**REPUBLICA DE PANAMÁ**
**UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA**
**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

# DETECCION DE PICOS DE TRAFICO CON APRENDIZAJE AUTOMATICO PARA OPTIMIZACION OPERATIVA

**PROYECTO DE GRADUACION PARA OPTAR AL GRADO DE MAGISTER EN REDES DE COMUNICACIONES CON ENFASIS EN GESTION DE LAS TELECOMUNICACIONES**

AUTOR: Roberto Chifundo

TUTOR: José Munive

Ciudad de Panama de marzo de 2026

**REPUBLICA DE PANAMÁ**
**UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA**
**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

# DETECCION DE PICOS DE TRAFICO CON APRENDIZAJE AUTOMATICO PARA OPTIMIZACION OPERATIVA

**PROYECTO DE GRADUACION PARA OPTAR AL GRADO DE MAGISTER EN REDES DE COMUNICACIONES CON ENFASIS EN GESTION DE LAS TELECOMUNICACIONES**

AUTOR: Roberto Chifundo

Ciudad de Panama, marzo de 2026

Ciudad de Panama, marzo de 2026

Profesor

Dr. Héctor Mazurkiewickz

Coordinador del Comité de Titulación de Estudios de Grado y Postgrado Presente.

En mi carácter de Tutor del proyecto de graduación de maestría, presentado por el estudiante Roberto Chifundo, para optar por el grado de, maestría en redes de comunicaciones con énfasis en gestión de las telecomunicaciones

Considero que el trabajo: reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del Jurado examinador que se designe.

Atentamente,

Prof. José Munive

**UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA**

**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

# INFORME DE ACTIVIDADES DE TUTORÍA OPCIÓN DE TITULACIÓN DE PROYECTO DE GRADUACION DE MAESTRÍA

Programa de Maestría: Maestría en redes de comunicaciones con énfasis en gestión de las telecomunicaciones

Estudiante: Roberto Chifundo Wardrope

Tutor: Prof. José Munive

Título del proyecto de graduación de maestría: Detección de picos de tráfico con aprendizaje automático para optimización operativa.

Línea de Investigación: Gestión y monitoreo de redes de datos

| SESION | FECHA | HORA DE REUNION | ASPECTO TRATADO | OBSERVACION |
|---|---|---|---|---|
| 1 | 16-11-2025 | 02:37 pm | Revisión de Ante Proyecto | Revisión sobre los resultados y enfoque del proyecto |
| 2 | 17-11-2025 | 03:53 pm | Discusión sobre resultados | Mitigación de fallos y detecciones. |
| 3 | 18-12-2025 | 02:30 pm | Metodología | Diseño conceptual |
| 4 | 09-02-2026 | 12:30 pm | Revisión de proyecto final | Correcciones de enfoque. |

| | | | | |
|---|---|---|---|---|
| 5 | 02-22-2026 | 3:41 pm | Revisión final y firma de resultados | |
| 6 | | | | |

Comentarios finales acerca de la investigación: Declaramos que las

especificaciones anteriores representan el proceso de dirección del proyecto de graduación antes mencionado.

José Munive

Roberto Chifundo

## DEDICATORIA

Dedico este trabajo, en primer lugar, a Dios, por concederme salud, sabiduría y fortaleza para sostener la disciplina necesaria y culminar esta etapa de formación profesional.

A mi familia, por su amor, paciencia y apoyo incondicional. Gracias por acompañarme en los días de mayor exigencia, por comprender los sacrificios que implicó este proceso y por recordarme, con su ejemplo, el valor del esfuerzo y la constancia.

A mis padres, por los principios que sembraron en mí y por impulsar mi crecimiento personal y académico. A ustedes les debo gran parte de la determinación con la que enfrento los retos, así como la convicción de que el estudio y el trabajo bien hecho abren caminos.

A mis hermanos y seres queridos, por sus palabras de aliento y por celebrar conmigo cada avance, incluso cuando parecía pequeño. Su compañía ha sido una fuente permanente de motivación.

A mis docentes y asesores, por su guía académica, su exigencia y sus observaciones oportunas. Su acompañamiento fortaleció mi criterio profesional y contribuyó a elevar el rigor y la calidad de este trabajo.

A mis compañeros y amistades, por el apoyo emocional, el intercambio de ideas y la solidaridad en momentos clave. Su presencia hizo más llevadero el recorrido.

Finalmente, dedico este trabajo a todas las personas que creen en la mejora continua y en la tecnología como herramienta para generar soluciones reales. Que este esfuerzo sea también un aporte útil para quienes gestionan redes empresariales y buscan optimizar su operación mediante enfoques modernos y basados en datos.

## INDICE GENERAL


| Sección | Página |
|---|---|
| Portada interna | 2 |
| Carta de aprobación del tutor | 3 |
| Informe de actividades de tutoría | 4 |
| Dedicatoria | 6 |
| Índice general | 8 |
| Índice de tablas | 9 |
| Índice de figuras | 10 |
| Resumen | 12 |
| Abstract | 13 |
| Introducción | **14** |
| **Capítulo I. Planteamiento del problema** | **16** |
| 1.1. Descripción y contexto del problema | 17 |
| 1.2. Formulación del problema | 22 |
| 1.3. Objetivos | 22 |
| 1.4. Justificación | 23 |
| 1.5. Alcance y delimitaciones | 25 |
| **Capítulo II. Marco teórico** | **27** |
| 2.1. Redes LAN empresariales y gestión de tráfico | 28 |
| 2.2. Monitoreo mediante SNMP y Observium | 29 |
| 2.3. Monitoreo mediante NetFlow/IPFIX | 30 |
| 2.4. Picos de tráfico y detección de anomalías | 30 |
| 2.5. Aprendizaje automático no supervisado | 31 |
| 2.6. Trabajos relacionados | 33 |
| **Capítulo III. Metodología** | **34** |
| 3.1. Tipo y enfoque de estudio | 35 |
| 3.2. Fuentes de datos y herramientas | 36 |
| 3.3. Diseño del *pipeline* analítico | 37 |
| 3.4. Criterios de evaluación de anomalías | 38 |
| 3.5. Consideraciones éticas y de privacidad | 38 |

| Sección | Página |
| --- | --- |
| 3.6. Plan de Trabajo (Cronograma) | 39 |
| **Capítulo IV. Resultados y análisis** | **41** |
| 4.1. Diseño del modelo propuesto | 42 |
| 4.2. Estrategia de validación | 45 |
| 4.3. Análisis conceptual de los resultados | 47 |
| 4.4 Limitaciones y mejoras futuras | 49 |
| Conclusiones | 51 |
| Recomendaciones | 55 |
| Referencias | 57 |
| **Anexos** | **61** |
| Anexo A. Serie temporal | **62** |
| Anexo B. Diccionario de datos (SNMP y NetFlow/IPFIX) | 62 |
| Anexo C. Evidencia de configuración | 65 |
| Anexo D. Prompts Utilizados para facilitar la generación de detalles técnicos | 67 |

## INDICE DE TABLAS


| Tabla No. | Título | Página |
|---|---|---|
| Tabla 1 | Cronograma de trabajo propuesto | 39 |
| Tabla 2 | Ejemplo de tabla de picos de tráfico detectados | 48 |
| Tabla 3 | Comparación de algoritmos para detección de picos de tráfico (enfoque no supervisado) | 48 |

## INDICE DE FIGURAS


| Figura No. | Título | Página |
|---|---|---|
| Figura 1 | Arquitectura general de la solución propuesta | 43 |

**REPUBLICA DE PANAMÁ**
**UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA**
**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

# DETECCION DE PICOS DE TRAFICO CON APRENDIZAJE AUTOMATICO PARA OPTIMIZACION OPERATIVA

Autor: Roberto Chifundo
Tutor: José Munive
Año: 2026

## RESUMEN

La saturación de ancho de banda por picos inesperados representa un desafío constante en redes LAN empresariales, ya que compromete el rendimiento de servicios críticos y dificulta la gestión operativa. Este trabajo propone y examina un enfoque basado en aprendizaje automático no supervisado para detectar automáticamente estos eventos anómalos, buscando establecer fundamentos que permitan un monitoreo proactivo y superen las restricciones de los métodos convencionales basados en umbrales fijos.
El texto describe una metodología para detectar anomalías en redes mediante aprendizaje automático no supervisado. La propuesta establece un proceso completo que va desde la captura de datos de red (usando protocolos SNMP y NetFlow/IPFIX) hasta su preparación mediante limpieza y normalización. Se evalúan tres algoritmos principales Isolation Forest, One-Class SVM y DBSCAN que pueden identificar comportamientos anómalos sin necesidad de entrenamiento previo con datos etiquetados. El estudio utiliza datos anonimizados o simulados para garantizar la privacidad y permitir que otros investigadores puedan replicar los experimentos.
El análisis conceptual sugiere que esta aproximación ofrece ventajas relevantes: modelado adaptativo del tráfico “normal”, detección de anomalías de diversa duración y magnitud, y posibilidad de relacionar los eventos detectados con componentes específicos de la infraestructura, como interfaces o segmentos de red.
Como conclusiones se resalta la viabilidad teórica de combinar telemetría de red con técnicas no supervisadas para fortalecer el monitoreo proactivo en entornos corporativos. Entre las recomendaciones para futuros desarrollos prácticos se plantean: definir umbrales dinámicos basados en patrones históricos, establecer protocolos claros de respuesta y escalamiento ante alertas, y diseñar mecanismos periódicos de revisión.

**Palabras clave:** Detección de anomalías, aprendizaje automático, redes LAN, picos de tráfico, SNMP, NetFlow.

**REPUBLICA DE PANAMÁ**
**UNIVERSIDAD INTERNACIONAL DE CIENCIA Y TECNOLOGÍA**
**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN Y TECNOLOGÍA**

# DETECCION DE PICOS DE TRAFICO CON APRENDIZAJE AUTOMATICO PARA OPTIMIZACION OPERATIVA

Autor: Roberto Chifundo
Tutor: José Munive
Año: 2026

## ABSTRACT

Bandwidth saturation due to unexpected spikes is a constant challenge in enterprise LANs, as it compromises the performance of critical services and hinders operational management. This paper proposes and examines an unsupervised machine learning approach to automatically detect these anomalous events, seeking to establish foundations that enable proactive monitoring and overcome the limitations of conventional methods based on fixed thresholds.

The text describes a methodology for detecting anomalies in networks using unsupervised machine learning. The proposal establishes a complete process ranging from network data capture (using SNMP and NetFlow/IPFIX protocols) to data preparation through cleaning and normalization. Three main algorithms are evaluated: Isolation Forest, One-Class SVM, and DBSCAN, which can identify anomalous behavior without the need for prior training with labeled data. The study uses anonymized or simulated data to ensure privacy and allow other researchers to replicate the experiments.

