciudades

inteligentes. El proceso también fomenta el trabajo en equipo, la investigación académica y la capacidad de transformar necesidades urbanas en propuestas tecnológicas viables.

En cuanto a la organización del contenido, el documento se organiza en cuatro apartados principales. En la sección de justificación del dominio, explicamos por qué el machine learning se identifica como el enfoque más idóneo para resolver la problemática de detección de anomalías, considerando tanto los retos técnicos como las necesidades del contexto urbano. En la parte de beneficios y KPIs, detallamos las métricas que permitirán evaluar el éxito del sistema, incluyendo reducción de tiempos de respuesta, precisión en la detección y disminución de costos operativos. Posteriormente, en la arquitectura de solución, presentamos un diagrama de alto nivel que ilustra el flujo completo de datos: desde la adquisición en los sensores, pasando por el proceso ETL y el entrenamiento de modelos, hasta el despliegue de alertas y dashboards de monitoreo. Finalmente, en el apartado de conclusiones y próximos pasos, reflexionamos sobre las aportaciones del proyecto y sugerimos líneas de mejora e investigación futura.

JUSTIFICACIÓN DEL DOMINIO

En este caso de estudio seleccionamos el Machine Learning (ML) como el dominio más idóneo para abordar la detección de anomalías en redes de sensores IoT aplicadas a smart cities. La decisión se fundamenta en tres razones principales: la naturaleza de los datos, la capacidad predictiva de los modelos y la pertinencia en términos de negocio y operación.

En primer lugar, los sensores de la iniciativa CitySense generan series temporales de gran volumen y variabilidad, lo que exige herramientas que puedan identificar patrones ocultos y detectar desviaciones de manera automática. Las técnicas tradicionales de estadística descriptiva o reglas estáticas de umbrales son insuficientes, pues no se adaptan a contextos dinámicos ni a variaciones inesperadas. Por el contrario, el ML construye modelos entrenados con datos históricos que aprenden a reconocer el comportamiento normal y, en consecuencia, pueden identificar de manera eficiente los outliers que corresponden a fallas técnicas, ataques o eventos ambientales críticos (Hidalgo et al., 2025).

En segundo lugar, el ML ofrece algoritmos robustos para entornos con alta incertidumbre, como DBSCAN, Isolation Forest o técnicas basadas en redes neuronales. Estos algoritmos no requieren necesariamente etiquetas previas de anomalías y son capaces de adaptarse a la evolución de los datos en tiempo real, lo que resulta particularmente útil en el contexto del IoT urbano, donde las condiciones cambian constantemente (Romero et al., 2018). Al mismo tiempo, este enfoque reduce falsos positivos y proporciona explicaciones más confiables para los stakeholders responsables de la toma de decisiones (Berríos et al., 2021).

Finalmente, desde el punto de vista de negocio, implementar un sistema basado en ML ayuda a las ciudades a optimizar recursos, anticipar problemas y reducir costos asociados a fallos o emergencias. Detectar anomalías en etapas tempranas significa menor tiempo de inactividad de los servicios urbanos, mayor seguridad para la ciudadanía y mejor capacidad de respuesta ante contingencias ambientales. De acuerdo con la literatura reciente, la incorporación de algoritmos

de ML en entornos IoT se reconoce como una de las tendencias más efectivas para garantizar el funcionamiento sostenible de las infraestructuras críticas en smart cities (Arnau Muñoz, 2023). Esto también se refleja en análisis más amplios sobre el papel del IoT y la inteligencia artificial en la transformación de los entornos urbanos, donde se subraya la importancia de estos enfoques para la resiliencia y la sostenibilidad (Rodríguez et al., 2023).

BENEFICIOS Y KPI'S

El diseño de este sistema de detección de anomalías tiene beneficios medibles que evalúan su impacto real en la operación urbana. Para nuestro caso, hemos definido tres métricas de desempeño (KPIs) que valoran tanto la eficiencia técnica como la relevancia práctica de la solución.