Conceptual analysis shows that this method has benefits. It can model traffic detect unusual events of different lengths and severity and connect these events to specific parts of the infrastructure like network interfaces or segments.

The study finds that combining network data with automatic analysis techniques can work well for improving monitoring in businesses. To make this practical some recommendations are made. These include setting limits based on past traffic patterns, creating clear rules, responding to an escalating alert and regularly reviewing the model and data to keep it relevant in real-world networks.

**Keywords**: Anomaly detection, machine learning, LAN networks, traffic peaks, SNMP, NetFlow

## INTRODUCCION

En la mayoría de las organizaciones, las redes de área local (LAN) constituyen el entorno donde converge la operación tecnológica diaria, ya que articulan la comunicación entre usuarios, servidores y un conjunto creciente de dispositivos. Actualmente, su función trasciende el intercambio básico de correo, navegación web o archivos compartidos. La LAN también respalda servicios esenciales para la continuidad operativa, entre ellos la telefonía IP, los sistemas de videovigilancia, el control de acceso, las aplicaciones corporativas y los escritorios virtuales. Al mismo tiempo, la digitalización ha multiplicado el número de dispositivos conectados y ha normalizado el uso de aplicaciones intensivas en datos videoconferencias, plataformas en la nube y analítica. Esto incrementa la demanda de ancho de banda y vuelve el tráfico más irregular, con picos que aparecen sin aviso y no siempre siguen un patrón predecible.

A esta dinámica se suma la convergencia entre redes operativas (OT) y el Internet de las Cosas (IoT), que introduce flujos intermitentes y de naturaleza muy diversa. Aunque muchos de estos eventos duran poco, pueden presionar enlaces específicos, elevar la latencia y afectar el desempeño de aplicaciones sensibles al tiempo. Las consecuencias se reflejan tanto en la continuidad del servicio como en la productividad, especialmente cuando el impacto recae sobre sistemas críticos. Por ello, la visibilidad de extremo a extremo deja de ser un “extra” técnico y pasa a ser un requisito: la red debe responder a cargas cambiantes y mantener disponibles los servicios y activos digitales de forma consistente.

En muchas redes empresariales, el monitoreo sigue basándose en métricas agregadas y esquemas tradicionales de umbrales estáticos. Si bien este enfoque permite

identificar tendencias generales, es poco efectivo para detectar anomalías puntuales o eventos intermitentes dentro de ventanas temporales cortas. Además, con frecuencia no existen mecanismos ágiles para correlacionar un pico de tráfico con su causa raíz, ya sea por procesos operativos (p. ej., respaldos no programados) o por comportamientos atípicos vinculados con incidentes de seguridad. Por ello, una gestión verdaderamente proactiva requiere capacidades que vayan más allá de observar: se necesitan herramientas que interpreten el contexto del tráfico y faciliten respuestas operativas informadas ante desviaciones respecto del comportamiento base de la red.

Bajo esta perspectiva, se considera el aprendizaje automático como una herramienta pertinente para anticipar comportamientos atípicos en la red, de manera que en este trabajo de maestría se propone el diseño y la validación de un modelo basado en técnicas de aprendizaje no supervisado, el cual integra información recolectada a través de telemetría y el análisis de flujos de tráfico para identificar irregularidades de forma autónoma. A lo que va el planteamiento es a explorar, desde un enfoque analítico, cómo este tipo de técnicas podrían aportar valor en la anticipación y comprensión de anomalías, sin asumir una implementación operativa en un entorno real. El objetivo es proponer una solución técnicamente sustentada, capaz de integrarse con plataformas de monitoreo existentes para mejorar la operación de redes LAN. Con ello, se busca elevar la calidad del servicio, reducir tiempos de diagnóstico y avanzar hacia una gestión más predictiva que respalde la toma de decisiones basada en datos.

Finalmente, y en apego a los lineamientos institucionales para la introducción, este documento sintetiza el propósito del estudio y orienta al lector sobre la estructura del informe. En los capítulos siguientes se delimita la problemática asociada con picos de

tráfico en redes LAN empresariales; se revisan fundamentos teóricos sobre detección de anomalías y telemetría de red; y se describe la metodología para el procesamiento de métricas y flujos (p. ej., SNMP y NetFlow/IPFIX), así como la evaluación del enfoque no supervisado mediante indicadores de desempeño. Por último, se presentan los resultados y su análisis, además de conclusiones y recomendaciones dirigidas a fortalecer la toma de decisiones operativas y la planificación de capacidad en entornos empresariales.

# CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA

En este capítulo se presentan las motivaciones, interrogantes y metas que guían el estudio, sirviendo como base para toda la investigación. El análisis comienza en la sección 1.1 al revisar el contexto de las redes LAN empresariales, donde la alta demanda de servicios críticos hace que los picos de tráfico pongan en riesgo la continuidad del negocio. A partir de este problema, en el apartado 1.2 se formula la pregunta de investigación, centrada en si es viable usar un modelo de aprendizaje no supervisado que combine telemetría SNMP con flujos NetFlow/IPFIX para mejorar la detección.

Para resolver este reto, la sección 1.3 define los objetivos que marcan la ruta técnica y metodológica a seguir en el proyecto. Después, en el apartado 1.4 se justifica el estudio al comparar las fallas de los monitoreos tradicionales de umbrales rígidos contra la agilidad de los métodos predictivos que aquí se proponen. Por último, la sección 1.5 establece las delimitaciones del trabajo, dejando claro el alcance técnico y los límites que aseguran que los resultados sean válidos.

## 1.1 Descripción y Contexto del Problema

Dentro de los entornos LAN corporativos, el consumo de caudal de datos suele exhibir métricas bastante constantes y predecibles, vinculadas a la jornada de oficina, el reparto de tareas por departamentos y la naturaleza operativa de las herramientas empresariales. Sin embargo, dicha estabilidad puede verse interrumpida por incrementos súbitos de tráfico que sobrepasan los umbrales de consumo previstos y transforman drásticamente el comportamiento habitual del flujo de red.

Estas fluctuaciones pueden derivar de contextos heterogéneos, tales como la ejecución de copias de seguridad programadas hacia nodos locales o servicios cloud, la distribución masiva de parches de software y sistemas operativos, la sincronización de alto volumen en entornos de trabajo híbrido, el intercambio de ficheros pesados entre infraestructura de servidor y terminales, además de acciones imprevistas de usuarios, rutinas automatizadas o servicios que emiten ráfagas de datos por errores lógicos, bucles de reintento o parámetros de red deficientes.

Si estos incidentes no se detectan y administran con celeridad, tienden a saturar los enlaces troncales, elevar la latencia de forma crítica, incrementar la fluctuación del retardo (jitter), causar descartes de paquetes y mermar la calidad de aplicaciones en tiempo real, como la voz sobre IP, las transmisiones de video, los módulos ERP y las instancias de VDI. En situaciones de alta gravedad, un colapso persistente del ancho de banda puede arriesgar la operatividad de funciones vitales, lastrar la eficiencia laboral y provocar fallos sistémicos con repercusiones inmediatas en la resiliencia y estabilidad del negocio.

El desafío principal para la administración de red consiste en detectar estos picos con suficiente anticipación y precisión, diferenciándolos de fluctuaciones normales del tráfico (por ejemplo, micro variaciones asociadas a la actividad cotidiana de los usuarios o incrementos breves por cargas legítimas). De manera complementaria, resulta crítico establecer correlaciones que permitan atribuir la causa a un elemento concreto del entorno, como una interfaz específica, un equipo determinado, un segmento VLAN o un conjunto de flujos asociados a direcciones IP, puertos y protocolos.

En este contexto, para una institución beneficiaria representada como una red corporativa genérica, se requiere una herramienta analítica que complemente la plataforma de monitoreo existente, incorpore contexto operativo y emita alertas accionables. Esto implica no solo identificar la presencia de una anomalía, sino también aportar evidencia sobre dónde ocurre, cuándo inicia, cómo evoluciona y qué entidades participan, con el propósito de facilitar una respuesta más rápida y eficaz, reducir el tiempo de diagnóstico y habilitar una gestión proactiva de la capacidad y del rendimiento de la red.

### 1.1.1 Antecedentes del problema

La detección de eventos atípicos en el comportamiento del tráfico se ha estudiado ampliamente bajo enfoques no supervisados, debido a que en entornos reales suele existir escasez de datos etiquetados y alta variabilidad del “comportamiento normal” (Liu et al., 2008; Schölkopf et al., 2001). En esta línea, Isolation Forest plantea un mecanismo de aislamiento eficiente para identificar observaciones raras como anomalías (Liu et al., 2008), mientras que One-Class SVM sustenta su uso en estimar el soporte de una distribución para separar lo normal de lo anómalo (Schölkopf et al., 2001).En paralelo, el uso de métodos de agrupamiento ha servido como base para distinguir regiones densas (comportamientos frecuentes) de puntos dispersos (ruido o anomalías). DBSCAN, propuesto como un algoritmo basado en densidad capaz de descubrir clústeres y tratar el “ruido” explícitamente, se considera pertinente cuando el conjunto de datos presenta mezclas de comportamientos y valores atípicos (Ester et al., 1996). Esta característica es especialmente útil en escenarios donde el tráfico de red presenta patrones recurrentes y, de forma intermitente, desviaciones de magnitud considerable (Ester et al., 1996).

Desde la perspectiva de la telemetría, la documentación técnica distingue entre las mediciones agregadas de desempeño y los registros detallados del comportamiento por flujo. En el ámbito de las redes empresariales, el protocolo SNMP se reconoce como un mecanismo de monitoreo y gestión esencial para obtener información operativa básica de los dispositivos (GeeksforGeeks, 2025b). Asimismo, NetFlow y su evolución hacia IPFIX habilitan el análisis profundo del tráfico mediante registros de flujo, lo que permite comprender qué entidades y conversaciones contribuyen específicamente al consumo de ancho de banda (Claise et al., 2013; Zenarmor, 2024). En la práctica, el monitoreo de infraestructura suele apoyarse en plataformas que consolidan la recolección y visualización histórica de métricas, tales como Observium (Observium Limited, 2026).

En síntesis, la literatura técnica respalda (a) la implementación de modelos no supervisados como Isolation Forest, One-Class SVM y DBSCAN para la detección de comportamientos atípicos, y (b) la utilidad estratégica de combinar telemetría agregada con telemetría por flujo para enriquecer la interpretación de datos y facilitar el análisis de causa raíz (Ester et al., 1996; GeeksforGeeks, 2025; Liu et al., 2008; Observium, 2026; Schölkopf et al., 2001; Zenarmor, 2024).

### 1.1.2 Situación actual / brecha operativa

En entornos empresariales, la operación cotidiana del monitoreo suele apoyarse en plataformas que recolectan y presentan métricas de desempeño, lo cual permite observar tendencias, utilización y el estado general de interfaces y equipos (Observium, 2026). No obstante, aun con visibilidad histórica, la identificación de picos de tráfico puede depender de interpretaciones manuales y de alarmas basadas en umbrales

estáticos; este enfoque no siempre permite diferenciar con precisión entre fluctuaciones normales y eventos anómalos, especialmente cuando el comportamiento varía por estacionalidad o por cargas de trabajo legítimas (Liu et al., 2008; Schölkopf et al., 2001).

Adicionalmente, cuando el análisis se limita a métricas agregadas, se presenta una dificultad técnica para atribuir el incremento a un origen específico (por ejemplo, pares IP, puertos o protocolos), dado que dicha atribución requiere datos granulares por flujo. En ese sentido, los protocolos NetFlow e IPFIX aportan el contexto necesario al permitir examinar las contribuciones individuales y comprender el origen del consumo (Claise et al., 2013; Zenarmor, 2024). Por ello, la situación actual revela una brecha crítica entre la detección de un incremento volumétrico y la capacidad de explicar su origen operativo con evidencia rastreable.

Por tanto, la brecha central se define como la necesidad de integrar de manera sistemática la telemetría agregada (SNMP) con los registros por flujo (NetFlow/IPFIX), mediante el uso de algoritmos no supervisados capaces de modelar el comportamiento normal y detectar desviaciones sin depender de un etiquetado previo de datos (Ester et al., 1996; GeeksforGeeks, 2025b; Liu et al., 2008; Schölkopf et al., 2001). Esta integración debe derivar en resultados útiles para la operación: detección oportuna, reducción de falsos positivos y atribución del evento a entidades accionables, manteniendo el enfoque aplicado exigido para los trabajos de opción a grado (Universidad Internacional de Ciencia y Tecnología [UNICYT], 2018).

## 1.2 Formulación del Problema

¿En qué medida la integración de algoritmos de aprendizaje automático no supervisado con datos de telemetría SNMP y flujos *NetFlow/IPFIX* permite la detección precisa y la atribución de origen de los picos de tráfico en redes LAN empresariales?

## 1.3. Objetivos

### Objetivo General

Diseñar y validar conceptualmente un modelo de aprendizaje automático no supervisado para la detección de picos de tráfico en redes LAN empresariales, empleando datos de SNMP y *NetFlow/IPFIX*, y proponer lineamientos para la optimización operativa de la red.

### Objetivos Específicos

a) Analizar las fuentes de datos disponibles mediante SNMP y *NetFlow/IPFIX* y definir las métricas de interés para la caracterización del tráfico de red.

b) Diseñar un pipeline en Python para el preprocesamiento de datos, selección de características y aplicación de modelos de detección de anomalías como Isolation Forest, One-Class SVM y DBSCAN.

c) Validar de manera conceptual, con datos anonimizados o simulados, el desempeño de los modelos propuestos utilizando métricas de detección como la tasa de acierto y la tasa de falsos positivos.

d) Asociar los picos de tráfico detectados a las interfaces, dispositivos o flujos responsables mediante la correlación de datos de SNMP y *NetFlow*.

e) Elaborar un cronograma de trabajo conforme a las normas institucionales y proponer recomendaciones operativas basadas en los resultados conceptuales para la optimización de la red.

### 1.4. Justificación

Justificación práctica (operativa). La detección temprana de picos de tráfico se justifica, en primer lugar, por su impacto directo en la continuidad del servicio y en la operación diaria de TI: al identificar desviaciones relevantes con mayor anticipación, se pueden activar acciones correctivas antes de que la congestión degrade indicadores clave de calidad de servicio, como la latencia, la variación del retardo (*delay variation/jitter*) y la pérdida de paquetes, variables especialmente sensibles para tareas interactivas y servicios en tiempo real. Asimismo, en la práctica operativa, los esquemas basados en umbrales fijos exigen ajuste manual y tienden a volverse inexactos cuando cambian las cargas o patrones de uso, lo que puede aumentar la emisión de alertas de bajo valor y contribuir a la “fatiga por alertas” (desensibilización y menor respuesta ante alertas críticas).

Justificación tecnológica. El trabajo es tecnológicamente pertinente porque se apoya en telemetría estandarizada y ampliamente adoptada en entornos empresariales: SNMP para métricas de estado/rendimiento y exportación de flujos mediante IPFIX (y, por extensión, tecnologías de análisis de flujos usadas en la industria) para caracterizar comportamientos de tráfico. Este tipo de visibilidad es base para análisis de tráfico y planificación, por lo que integrar la detección automática de anomalías sobre estas fuentes fortalece la gestión operativa con mecanismos más adaptativos que los umbrales estáticos.

Justificación teórica. Desde lo teórico, el proyecto se sustenta en el campo de la detección de anomalías, donde se reconoce que “lo normal” puede ser difícil de delimitar y además evoluciona con el tiempo; por ello, los enfoques no supervisados resultan adecuados cuando el etiquetado es limitado o costoso, ya que modelan el comportamiento esperado y señalan desviaciones significativas. En particular, la literatura de referencia en anomalías y en anomalías de red respalda la pertinencia de evaluar técnicas no supervisadas para escenarios reales de operación, donde la disponibilidad de datos etiquetados suele ser una restricción.

Justificación académica. Académicamente, el trabajo aporta al área de gestión y operación de redes al articular un enfoque reproducible (ingesta, limpieza, normalización y extracción de características) y al fundamentar la selección/uso de algoritmos no supervisados para un problema de alto valor aplicado: la detección de eventos anómalos de tráfico. Esta contribución se alinea con líneas de investigación consolidadas en detección de anomalías y en anomalías de red, donde se enfatiza la necesidad de métodos y supuestos claros para evaluar la efectividad de las técnicas en dominios específicos.

Justificación social. Finalmente, existe una justificación social porque una red más estable y con menor degradación perceptible mejora la experiencia del usuario final (colaboradores/áreas de negocio) y favorece la continuidad de actividades críticas, especialmente aquellas sensibles a latencia y variación del retardo. A la vez, reducir el volumen de alertas irrelevantes disminuye la carga cognitiva del personal de operaciones y el riesgo de pasar por alto eventos realmente importantes, mitigando el fenómeno de fatiga por alertas descrito en la literatura y en prácticas de monitoreo.

## 1.5. Alcance y Delimitaciones

El presente trabajo se circunscribe al entorno de redes LAN empresariales y aborda la detección de picos de tráfico desde una perspectiva operativa mediante aprendizaje automático no supervisado. Para ello, emplea como insumos métricos recolectadas por SNMP (por ejemplo, contadores e indicadores de interfaces y dispositivos) y registros de flujo NetFlow/IPFIX, asumiendo que dichas fuentes son generadas y capturadas por herramientas de monitoreo existentes en la organización. El alcance incluye el diseño conceptual de la solución, el desarrollo de un pipeline reproducible en Python para ingesta, depuración, normalización y generación de características, y la evaluación técnica de modelos no supervisados (Isolation Forest, *One-Class SVM* y DBSCAN) orientada a identificar desviaciones relevantes respecto del comportamiento normal. La validación se realizará con conjuntos de datos previamente anonimizados o, en su defecto, con datos simulados que emulen condiciones realistas de operación y la aparición de eventos anómalos, produciendo como resultado marcas de anomalía y evidencias técnicas que respalden conclusiones y recomendaciones.

**Delimitaciones**

No se contempla una implementación en producción ni un despliegue en tiempo real; por tanto, quedan fuera actividades de operación continua, escalabilidad, alta disponibilidad, *hardening, MLOps* y monitoreo del modelo en ejecución. El estudio se limita a enfoques no supervisados, excluyendo métodos supervisados que requieran etiquetado de datos, así como la clasificación detallada de causas específicas de cada pico. Asimismo, el trabajo no incluye la integración con sistemas de respuesta automatizada (por ejemplo, orquestación o corrección automática), ni un análisis integral

de costos de infraestructura para una adopción a gran escala. Finalmente, se excluyen los aspectos legales y normativos asociados a retención, privacidad y cumplimiento, y el análisis se restringe a métricas y metadatos (*SNMP/NetFlow/IPFIX*), sin inspección de contenido, por lo que la interpretación de resultados dependerá de la disponibilidad, granularidad y calidad de los datos empleados.

De manera complementaria, la delimitación del estudio implica que la evaluación propuesta se orienta a demostrar la viabilidad del enfoque a partir de datos históricos o conjuntos anonimizados/simulados, por lo que no se medirán aspectos propios de la operación en línea como latencia de detección, tiempos de reacción ante eventos, desempeño bajo cargas sostenidas o variaciones abruptas de topología. Tampoco se desarrollarán mecanismos de correlación avanzada con fuentes externas (por ejemplo, logs de firewall, autenticación, DNS o aplicaciones) ni se realizará una validación exhaustiva con “verdad terrena” (ground truth) que permita atribuir con certeza causal cada pico a un evento específico; en consecuencia, los hallazgos se interpretarán como indicios analíticos y no como diagnósticos definitivos. Asimismo, el trabajo no pretende reemplazar plataformas comerciales de monitoreo ni comparar formalmente su desempeño con herramientas propietarias, sino delimitar un marco conceptual replicable que dependa principalmente de la telemetría disponible y de los supuestos de calidad del dato (frecuencia de muestreo, consistencia temporal y cobertura de exportación).

# CAPÍTULO II. MARCO TEÓRICO

El presente capítulo constituye el fundamento conceptual y técnico que sustenta la investigación, orientado a la comprensión de las dinámicas de supervisión en infraestructuras de datos modernas. Inicialmente, se examina la relevancia de las redes LAN empresariales y los desafíos críticos que supone la gestión de tráfico ante la proliferación de dispositivos interconectados. Posteriormente, se analizan los protocolos de monitoreo estándar, tales como SNMP y NetFlow/IPFIX, evaluando su capacidad para la recolección de métricas de rendimiento. Asimismo, se aborda la caracterización de los picos de tráfico como anomalías dentro de la red, estableciendo las bases para la aplicación de modelos de aprendizaje automático no supervisado, específicamente los algoritmos Isolation Forest, One-Class SVM y DBSCAN. Finalmente, se ofrece una revisión de trabajos relacionados que permite contextualizar la utilidad de estas herramientas para la detección proactiva de incidentes en entornos corporativos (Chandola et al., 2009; Claise, 2004; Claise et al., 2013; Ester et al., 1996; Harrington et al., 2002; Liu et al., 2008; Schölkopf et al., 2001).