Mejora de la Precisión en la Detección de Anomalías

Uno de los principales beneficios esperados es incrementar la precisión de los modelos de Machine Learning al identificar outliers en los datos de sensores. La efectividad de algoritmos como DBSCAN o Isolation Forest en escenarios urbanos se ha demostrado en estudios previos, donde su aplicación reduce significativamente los falsos positivos y aumenta la confiabilidad del sistema (Romero et al., 2018). Este indicador será medido a través del porcentaje de aciertos en la clasificación de eventos anómalos frente a datos de referencia validados.

Reducción del Tiempo de Respuesta en la Gestión de Incidentes

El segundo beneficio esperado es la capacidad de reaccionar con mayor rapidez ante situaciones críticas, como fallos de sensores o niveles elevados de contaminantes. La literatura señala que los sistemas automáticos de análisis y alerta les da oportunidad a los equipos de actuar en tiempo casi real, disminuyendo el riesgo para la ciudadanía y el entorno (Hidalgo Vargas et al., 2025). El KPI asociado corresponde al tiempo promedio de detección y notificación de una anomalía, comparado con los métodos tradicionales de supervisión manual.

Optimización de Costos Operativos en el Mantenimiento de la Red

Un tercer beneficio consiste en la reducción de costos vinculados al monitoreo y mantenimiento de los sensores, ya que la detección temprana de anomalías evita reparaciones tardías o pérdidas de datos significativas. Este beneficio se vincula directamente con la sostenibilidad de las ciudades inteligentes, al maximizar la eficiencia en el uso de recursos tecnológicos y financieros (Rodríguez et al., 2023). La métrica que se empleará en este caso es el porcentaje de disminución de costos asociados a la operación y mantenimiento de la red en comparación con escenarios previos.