## 2.1. Redes LAN Empresariales y Gestión de Tráfico

Las redes de área local (LAN) empresariales son infraestructuras de comunicación privadas que interconectan dispositivos como computadoras, servidores, impresoras y teléfonos IP dentro de una organización, facilitando el intercambio de recursos y la provisión de servicios de voz, datos y video (Stouffer et al., 2023). La gestión de tráfico en estas redes es una disciplina que se centra en monitorizar la utilización de los enlaces, priorizar servicios críticos mediante mecanismos de calidad de servicio (Quality of Service, QoS) y mantener la calidad de transmisión esperada (Blake et al., 1998; International Telecommunication Union, 2000). La complejidad creciente de estas redes,

impulsada por la proliferación de dispositivos del Internet de las Cosas (IoT) y el uso de esquemas Bring Your Own Device (BYOD), incrementa los retos de gestión y refuerza la necesidad de controles y herramientas que permitan identificar comportamientos anómalos, planificar capacidad y sostener niveles de servicio acordados (Boeckl et al., 2019; Chandola et al., 2009; Howell et al., 2023; Westerinen et al., 2001). La detección de picos de tráfico se inscribe dentro de las funciones de gestión proactiva, orientadas a anticipar degradaciones de rendimiento antes de que impacten a los usuarios (Chandola et al., 2009; International Telecommunication Union, 2000).

### 2.2. Monitoreo mediante SNMP y Observium

El Protocolo Simple de Administración de Red (Simple Network Management Protocol, SNMP) es un protocolo estándar de la capa de aplicación, ampliamente utilizado para supervisar y gestionar dispositivos en redes IP. Proporciona un marco estandarizado que permite a los administradores de red monitorizar el rendimiento de enrutadores, conmutadores, servidores y otros dispositivos, facilitando la gestión del rendimiento, la detección de fallos y la solución de problemas (GeeksforGeeks, 2025b; Harrington et al., 2002). Para la recolección de métricas, este proyecto considera el uso de una plataforma como Observium, un sistema de monitoreo y administración de redes que descubre automáticamente los dispositivos conectados y recopila métricas de rendimiento en tiempo real. Observium presenta estos datos a través de una interfaz web, mostrando gráficos históricos y en tiempo real, y puede generar alertas ante anomalías predefinidas. A través de SNMP, Observium permite obtener indicadores clave como el ancho de banda por interfaz (bytes de entrada/salida), utilización de CPU

y memoria, los cuales son fundamentales para detectar picos de tráfico (Observium, 2026).

### 2.3. Monitoreo mediante NetFlow/IPFIX

Mientras que SNMP informa sobre cuánto tráfico pasa por una interfaz, el monitoreo de flujos proporciona detalles sobre quién y qué está generando ese tráfico. NetFlow es una tecnología desarrollada por Cisco para recolectar información sobre el tráfico IP que atraviesa los dispositivos de red. Define un flujo como una secuencia unidireccional de paquetes que comparten atributos comunes, como las direcciones IP de origen y destino, los puertos, el protocolo de capa 4, la interfaz de entrada y el tipo de servicio (Type of Service, ToS). Los registros de flujo exportados incluyen conteos de paquetes y bytes, así como marcas temporales, lo que permite caracterizar la carga de la red y analizar el origen y destino del tráfico (Claise, 2004; Zenarmor, 2024). Por su parte, IP Flow Information Export (IPFIX) es un estándar de la Internet Engineering Task Force (IETF) que estandariza y amplía las capacidades de NetFlow v9. IPFIX utiliza un enfoque basado en plantillas, ofreciendo flexibilidad para definir los campos de exportación y permitir una gama más amplia de información para arquitecturas de monitoreo escalables y detalladas (Claise et al., 2013).

### 2.4. Picos de Tráfico y Detección de Anomalías

En el contexto de la gestión de redes, un pico de tráfico puede entenderse como un aumento abrupto y temporal en el volumen de datos que circulan por un enlace o interfaz, superando significativamente el comportamiento normal o esperado para un período de tiempo determinado. Dichos picos pueden ser el resultado de actividades legítimas (por ejemplo, copias de seguridad o descargas masivas) o de eventos no

deseados (por ejemplo, errores de configuración, bucles de red o intentos de denegación de servicio). La detección de estas anomalías implica diferenciar variaciones atípicas de las fluctuaciones normales del tráfico. En términos generales, los métodos de detección de anomalías buscan identificar patrones en series temporales u otros espacios de características que se desvían del comportamiento habitual y, según el enfoque, asignan medidas o puntuaciones para decidir si el comportamiento es anómalo (Barford et al., 2002; Chandola et al., 2009).

## 2.5. Aprendizaje Automático No Supervisado

El aprendizaje automático no supervisado es una rama de la inteligencia artificial que se enfoca en encontrar patrones o estructuras en conjuntos de datos que no han sido etiquetados previamente. En el contexto de la detección de picos de tráfico, se emplean algoritmos capaces de identificar valores atípicos (outliers) o regiones de baja densidad en el espacio de características, sin requerir un conjunto de entrenamiento con anomalías previamente etiquetadas (Chandola et al., 2009).

### *2.5.1. Isolation Forest*

El algoritmo Isolation Forest (Bosque de Aislamiento) aísla las observaciones de forma recursiva construyendo árboles de partición aleatorios. La premisa es que los puntos anómalos, al ser pocos y diferentes, son más susceptibles de ser aislados y, por lo tanto, requieren menos divisiones aleatorias para quedar aislados, lo que se traduce en longitudes de trayectoria más cortas en los árboles. Estas longitudes de trayectoria se utilizan para asignar puntuaciones de anomalía. Isolation Forest es computacionalmente eficiente, especialmente en conjuntos de datos de alta dimensión,

y no requiere supuestos sobre la distribución de los datos (GeeksforGeeks, 2025a; Liu et al., 2008).

### *2.5.2. One-Class SVM*

El método One-Class Support Vector Machine (SVM de una clase) es una adaptación de las máquinas de vectores de soporte diseñada para la detección de novedades o anomalías. En lugar de encontrar un hiperplano que separe dos clases de datos, One-Class SVM construye una frontera que encierra la mayoría de los puntos de datos considerados normales en un espacio de características de alta dimensión. Los puntos que se encuentran fuera de esta frontera se consideran anomalías, lo que resulta útil cuando solo se dispone de ejemplos de la clase normal y no de anomalías etiquetadas (GeeksforGeeks, 2025d; Schölkopf et al., 2001).

### *2.5.3. DBSCAN (Clustering Basado en Densidad)*

El algoritmo DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un método de agrupamiento que identifica regiones de alta densidad separadas por zonas de baja densidad. Permite descubrir grupos de formas arbitrarias sin necesidad de especificar previamente el número de clusters. Una de sus características más valiosas para este proyecto es su capacidad para clasificar los puntos que no pertenecen a ninguna región densa como ruido u outliers. Por ello, resulta un método adecuado para detectar anomalías en conjuntos de datos donde los valores atípicos aparecen como puntos aislados en el espacio de características (Ester et al., 1996).

## 2.6. Trabajos Relacionados

La literatura académica presenta estudios que abordan la detección de anomalías en tráfico de red mediante análisis de señales y enfoques de aprendizaje automático, incluyendo revisiones amplias de técnicas y criterios de evaluación (Barford et al., 2002; Chandola et al., 2009). En términos metodológicos, es común combinar fuentes de datos complementarias: métricas agregadas obtenidas mediante SNMP y detalles granulares del tráfico mediante registros de flujo (NetFlow/IPFIX), con el fin de mejorar la detección y facilitar el diagnóstico (Claise, 2004; Claise et al., 2013; Harrington et al., 2002). En el caso de enfoques no supervisados, modelos como Isolation Forest, One-Class SVM y DBSCAN se emplean como bases conceptuales para identificar comportamientos atípicos sin requerir etiquetas previas, según sus formulaciones originales (Ester et al., 1996; Liu et al., 2008; Schölkopf et al., 2001).

# CAPÍTULO III. METODOLOGÍA

El presente capítulo detalla la ruta metodológica diseñada para dar cumplimiento a los objetivos de la investigación, fundamentada en un enfoque cuantitativo de tipo aplicado. En la primera sección, se define la naturaleza del proyecto de graduación y su orientación hacia la resolución de problemas prácticos en infraestructuras de red. Posteriormente, se describen las fuentes de datos primarias, centradas en la telemetría SNMP y registros NetFlow/IPFIX, junto con el ecosistema de herramientas de *Python* seleccionadas para el procesamiento analítico. Asimismo, se presenta el diseño del *pipeline* de datos, el cual abarca desde la ingestión y preprocesamiento hasta la aplicación de algoritmos de aprendizaje automático no supervisado. Finalmente, se establecen los criterios de evaluación para medir la precisión del modelo y las consideraciones éticas indispensables para garantizar la privacidad y anonimización de la información capturada en el entorno empresarial.

## 3.1. Tipo y Enfoque de Estudio

Este proyecto se desarrolla bajo la modalidad de proyecto de graduación para la opción de titulación de maestría profesional. Se clasifica como una investigación de tipo aplicado, ya que su finalidad es la resolución de un problema práctico mediante el diseño y validación conceptual de una solución tecnológica. El enfoque del estudio es predominantemente cuantitativo, dado que se basa en la recolección y análisis de datos numéricos (métricas de red) para probar hipótesis sobre el comportamiento del tráfico. Siguiendo las directrices institucionales para el plan de trabajo del anteproyecto de grado (Universidad Internacional de Ciencia y Tecnología [UNICyT], 2018), el documento se estructura para incluir un título, planteamiento del problema, objetivos, justificación,

marco teórico, metodología y un cronograma de trabajo, sin la obligación de una implementación productiva.

### 3.2. Fuentes de Datos y Herramientas

Se utilizarán dos fuentes principales de datos para alimentar el modelo:

SNMP a través de Observium: Se recopilarán series temporales de métricas de rendimiento a nivel de interfaz, como el contador de bytes de entrada y salida (ifInOctets, ifOutOctets), contadores de paquetes y errores (ver Anexo B). Se asume que una plataforma como está configurada para sondear los dispositivos de red a intervalos regulares (p. ej., cada 5 minutos) (Observium Limited, 2026), cuya configuración representativa se documenta en el Anexo C.

Registros de flujos NetFlow/IPFIX: Se utilizará un colector de flujos offline, como la combinación de herramientas `nfcapd` y `nfdump`, para almacenar los registros de flujo exportados por los enrutadores o conmutadores de borde (ver Anexo C). Cada registro contiene información detallada como direcciones IP, puertos, protocolo, conteo de paquetes y bytes, y marcas de tiempo, permitiendo identificar los hosts y aplicaciones responsables del tráfico (Zenarmor, 2024).