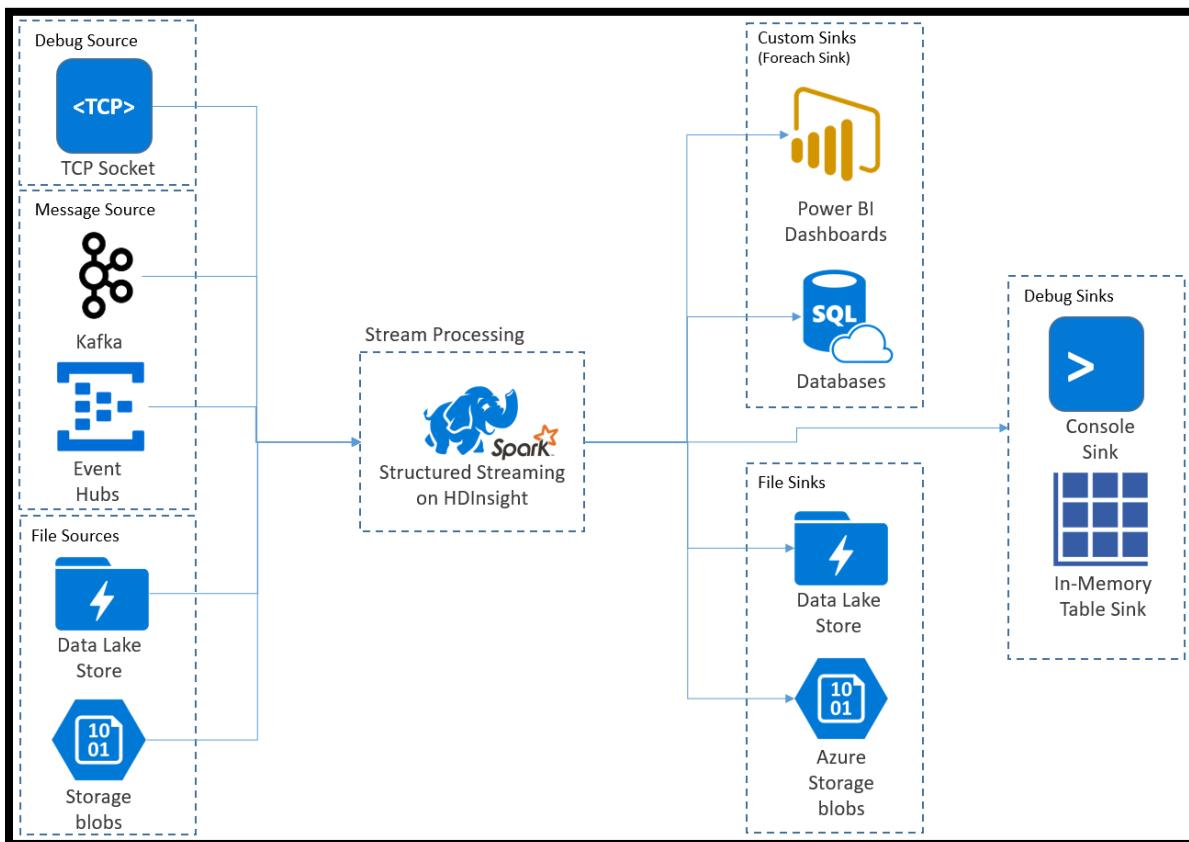
HERRAMIENTAS Y LENGUAJES

1. Apache Spark (Structured Streaming + batch/ETL)

Lo usaremos para ingerir el flujo continuo de sensores (puenteando MQTT→Kafka cuando aplique), aplicar ventanas y watermarks, y cargar tanto los agregados de tiempo real como el histórico al DW. Spark nos da elasticidad para crecer desde un piloto a ciudad completa y, además, facilita calcular KPIs diarios (por zona y tipo de sensor) sin duplicar lógica entre streaming y batch. La literatura sobre smart cities resalta precisamente el papel de big data y el análisis de datos como habilitadores de decisiones oportunas y escalables en entornos urbanos, lo que encaja con este uso de Spark.