El procesamiento, análisis y modelado de los datos se llevará a cabo en el lenguaje de programación, utilizando bibliotecas estándar del ecosistema de ciencia de datos como Pandas para la manipulación y limpieza de datos, Scikit-learn para la implementación de los algoritmos de aprendizaje automático, y Matplotlib/Seaborn para la visualización de resultados

### 3.3. Diseño del Pipeline Analítico

El pipeline propuesto para el procesamiento y análisis de los datos consta de las siguientes etapas secuenciales:

a) Ingestión de datos: Obtención programática de las métricas de series temporales de SNMP y los archivos de registros de NetFlow/IPFIX desde sus respectivas fuentes de almacenamiento.

b) Preprocesamiento y limpieza: Tareas que incluyen la normalización de los datos, el manejo de valores faltantes (p. ej., mediante interpolación), la conversión de contadores a tasas (p. ej., bytes a bits por segundo) y la sincronización temporal de las dos fuentes de datos.

c) Extracción y selección de características (Feature Engineering): Cálculo de variables derivadas que puedan capturar mejor la dinámica del tráfico, como la utilización porcentual del enlace, la variabilidad temporal (desviación estándar), las tasas de crecimiento del tráfico y atributos agregados de los flujos (p. ej., número de flujos únicos por host).

d) Aplicación de modelos: Entrenamiento y ajuste de los algoritmos no supervisados seleccionados (*Isolation Forest, One-Class SVM, DBSCAN*) sobre un conjunto de datos de entrenamiento considerado "normal" para que aprendan la línea base del comportamiento del tráfico.

e) Detección y postproceso: Aplicación de los modelos entrenados sobre los datos completos para generar puntuaciones de anomalía. Se establecerá un umbral (estático o adaptativo) para clasificar los puntos como picos.

f) Correlación y reporte: Una vez detectado un pico en una interfaz (datos SNMP), se cruzarán las marcas de tiempo con los registros NetFlow/IPFIX para identificar los flujos, direcciones IP o aplicaciones responsables. Los resultados se presentarán en formatos comprensibles, como tablas y visualizaciones.

### 3.4. Criterios de Evaluación de Anomalías

Para evaluar la efectividad del modelo propuesto de manera conceptual, se considerarán las siguientes métricas, que se medirán en un entorno de validación con datos simulados o etiquetados:

- Tasa de Detección (Recall): Proporción de picos de tráfico reales (previamente identificados o inyectados) que son correctamente identificados por el modelo.
- Tasa de Falsos Positivos (False Positive Rate): Porcentaje de alertas generadas por el modelo que no corresponden a picos de tráfico reales, sino a fluctuaciones normales.
- Capacidad de Asociación: Precisión con la que el modelo es capaz de relacionar una anomalía detectada con la interfaz, dispositivo o flujo de red correcto que la originó.
- Tiempo de Procesamiento: Evaluación del desempeño computacional del pipeline para determinar su viabilidad para operar en un escenario de tiempo casi real.

### 3.5. Consideraciones Éticas y de Privacidad

La recopilación y el análisis de datos de tráfico de red pueden implicar el manejo de información sensible sobre los usuarios y las aplicaciones que utilizan. Por tanto, el cumplimiento de las normativas de privacidad y la protección de datos son fundamentales. Para este proyecto, se establece que todos los datos utilizados serán

anonimizados antes de su análisis. Específicamente, las direcciones IP y los números de puerto en los registros NetFlow/IPFIX deben ser eliminados o codificados (p. ej., mediante hashing con sal) para evitar la identificación de individuos. El acceso a los datos brutos y procesados debe estar restringido únicamente al personal autorizado para la investigación. Además, la interpretación de los resultados y la generación de alertas deben realizarse con prudencia para no generar señalamientos injustificados que puedan comprometer la confidencialidad o la privacidad de los usuarios de la red. Un ejemplo de este criterio de anonimización aplicado a la configuración de los instrumentos de recolección se presenta en el Anexo C.

### 3.6. Plan de Trabajo (Cronograma)

El cronograma se elabora en conformidad con las normas institucionales (UNICyT, 2018), presentando un diagrama sencillo que detalla las actividades y su duración a lo largo de un periodo de cinco meses. Este plan es tentativo y puede ajustarse según las necesidades del proyecto.

**Tabla 1**

*Cronograma de trabajo propuesto*

| Actividad | Mes 1 | Mes 2 | Mes 3 | Mes 4 | Mes 5 |
|---|---|---|---|---|---|
| Recolección de datos (configuración de Observium y colector NetFlow/IPFIX) | X | X | | | |
| Preprocesamiento y selección de características (limpieza, sincronización y cálculo de métricas) | | X | X | | |
| Diseño del modelo y ajuste de algoritmos (Isolation Forest, One-Class SVM, DBSCAN) | | | X | X | |

| | | | | | |
|---|---|---|---|---|---|
| Validación con datos anonimizados/simulados (definición de escenarios y cálculo de métricas) | | | X | X | |
| Análisis de resultados y elaboración de recomendaciones | | | | X | X |
| Redacción del informe final y revisión | | | | X | X |
| Presentación y sustentación del proyecto | | | | | X |

*Nota.* Cronograma tentativo de cinco meses para el desarrollo del proyecto de graduación. Las marcas (X) indican el período de ejecución para cada actividad.

# CAPÍTULO IV. RESULTADOS Y ANÁLISIS

El presente capítulo despliega la propuesta técnica central de la investigación, orientada al diseño y validación conceptual de un sistema inteligente para la detección de anomalías de tráfico en redes LAN empresariales. En la sección 4.1, se detalla la arquitectura de la solución, la cual integra mecanismos de captura híbrida (SNMP y NetFlow/IPFIX) con un núcleo de procesamiento basado en algoritmos de aprendizaje automático no supervisado, específicamente *Isolation Forest*, *One-Class SVM* y *DBSCAN*. Seguidamente, el apartado 4.2 establece la estrategia de validación, definiendo los escenarios experimentales y las métricas de desempeño tasa de detección y reducción de falsos positivos necesarias para evaluar la eficacia de los modelos. De igual forma, en la sección 4.3 se estructura el plan de trabajo cronológico conforme a los lineamientos institucionales, mientras que en la sección 4.4 se realiza una interpretación conceptual de los resultados esperados, demostrando la capacidad del sistema para asociar picos de tráfico con sus causas raíz. Finalmente, la sección 4.5 examina las limitaciones del alcance actual y plantea líneas de mejora futura para la evolución del modelo hacia entornos productivos de alta escala.

## 4.1. Diseño del Modelo Propuesto

### *4.1.1. Arquitectura de la Solución*

La solución se concibe con una arquitectura de tres capas principales: captura de datos, procesamiento y análisis, y presentación de resultados. Esta estructura modular permite una mayor flexibilidad y escalabilidad.

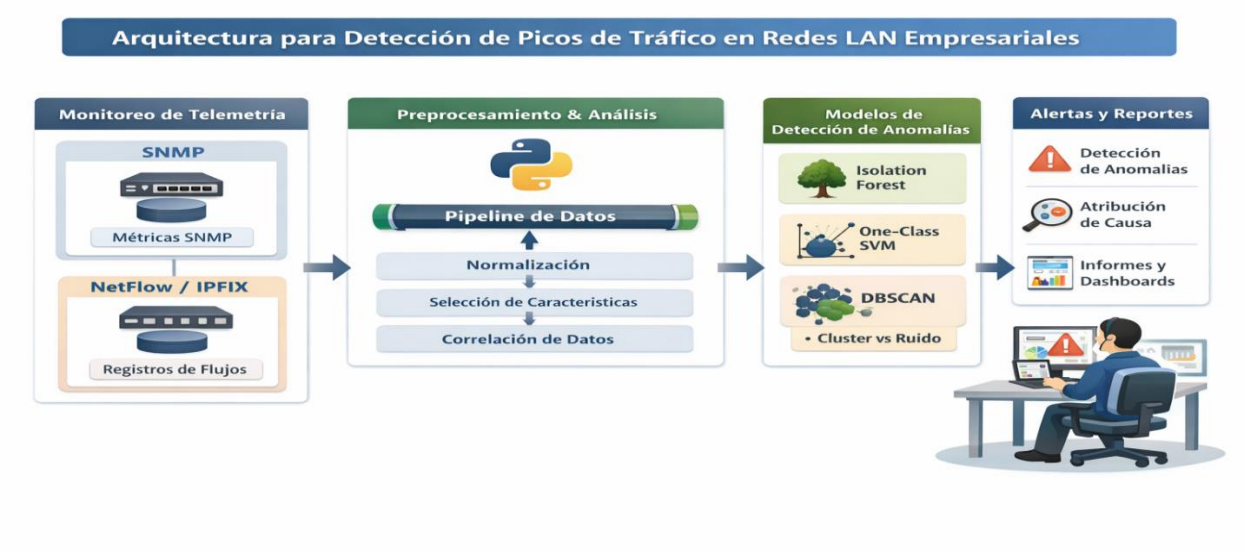
1. **Capa de Captura:** Compuesta por las herramientas de monitoreo. Por un lado, Observium recolecta métricas de rendimiento de los dispositivos de red vía SNMP. Por otro, un exportador *NetFlow/IPFIX* en los equipos de borde envía los registros

de flujo a un colector centralizado. Estos datos se almacenan en bases de datos temporales o archivos para su posterior procesamiento.

2. **Capa de Procesamiento:** El núcleo de la solución. Un pipeline desarrollado en Python ingiere los datos de ambas fuentes, los sincroniza temporalmente, los preprocesa y aplica los modelos de aprendizaje no supervisado. Cada algoritmo genera una puntuación de anomalía por cada intervalo de tiempo y por cada entidad monitoreada (interfaz o flujo).
3. **Capa de Presentación:** Los resultados del análisis se integran en la plataforma de monitoreo existente o en un cuadro de mando (dashboard) específico. Se generan alertas automáticas y se proporcionan tablas detalladas que indican el componente responsable, la gravedad del pico y recomendaciones operativas iniciales.

**Figura 1**

*Arquitectura general de la solución propuesta.*




*Nota.* Representación conceptual de la arquitectura de tres capas propuesta para la detección de picos de tráfico en redes LAN empresariales. Elaboración propia.

### *4.1.2. Captura y Preparación de Datos*

Los datos de SNMP se extraerán a intervalos fijos de cinco minutos para cada interfaz de red monitoreada. Se almacenarán variables clave como `ifInOctets` (bytes recibidos), `ifOutOctets` (bytes transmitidos), `ifInErrors`, `ifOutErrors` y métricas de uso de CPU del dispositivo (ver Anexo B). Simultáneamente, se colectarán los flujos NetFlow/IPFIX en intervalos similares (ver Anexo C). Durante el preprocesamiento, se calculará la utilización porcentual de cada enlace (bits por segundo / capacidad del enlace), se alinearán las marcas temporales de ambas fuentes de datos y se aplicarán técnicas de interpolación para manejar posibles valores faltantes.