Ilustración 1

Ejemplo de flujo de procesamiento de datos

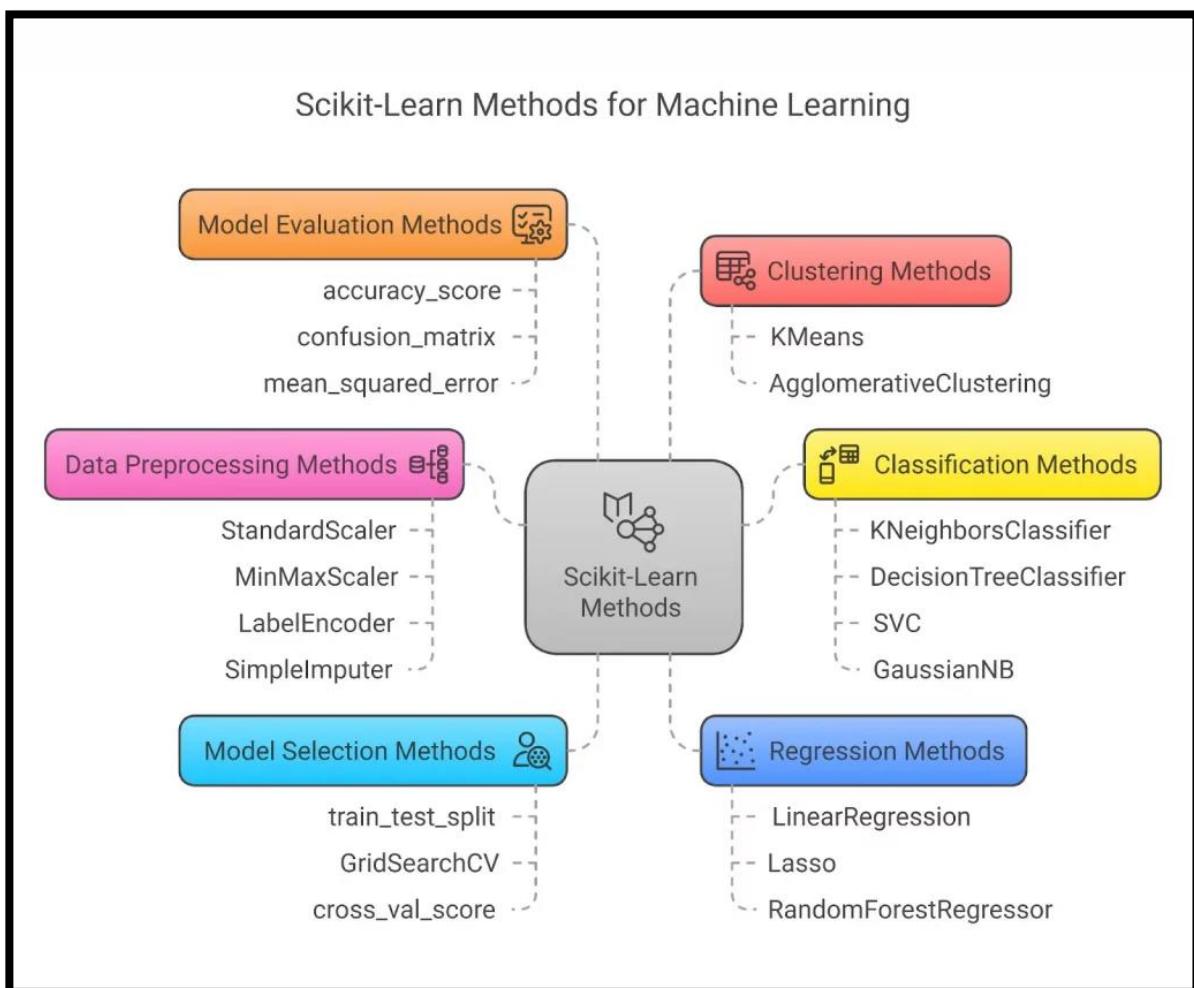


2. Python + scikit-learn (modelo de detección)

Prototiparemos y entrenaremos los detectores con Python y scikit-learn (p. ej., DBSCAN e Isolation Forest) por su velocidad de iteración y amplio ecosistema. En escenarios de ciudades inteligentes, DBSCAN se ha usado con éxito para detectar anomalías a partir de agrupamiento, lo que sustenta nuestra elección para datos no etiquetados. Hay trabajos académicos en IoT/Smart-City que emplean Python como lenguaje principal para el pipeline de análisis y modelado, lo cual nos da confianza en la mantenibilidad del stack.

Ilustración 2

Métodos para el Machine Learning en Scikit

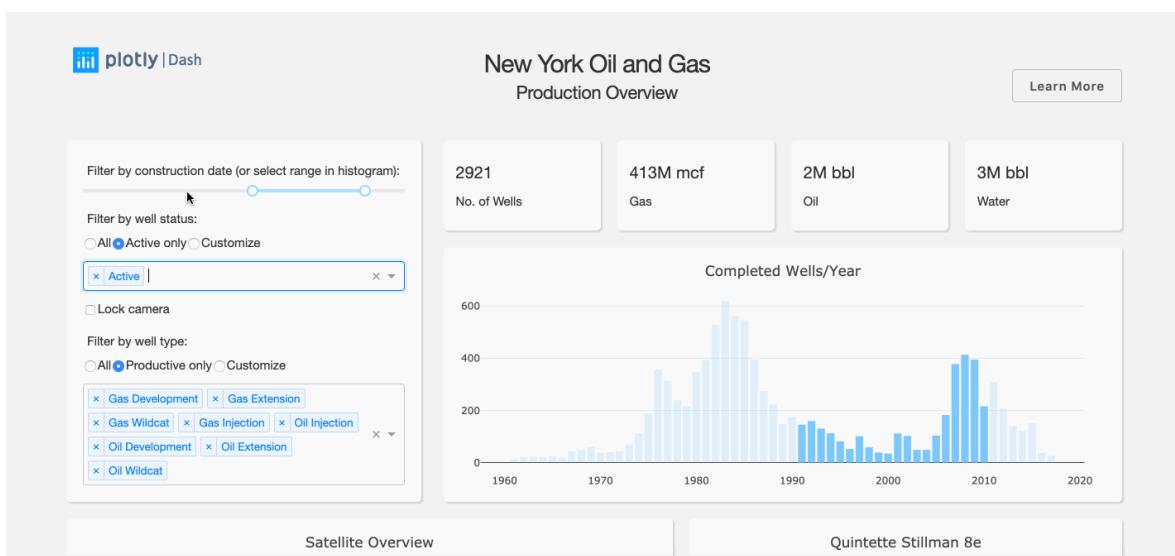


3. Dash (Plotly) para visualización operativa

Construiremos un dashboard web con gráficas de series temporales y resaltado de eventos, filtros por zona y tipo de sensor, y vista de alertas en tiempo casi real (consumiendo la API de inferencia). Elegimos Dash porque nos permite permanecer en el ecosistema Python (misma base de datos de métricas y mismo entorno de modelos) y publicar una interfaz interactiva sin levantar un frontend complejo, siendo consistente con soluciones previas que integran análisis en Python y una interfaz web ligera para explorar resultados en contextos de smart cities.

Ilustración 3

Ejemplo de uso de Dash (Plotly)

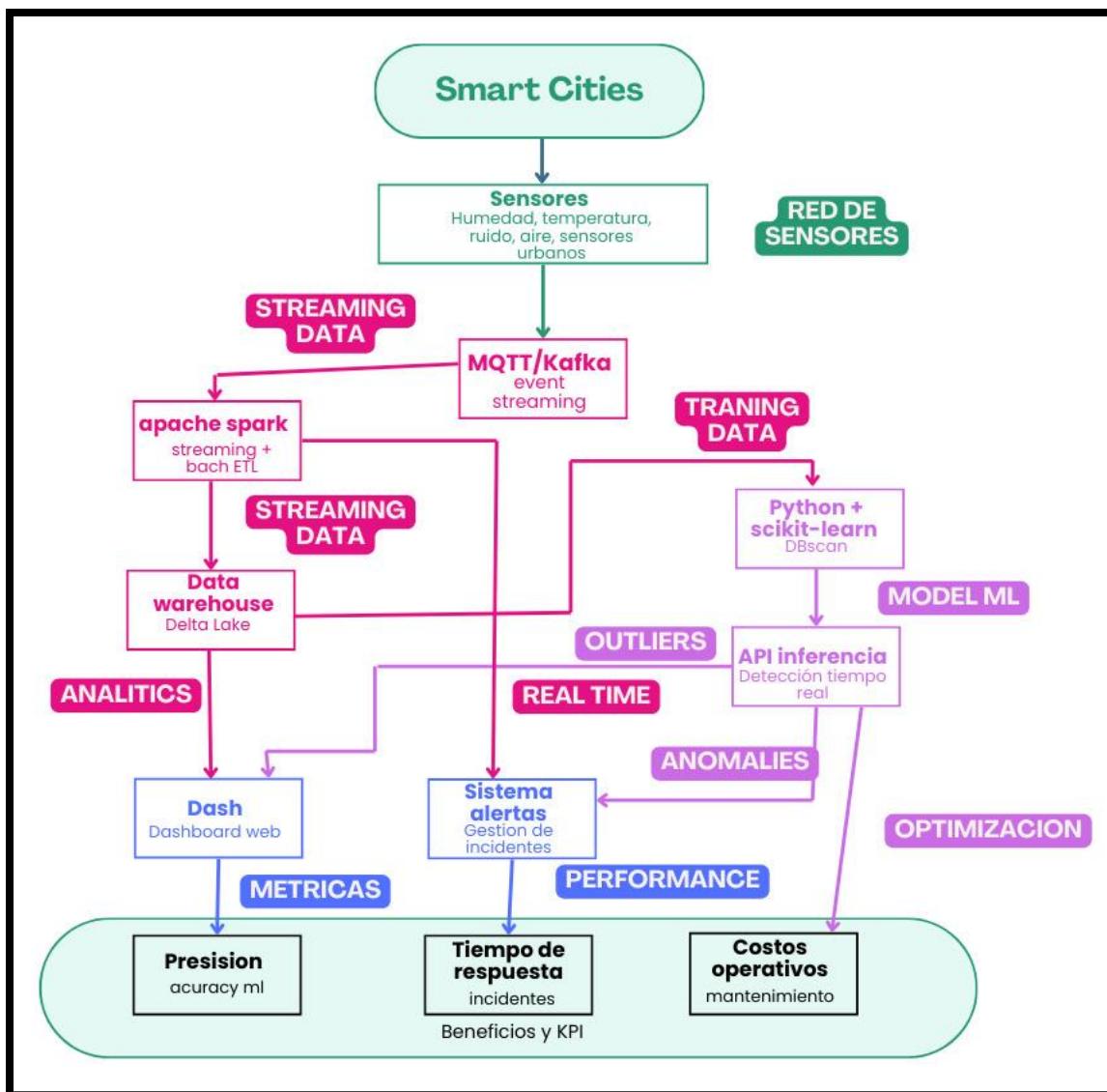


Mapa al pipeline (de punta a punta): sensores → (MQTT/Kafka) → Spark Structured Streaming (limpieza, ventanas, umbrales básicos) → DW histórico (parquet/Delta) → Python + scikit-learn (entrenamiento, validación y servicio de inferencia) → Dash (monitor, KPIs y anomalías resaltadas).

ARQUITECTURA DE SOLUCIÓN (DIAGRAMA)

Ilustración 4

Diagrama de la arquitectura que desarrollamos para solucionar la problemática



En el diagrama mostramos la arquitectura de solución que diseñamos para la detección de anomalías en redes de sensores IoT aplicadas a smart cities. La lectura del flujo va de arriba hacia abajo y de izquierda a derecha, representando el recorrido que siguen los datos desde su origen hasta la visualización de resultados y métricas.

Todo inicia en la red de sensores urbanos, que recogen información de variables ambientales como humedad, temperatura, ruido y calidad del aire. Estos sensores envían los datos en tiempo real a través de protocolos de mensajería como MQTT o Kafka, que permiten transmitir flujos de eventos de manera continua y confiable.

A partir de este punto, el flujo se divide en dos caminos complementarios. Por un lado, los datos en streaming ingresan a Apache Spark, donde se ejecutan procesos ETL tanto en tiempo real como en lotes. Esto garantiza la limpieza, transformación y almacenamiento de la información en un Data Warehouse tipo Delta Lake, que sirve como repositorio histórico para análisis posteriores. Desde ahí se generan métricas y visualizaciones en Dash, un dashboard web que facilita la exploración de resultados.

Por otro lado, los datos también alimentan la etapa de entrenamiento de modelos con Python y scikit-learn, donde aplicamos algoritmos como DBSCAN para detectar outliers. Estos modelos se integran en una API de inferencia que opera en tiempo real, identificando anomalías y enviando alertas cuando se superan ciertos umbrales críticos. El componente de gestión de alertas se conecta directamente con los equipos de respuesta, asegurando que la información detectada se traduzca en acciones concretas.