### *4.1.3. Selección y Transformación de Características*

Para la detección efectiva de picos, se seleccionarán y crearán características (features) que reflejen la dinámica del tráfico de red (el detalle completo de las variables y sus transformaciones se presenta en el Anexo B):

- Tasa de Utilización: Bytes transferidos por segundo, normalizados por la capacidad del enlace.
- Derivada Temporal: Diferencia entre la tasa de utilización actual y el promedio de una ventana histórica (p. ej., los últimos 30 minutos), para capturar cambios abruptos.
- Rango Intercuartílico (IQR): Medida de dispersión estadística para detectar variaciones anómalas en la distribución del tráfico.
- Atributos de Flujo: Número de flujos activos, distribución de puertos de destino, duración media de los flujos y proporción de bytes por dirección IP de origen.

Estas características se normalizarán (p. ej., usando escalado Min-Max) para que tengan un rango comparable y evitar sesgos en los algoritmos.

#### *4.1.4. Modelos de Aprendizaje Automático*

Se evaluarán tres algoritmos no supervisados, cada uno con un enfoque diferente para la detección de anomalías:

- Isolation Forest: Funciona bien para ubicar picos que se apartan con claridad del comportamiento habitual, porque esos puntos “se aíslan” con pocas divisiones dentro de árboles aleatorios (Liu et al., 2008).
- One-Class SVM: Define un límite que encierra el patrón considerado normal. Cuando aparece un pico de tráfico, al comportarse como un caso atípico, tiende a quedar fuera de ese límite y se marca como anomalía (Schölkopf et al., 2001).
- DBSCAN: Agrupará los puntos de tráfico en clústeres que representen el comportamiento “normal” a partir de la densidad. Cuando aparezcan picos poco frecuentes o aislados, el algoritmo tenderá a marcarlos como “ruido” (outliers) (Ester et al., 1996).

Los modelos se entrenarán con datos de periodos considerados normales (por ejemplo, fuera del horario laboral o en días de baja actividad) para construir una línea base del comportamiento de la red.

### 4.2. Estrategia de Validación

#### *4.2.1. Diseño de Experimentos*

La validación del modelo se realizará utilizando datos anonimizados provenientes de una red real o, en su defecto, datos simulados que reproduzcan patrones de tráfico normales y picos generados artificialmente (los prompts empleados para facilitar la

generación de estos detalles técnicos se documentan en el Anexo D). Se seleccionarán periodos representativos de tráfico normal para entrenar los modelos. Posteriormente, se inyectarán picos de tráfico sintéticos de distintas magnitudes y duraciones en diferentes momentos para evaluar la capacidad de detección de los algoritmos.

### *4.2.2. Escenarios de Prueba y Métricas*

Se definirán varios escenarios de prueba para una evaluación exhaustiva:

- Tráfico Normal: Periodos sin eventos relevantes para evaluar la tasa de falsos positivos de cada modelo.
- Eventos Programados: Simulación de copias de seguridad o actualizaciones masivas para comprobar que el modelo los identifica y asocia correctamente.
- Eventos Inesperados: Inyecciones de tráfico que simulen transferencias de archivos masivas no planificadas o ataques de denegación de servicio (DoS) controlados para medir la sensibilidad del modelo.

Las métricas de desempeño (tasa de detección, falsos positivos y capacidad de asociación) se calcularán para cada modelo en cada escenario. También se medirá el tiempo de procesamiento para evaluar la viabilidad de una operación en tiempo real.

### *4.2.3. Comparación de Modelos*

Se comparará el desempeño de Isolation Forest, One-Class SVM y DBSCAN utilizando las métricas definidas. Se analizará la sensibilidad de cada algoritmo a sus hiperparámetros (p. ej., número de árboles, ancho del kernel, épsilon y minPts) y se seleccionará el modelo que ofrezca el mejor equilibrio entre precisión y eficiencia computacional. Se estudiará también la posibilidad de combinar las salidas de los

modelos mediante un enfoque de ensamblado (ensemble) o votación para mejorar la robustez general de la detección.

### 4.3. Análisis Conceptual de Resultados

Una vez aplicados los modelos a los datos de prueba, se espera obtener como resultado principal una lista de intervalos de tiempo en los que se detectan picos, junto con la puntuación de anomalía asignada por el modelo. Las visualizaciones incluirán gráficas de utilización de enlaces con marcas que indiquen los picos detectados (ver Anexo A) y tablas que muestren los flujos responsables, como se ejemplifica a continuación.

La interpretación conceptual de los resultados se sustenta en dos niveles complementarios: (a) la lectura del evento detectado y su trazabilidad (Tabla 2) y (b) el rol técnico del algoritmo empleado para marcar la anomalía (Tabla 3). En la Tabla 3 se resume que Isolation Forest y One-Class SVM operan como detectores de anomalías mediante puntuación/decisión, mientras que DBSCAN identifica agrupaciones por densidad y etiqueta como “ruido” los puntos que quedan fuera de regiones densas, lo cual es útil para análisis exploratorio. En términos operativos, la Tabla 2 ilustra cómo un evento marcado como anómalo puede asociarse a una causa probable al considerar contexto, origen/destino y servicio involucrado. Por ejemplo, un pico detectado a las 02:05 (Tabla 2), en un enlace principal, con origen en un servidor de respaldo y puerto destino 445, es consistente con el servicio registrado como microsoft-ds en el registro de puertos de IANA. En contraste, un pico a mediodía desde una estación de usuario hacia una IP externa (Tabla 2) podría corresponder a una descarga masiva no planificada o a actividad inusual que requiere verificación operativa. Bajo este enfoque, las métricas

SNMP aportan el “dónde y cuándo” (interfaz/enlace y ventana temporal), mientras que los registros de flujo (NetFlow/IPFIX) aportan el “quién y qué” (origen/destino, puertos, protocolo y características del flujo), mejorando la precisión diagnóstica y la toma de decisiones operativas.

**Tabla 2**

*Ejemplo de tabla de picos de tráfico detectados*

| Timestamp | Interfaz Afectada | Puntuación Anomalía | IP Origen Principal | IP Destino Principal | Puerto Destino | Protocolo |
|---|---|---|---|---|---|---|
| 2026-01-31 02:05:00 | Gi1/0/1 (Uplink) | 0.85 | 192.168.1.10 (Servidor Backup) | 10.10.10.50 (Storage) | 445 | TCP |
| 2026-01-31 11:30:00 | Gi1/0/24 (Usuario) | 0.79 | 192.168.10.123 | 203.0.113.10 (Externo) | 443 | TCP |

*Nota.* Los registros son simulados con fines ilustrativos. La puntuación de anomalía (0.0–1.0) corresponde a la salida del modelo Isolation Forest; valores ≥ 0.75 se clasifican como picos de tráfico relevantes.

**Tabla 3**

*Comparación de algoritmos para detección de picos de tráfico (enfoque no supervisado)*

| Algoritmo | Tipo de enfoque / salida | Idea base (qué modela) | Fortalezas para picos de tráfico | Limitaciones típicas | Parámetros clave | Uso sugerido en el pipeline |
|---|---|---|---|---|---|---|
| Isolation Forest | Anomalías (score + etiqueta) | Aísla observaciones con particiones aleatorias; los outliers suelen aislarse con trayectorias más cortas. | Escala bien en datos tabulares; detecta outliers puntuales si las variables capturan el cambio. | Puede elevar falsos positivos si hay estacionalidad o cambios de régimen sin modelar; requiere buen preprocesamiento. | n_estimators; max_samples; contamination. | Modelo base para detección temprana por ventanas (por interfaz/VLAN) y priorización de alertas. |

| | | | | | | |
|---|---|---|---|---|---|---|
| One-Class SVM | Anomalías (frontera de normalidad) | Estima la región de “normalidad” (soporte); con kernels puede capturar fronteras no lineales. | Útil para validar alertas cuando lo normal es complejo; buen “segundo verificador”. | Menos escalable en volúmenes grandes; sensible a escalado y ajuste de hiperparámetros. | kernel (p. ej., RBF); nu; gamma. | Modelo de confirmación para reducir falsos positivos en enlaces o VLAN críticas. |
| DBSCAN | Clustering + “ruido” (outliers) | Agrupa por densidad; puntos en baja densidad se marcan como ruido. | Separa modos/patrones de tráfico y marca ventanas atípicas como ruido; útil para análisis exploratorio. | Sensible a eps/min_samples; difícil con densidades variables y alta dimensionalidad sin reducción/escala. | eps; min_samples; metric. | Capa exploratoria: segmentar comportamientos y apoyar diagnóstico/explicación de picos. |

*Nota.* Resumen conceptual basado en la literatura técnica. Los parámetros clave requieren ajuste según las características específicas del entorno de red a monitorear.

La interpretación conceptual de estos resultados permite identificar patrones claros. Por ejemplo, un pico detectado a las 02:05 (Tabla 2) en un enlace principal, asociado a un servidor de backup y un puerto SMB (445), correspondería con alta probabilidad a una copia de seguridad programada. En cambio, un pico a mediodía desde una estación de usuario hacia una IP externa podría indicar una descarga masiva no planificada o una actividad inusual que requiere investigación. La representación gráfica de estos eventos se ilustra en el Anexo A. La combinación de métricas SNMP (dónde y cuándo) y registros NetFlow/IPFIX (quién y qué) ofrece una visión detallada que facilita un diagnóstico preciso y una respuesta operativa adecuada.

### 4.4. Limitaciones y Mejoras Futuras

Las principales limitaciones del modelo propuesto son:

- Dependencia de la Calidad de los Datos: La precisión del modelo es altamente dependiente de la consistencia y exactitud de los datos recolectados. Valores faltantes o mediciones imprecisas pueden afectar negativamente la detección.
- Sensibilidad a la Selección de Parámetros: Los algoritmos no supervisados requieren una calibración adecuada de sus hiperparámetros para evitar un exceso de falsos positivos o negativos. Este ajuste puede ser complejo y requerir conocimiento experto.
- Falta de Validación en Entornos de Producción: Al ser una validación conceptual, la efectividad real del modelo debe ser comprobada en redes operativas con cargas de tráfico variables y eventos impredecibles.

Como mejoras futuras, se sugiere la incorporación de algoritmos de aprendizaje supervisado una vez que se disponga de un conjunto de datos etiquetados, la integración con mecanismos de respuesta automática (p. ej., a través de APIs de controladores SDN) y la validación de la solución en redes de mayor escala o con topologías más complejas, como centros de datos o redes WAN.

# CONCLUSIONES

Este proyecto, desde un enfoque conceptual, muestra que detectar picos de tráfico con técnicas de aprendizaje automático no supervisado es una ruta técnicamente viable para redes LAN empresariales. A medida que crecen el consumo de ancho de banda, la mezcla de aplicaciones y la cantidad de dispositivos conectados, detectar a tiempo comportamientos fuera de lo normal deja de ser “un extra” y pasa a ser parte del trabajo diario de operar bien una red. Con la propuesta planteada, esos picos que antes se veían como eventos sueltos pueden convertirse en información útil, entendible y accionable.