Finalmente, vinculamos toda la arquitectura con tres métricas de desempeño:

- ✓ Precisión del modelo (accuracy ML) para evaluar la calidad de la detección.
- ✓ Tiempo de respuesta ante incidentes, que mide la rapidez con la que se genera y comunica una alerta.

- ✓ Costos operativos de mantenimiento, que reflejan la optimización lograda al detectar fallas temprano y evitar reparaciones tardías.

CONCLUSIONES Y PRÓXIMOS PASOS

Con este caso de estudio logramos responder la pregunta: ¿es posible detectar anomalías en una red de sensores IoT aplicada a una ciudad inteligente de manera eficaz? Nuestra conclusión es que sí y que el Machine Learning es el dominio adecuado para hacerlo. Identificamos que algoritmos como DBSCAN e Isolation Forest pueden adaptarse a datos en tiempo real y encontrar irregularidades que no serían visibles con reglas fijas.

Durante el desarrollo del trabajo comprobamos que diseñar una solución de este tipo requiere pensar en todo el flujo: la recolección de datos desde sensores, el preprocesamiento y almacenamiento, el entrenamiento de modelos y, finalmente, la visualización y monitoreo de resultados. Esta secuencia nos permitió estructurar una propuesta clara y viable que cubre las necesidades de un proyecto como el de CitySense.

El aprendizaje para nosotros fue doble. Por un lado, reforzamos conocimientos técnicos al aplicar conceptos de ingestión de datos, ETL y algoritmos de detección de anomalías en un contexto real. Por otro, nos dimos cuenta de que trabajar en equipo con un objetivo común nos ayudó a organizar mejor el proyecto y a comunicar nuestras ideas de manera más precisa. La experiencia nos acerca al tipo de retos que vamos a enfrentar en nuestra profesión.

En cuanto a los próximos pasos, identificamos varios puntos a desarrollar. Primero, probar el sistema con datos de sensores reales para validar la precisión de los modelos en condiciones de ruido e inestabilidad. Segundo, explorar algoritmos más avanzados que nos permitan detectar patrones complejos en series temporales largas. Tercero, mejorar los mecanismos de visualización y alertas, con el fin de que los resultados sean accesibles no solo para especialistas, sino también para equipos operativos encargados de la respuesta. Cuarto, integrar un monitoreo histórico que sirva para generar reportes diarios por zona y tipo de sensor, lo que complementaría la detección en tiempo real con una visión de largo plazo.

La propuesta desarrollada expone cómo la aplicación de inteligencia artificial machine learning y técnicas de minería de datos permite detectar anomalías en redes de sensores IoT en el contexto de smart cities. El uso de algoritmos como DBSCAN e Isolation Forest, en conjunto con herramientas de big data y un pipeline de análisis, proporciona datos confiables que respaldan la toma de decisiones y facilitan una respuesta oportuna ante fallas técnicas o eventos ambientales críticos. Con esta solución se logra optimizar recursos, reducir riesgos y fortalecer la sostenibilidad de las infraestructuras urbanas.

El proyecto también fortaleció las competencias del equipo en programación, análisis de datos y visualización, además de impulsar la investigación y el trabajo colaborativo. El ejercicio de trasladar conceptos teóricos al diseño de una solución práctica mostró el valor de la formación profesional para generar propuestas tecnológicas con impacto en el entorno social. La solución planteada establece un punto de partida para avanzar hacia sistemas de análisis predictivo y estrategias de gestión que mejoren el funcionamiento de las ciudades inteligentes.

REFERENCIAS Y FUENTES CONSULTADAS

- Arnau Muñoz, L. (2023). *Sistema de detección de anomalías para infraestructuras IoT* [Trabajo de fin de máster, Universidad de Alicante]. Repositorio Institucional RUA. <http://hdl.handle.net/10045/135258>
- Berríos, S., Hermosilla, P., Parra, H. B., y Oñate, P. (2021). *Modelo híbrido de dos niveles para la detección de anomalías con técnicas de machine learning en redes IoT* [Tesis de máster, Pontificia Universidad Católica de Valparaíso y Universidad Católica de Salta].
https://cms.digilab.ucasal.edu.ar/uploads/42_Modelo_Hibrido_de_dos_niveles_para_la_deteccion_1e08033c3e.pdf
- Fernández Lobo, A. E. (2021). *Detección de patrones en datos espaciales para aplicaciones con geolocalización en smart cities* [Tesis de máster, Universidad Politécnica de Madrid, Escuela Técnica Superior de Ingenieros Informáticos]. Archivo Digital UPM. <https://oa.upm.es/67907/>
- Hidalgo Vargas, L. U., León Borges, J. A., Ramírez Pacheco, J. C., Toral Cruz, H., Makita Balcorta, T. G., y Osuna Galán, I. (2025). Análisis de revisión sistemática de la aplicación de algoritmos de aprendizaje automático en sistemas de detección de intrusión en Internet de las cosas para ciudades inteligentes. *Ciencia Latina Revista Científica Multidisciplinaria*, 8(6), 11500-11517. https://doi.org/10.37811/cl_rcm.v8i6.15929
- Rodríguez Rodríguez, I., Campo Valera, M., y Calderón Fajardo, V. (2023). *Conectando el futuro: Ciudades inteligentes, IoT y la transformación de la sociedad urbana*. UMA Editorial.
<https://dialnet.unirioja.es/descarga/libro/973010.pdf>
- Romero, I., Alonso, C., Villagrá, V., Vázquez, L., y Holgado, P. (2018). Aplicación de técnicas de detección de anomalías a escenarios de ciudades inteligentes. En *XIII Jornadas de Ingeniería Telemática (JITEL 2017)* (pp.

182-188). Editorial Universitat Politècnica de València.

<https://doi.org/10.4995/JITEL2017.2017.6574>