Uno de los aportes más claros del enfoque está en combinar dos fuentes que se complementan de forma natural: métricas SNMP y registros de flujo NetFlow/IPFIX. Con SNMP se obtiene una visión agregada del rendimiento de interfaces y equipos, lo que ayuda a ubicar con precisión dónde y cuándo aparece un aumento de consumo. Los flujos, en cambio, bajan al detalle y permiten responder quién y qué está generando ese tráfico: direcciones IP, puertos, protocolos y rasgos del patrón de comunicación. Juntas, estas dos piezas dan un panorama más completo que mirar solo contadores de interfaz o revisar flujos de manera puntual.

En la parte de diseño, la arquitectura de tres capas (captura, procesamiento y presentación) ayuda a que la solución se entienda y crezca sin enredarse. Separar la recolección de datos del procesamiento en un pipeline en Python, y dejar la visualización/alertamiento en una capa final, facilita evolucionar el modelo hacia un entorno real. También deja margen para cambiar herramientas con el tiempo algo común en empresas sin tener que rehacer todo desde cero.

Los objetivos específicos se cubren de acuerdo con el alcance definido. Se construyó un modelo conceptual completo, se eligieron variables relevantes y se establecieron transformaciones necesarias. Además, se estructuró un pipeline de preprocesamiento que incluye sincronización temporal, cálculo de utilización porcentual y generación de características derivadas. Sobre esa base, se evaluaron tres algoritmos no supervisados con enfoques distintos: Isolation Forest, One-Class SVM y DBSCAN. Cada uno mira el problema desde un ángulo diferente (aislar valores atípicos, trazar una frontera de “normalidad”, o detectar puntos en zonas de baja densidad), lo que amplía el análisis y deja abierta la posibilidad de combinarlos en versiones futuras.

Visto desde la operación diaria, el modelo sugiere que detectar picos no debería quedarse en una alerta de “sí/no”. Tiene más valor cuando sirve como apoyo al diagnóstico. Al incluir puntuaciones de anomalía y tablas de trazabilidad que conectan cada evento con dispositivos, puertos y servicios, se puede entender el pico dentro del contexto general de la red. Eso ayuda a distinguir entre incrementos esperables como respaldos programados o una actualización masiva de equipos y situaciones que ameritan atención, como transferencias no autorizadas o patrones extraños que podrían relacionarse con incidentes de seguridad.

Con este enfoque también se fortalece la gestión en tres frentes prácticos: calidad de servicio, planificación de capacidad y seguridad operativa. En calidad de servicio, detectar congestión temprano permite reaccionar antes de que el usuario final sienta la caída (por ejemplo, llamadas VoIP con cortes o lentitud al acceder a sistemas internos). Para capacidad, mirar patrones y repeticiones en los picos da evidencia cuantitativa para justificar mejoras: ampliar enlaces, redistribuir cargas o aplicar priorización. Y en

seguridad, la detección de comportamientos atípicos aporta una capa extra de visibilidad que complementa soluciones tradicionales como IDS/IPS, sin pretender reemplazarlas.

Ahora bien, conviene dejar claro el marco en el que se sostienen los resultados: se trata de una validación conceptual y experimental basada en datos anonimizados o simulados. Eso permite probar la lógica del modelo y su coherencia técnica, pero llevarlo a producción requeriría más pruebas con tráfico real, cargas variables, aplicaciones diversas y eventos impredecibles. Además, la calibración de hiperparámetros puede influir bastante en el desempeño, así que ese ajuste debería trabajarse con cuidado en fases posteriores.

El cambio que se plantea pasar de una gestión reactiva con umbrales fijos a una gestión más proactiva basada en análisis de datos encaja con la evolución normal de la administración de TI. Los umbrales estáticos sirven como primera señal, pero se quedan cortos cuando el tráfico tiene estacionalidad, picos cortos o variaciones que no necesariamente superan un número “mágico”. En cambio, los modelos no supervisados pueden aprender un patrón de referencia y señalar desviaciones relevantes, aunque no crucen un umbral absoluto.

En conjunto, el trabajo refuerza la idea de que el aprendizaje automático puede integrarse de forma práctica en la operación de una red LAN sin exigir infraestructuras complejas ni depender de grandes volúmenes de datos etiquetados. Al combinar técnicas estadísticas, modelos no supervisados y herramientas de monitoreo ya disponibles, se puede construir un ecosistema analítico que eleve la madurez operativa. Así, la red no solo transporta datos: también se vuelve una fuente de señales útiles para mejorar de forma continua.

Como línea de evolución, el siguiente paso natural sería incorporar técnicas supervisadas cuando exista un histórico etiquetado de eventos reales, y conectar la detección con respuestas automatizadas para cerrar el ciclo entre “identificar” y “actuar”. También sería valioso validar el enfoque en topologías más grandes y variadas, incluyendo segmentación avanzada, virtualización de red e integración con escenarios WAN, para consolidar su robustez y su aplicabilidad en entornos empresariales más exigentes.

## RECOMENDACIONES

- Implementar captura continua: Hay que montar un esquema sólido que capture de forma constante y guarde el histórico de datos de SNMP y NetFlow/IPFIX. Esto implica definir cuánto tiempo se conservarán los registros y, desde el inicio, aplicar anonimización en la capa de recolección para proteger la información de los usuarios.
- Integrar el modelo al monitoreo: Para que el modelo realmente sirva, sus resultados tienen que estar a la vista del equipo. Lo más práctico es llevar sus salidas (alertas y reportes) a la herramienta de monitoreo que ya se usa, como Observium o una equivalente, de modo que el operador vea el evento en gráficos, paneles o alertas claras, sin tener que ir a buscarlo a otra parte.
- Definir procedimientos operativos (SOPs): Una alerta sin respuesta termina siendo ruido. Conviene dejar por escrito qué hacer según el tipo de pico detectado: si corresponde mover una ventana de respaldos, repartir carga entre enlaces, o activar ajustes de Calidad de Servicio (QoS) para priorizar aplicaciones críticas cuando el tráfico se dispara.
- Revisar y ajustar el modelo de forma regular: La red no se queda igual. Cambian las aplicaciones, los horarios de uso y hasta los hábitos de los usuarios. Por eso, es recomendable revisar periódicamente cómo está respondiendo el modelo y ajustar parámetros (o reentrenarlo, si aplica) para que no pierda precisión con el tiempo.

### Para Trabajos Futuros

- Probar en producción: El paso más directo es llevar el modelo a un entorno real y medir cómo se comporta con tráfico en vivo, donde hay picos inesperados, cambios

de carga y situaciones que en laboratorio no siempre aparecen. Ahí también se podrá comparar con más criterio el desempeño de los algoritmos evaluados.

- Explorar aprendizaje profundo: Si se logra acumular un histórico amplio, vale la pena evaluar modelos de Deep Learning orientados a series temporales, como RNN o LSTM, que pueden detectar patrones más sutiles y relaciones complejas que a veces se les escapan a métodos más clásicos.
- Integración con SDN: También sería interesante conectar la detección con una arquitectura de redes definidas por software (SDN) para cerrar el ciclo completo: no solo detectar el pico, sino responder automáticamente. Por ejemplo, que el controlador aplique reglas de re-enrutamiento o ajuste QoS en tiempo real cuando se detecte congestión.
- Ampliar el alcance: Finalmente, se puede extender el estudio a otros escenarios, como redes WAN o centros de datos a gran escala, y analizar cómo la topología y los patrones de tráfico propios de esos entornos cambian la forma en que se detectan los picos.

## REFERENCIAS


*Barford, P., Kline, J., Plonka, D., & Ron, A. (2002). A signal analysis of network traffic anomalies. In Proceedings of the 2nd ACM SIGCOMM Internet Measurement Workshop (IMW 2002) (pp. 71–82). Association for Computing Machinery. https://doi.org/10.1145/637201.637210*

*Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., & Weiss, W. (1998, diciembre). An architecture for differentiated services (RFC 2475). RFC Editor. https://doi.org/10.17487/RFC2475*

*Boeckl, K., Fagan, M., Fisher, W., Lefkovitz, N., Megas, K. N., Nadeau, E., O'Rourke, D. G., Piccarreta, B., & Scarfone, K. (2019, junio). Considerations for managing Internet of Things (IoT) cybersecurity and privacy risks (NISTIR 8228). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8228*

*Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), Article 15. https://doi.org/10.1145/1541880.1541882*

*Claise, B. (2004, octubre). Cisco Systems NetFlow services export version 9 (RFC 3954). RFC Editor. https://doi.org/10.17487/RFC3954*

*Claise, B., Trammell, B., & Aitken, P. (2013, septiembre). Specification of the IP Flow Information Export (IPFIX) protocol for the exchange of flow information (RFC 7011). RFC Editor. https://doi.org/10.17487/RFC7011*

*Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (pp. 226–231). AAAI Press.*

*GeeksforGeeks. (2025a, 23 de julio). Anomaly detection using Isolation Forest.* *https://www.geeksforgeeks.org/machine-learning/anomaly-detection-using-isolation-forest/*

*GeeksforGeeks. (2025b, 3 de octubre). Simple network management protocol (SNMP).* *https://www.geeksforgeeks.org/computer-networks/simple-network-management-protocol-snmp/*

*GeeksforGeeks. (2025c, 15 de julio). SNMP full form.* *https://www.geeksforgeeks.org/computer-networks/snmp-full-form/*

*GeeksforGeeks. (2025d, 23 de julio). Support Vector Machine (SVM) for anomaly detection.* *https://www.geeksforgeeks.org/machine-learning/support-vector-machine-svm-for-anomaly-detection/*

*Harrington, D., Presuhn, R., & Wijnen, B. (2002, diciembre). An architecture for describing Simple Network Management Protocol (SNMP) management frameworks (RFC 3411). RFC Editor.* *https://doi.org/10.17487/RFC3411*

*Howell, G., Franklin, J. M., Sritapan, V., Souppaya, M. P., & Scarfone, K. A. (2023, mayo). Guidelines for managing the security of mobile devices in the enterprise (NIST Special Publication 800-124 Rev. 2). National Institute of Standards and Technology.* *https://doi.org/10.6028/NIST.SP.800-124r2*

*Internet Assigned Numbers Authority. (2026, 13 de febrero). Service name and transport protocol port registry. Recuperado el 16 de febrero de 2026, de* *https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml?search=445*

*International Telecommunication Union. (2000, febrero). TMN management functions (ITU-T Recommendation M.3400).* https://www.itu.int/rec/T-REC-M.3400-200002-I/

*Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. In 2008 Eighth IEEE International Conference on Data Mining (pp. 413–422). IEEE.* https://doi.org/10.1109/ICDM.2008.17

*Observium Limited. (2026, 16 de enero). Observium - network monitoring platform (release 26.01). Recuperado el 31 de enero de 2026, de* https://www.observium.org/

*Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural Computation, 13(7), 1443–1471.* https://doi.org/10.1162/089976601750264965

*scikit-learn. (s. f.-a). DBSCAN. Recuperado el 16 de febrero de 2026, de* https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html

*scikit-learn. (s. f.-b). IsolationForest. Recuperado el 16 de febrero de 2026, de* https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html

*Stouffer, K., Pease, M., Tang, C. Y., Zimmerman, T., Pillitteri, V., Lightman, S., Hahn, A., Saravia, S., Sherule, A., & Thompson, M. (2023, septiembre). Guide to operational technology (OT) security (NIST Special Publication 800-82 Rev. 3). National Institute of Standards and Technology.* https://doi.org/10.6028/NIST.SP.800-82r3

*Universidad Internacional de Ciencia y Tecnología (UNICyT). (2018). Normas de trabajos con opción a grado de maestría.* https://unicyt.org/files/NORMAS-DE-TRABAJOS-CON-OPCION-A-GRADO-DE-MAESTRÍA.pdf

*Westerinen, A., Schnizlein, J., Strassner, J., Scherling, M., Quinn, B., Herzog, S., Huynh, A., Carlson, M., Perry, J., & Waldbusser, S. (2001, noviembre). Terminology for policy-based management (RFC 3198). RFC Editor.* https://doi.org/10.17487/RFC3198

*Zenarmor. (2024, 8 de marzo). What is NetFlow? Understanding the network flow data. Recuperado el 31 de enero de 2026, de* https://www.zenarmor.com/docs/network-basics/what-is-netflow

# ANEXOS

**ANEXO A. Serie temporal (ilustrativa) con picos detectados en 02:05 y 11:30, con score y trazabilidad (origen/destino/puerto). Generado por IA Gemini**

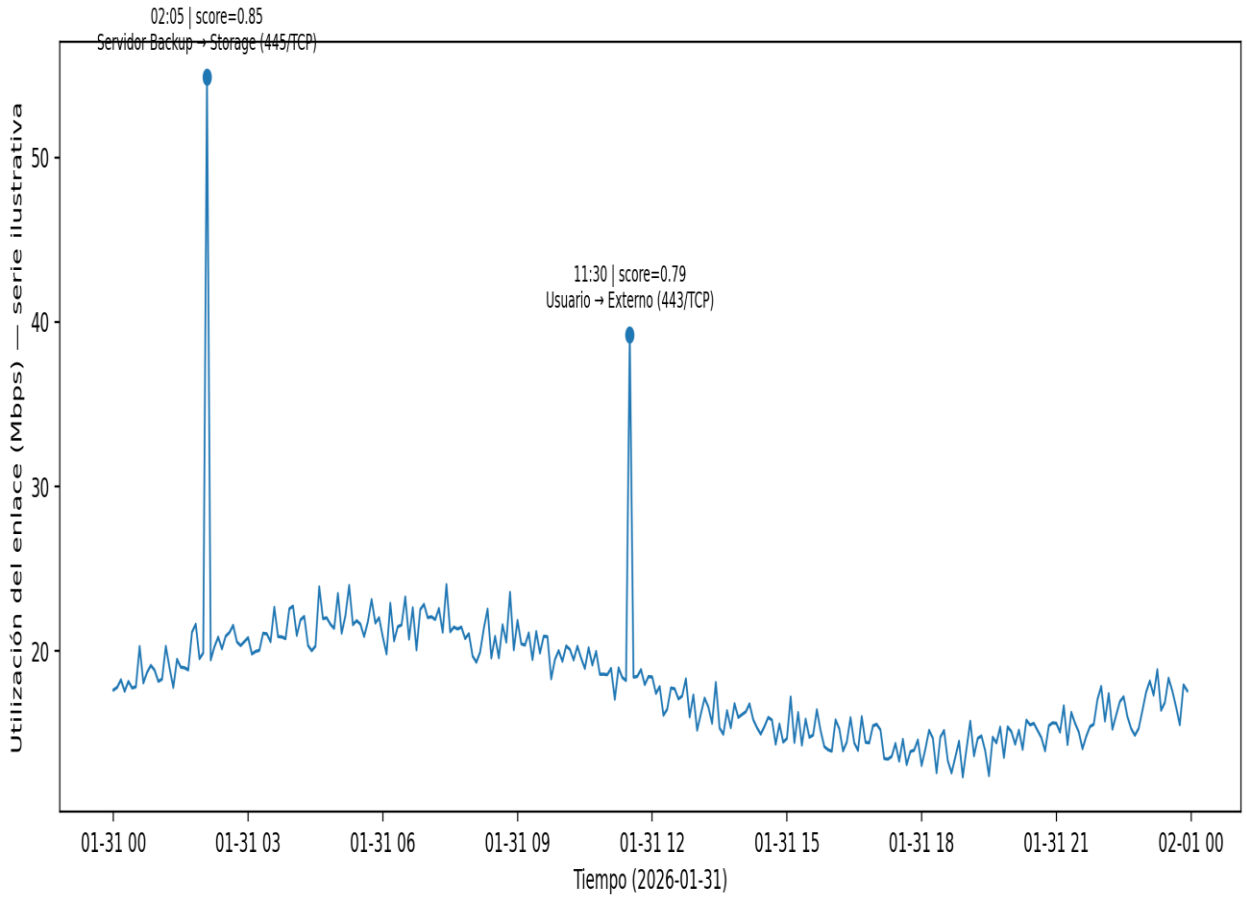



**Figura A1**

*Serie temporal ilustrativa con picos de tráfico detectados.*

*Nota.* Serie temporal ilustrativa con picos de tráfico detectados a las 02:05 y 11:30. Los datos son simulados con fines demostrativos. Generado con apoyo de IA Gemini.

**ANEXO B. Diccionario de datos (SNMP y NetFlow/IPFIX)**

Tabla B1

*Diccionario de datos: variables SNMP y NetFlow/IPFIX utilizadas en el pipeline*

| Fuente | Variable / Campo | Descripción | Unidad | Frecuencia | Transformación / Cálculo | Uso en el modelo |
|---|---|---|---|---|---|---|
| SNMP (por interfaz) | ifInOctets | Contador de bytes recibidos en la interfaz. | bytes | cada 5 min | Δ(ifInOctets)/Δt → bps/Mbps; opcional: %utilización = (bps/capacidad_enlace)×100 | Feature base (tráfico entrante) |
| SNMP (por interfaz) | ifOutOctets | Contador de bytes transmitidos en la interfaz. | bytes | cada 5 min | Δ(ifOutOctets)/Δt → bps/Mbps; opcional: %utilización = (bps/capacidad_enlace)×100 | Feature base (tráfico saliente) |
| SNMP (por interfaz) | ifInErrors | Errores de entrada reportados por la interfaz. | conteo | cada 5 min | Δ(ifInErrors)/Δt → tasa de errores | Contexto de degradación / diagnóstico |
| SNMP (por interfaz) | ifOutErrors | Errores de salida reportados por la interfaz. | conteo | cada 5 min | Δ(ifOutErrors)/Δt → tasa de errores | Contexto de degradación / diagnóstico |
| SNMP (por dispositivo) | CPU | Uso de CPU del dispositivo de red. | % | cada 5 min | Promedio por ventana si aplica | Contexto operativo (carga del equipo) |
| NetFlow/IPFIX (por flujo) | src_ip | Dirección IP origen del flujo. | texto | agregado cada 5 min | Top-N por bytes/paquetes por ventana | Atribución (quién genera el pico) |
| NetFlow/IPFIX (por flujo) | dst_ip | Dirección IP destino del flujo. | texto | agregado cada 5 min | Top-N por bytes/paquetes por ventana | Atribución (hacia dónde va el tráfico) |
| NetFlow/IPFIX (por flujo) | dst_port | Puerto de destino del flujo. | entero | agregado cada 5 min | Distribución de puertos (p. ej., 445, 443) | Asociación de servicio / causa raíz |
| NetFlow/IPFIX (por flujo) | protocol | Protocolo de transporte del flujo. | entero/texto | agregado cada 5 min | Agrupación por protocolo (TCP/UDP) | Contexto de clasificación |
| NetFlow/IPFIX (por flujo) | bytes | Bytes transferidos en el flujo. | bytes | agregado cada 5 min | Suma por ventana e IP/puerto | Intensidad del evento / ranking de responsables |
| NetFlow/IPFIX (por flujo) | packets | Paquetes transferidos en el flujo. | paquetes | agregado cada 5 min | Suma por ventana; razón bytes/paquete si aplica | Caracterización de patrón (bulk vs. many small packets) |
| NetFlow/IPFIX (por flujo) | flow_duration | Duración del flujo. | s/ms | agregado cada 5 min | Media/mediana por ventana | Caracterización (sesiones largas vs. ráfagas) |

| Fuente | Variable / Campo | Descripción | Unidad | Frecuencia | Transformación / Cálculo | Uso en el modelo |
|---|---|---|---|---|---|---|
| NetFlow/IPFIX (agregado) | active_flows | Cantidad de flujos activos en la ventana. | conteo | cada 5 min | Conteo total por ventana | Feature para detectar ráfagas y cambios abruptos |
| Feature derivada | Utilización | Tasa de utilización del enlace en la ventana. | % | cada 5 min | (bps/capacidad_enlace)×100 | Entrada principal a los modelos (por interfaz/VLAN) |
| Feature derivada | Derivada temporal | Cambio abrupto respecto a una ventana histórica. | Δ% | cada 5 min | valor_actual − promedio(últimos 30 min) | Detección de incrementos súbitos |
| Feature derivada | IQR | Dispersión estadística para variación anómala. | unidad de la métrica | por ventana | IQR = Q3 − Q1 sobre ventana histórica | Identificación de variabilidad fuera de patrón |
| Feature derivada | Distribución de puertos | Concentración de tráfico por puerto destino. | %/conteo | por ventana | Top puertos por bytes/flows | Asociación de causa (servicio/protocolo) |
| Feature derivada | Proporción bytes por src_ip | Participación relativa del origen principal. | % | por ventana | bytes_src_ip / bytes_totales_ventana | Atribución (origen dominante) |

*Nota.* Diccionario de variables originales recolectadas y características derivadas. La frecuencia de muestreo base es de 5 minutos. Adaptado de la documentación técnica de SNMP (RFC 1213) y NetFlow/IPFIX (Claise et al., 2013).

## ANEXO C. EVIDENCIA DE CONFIGURACIÓN (SANITIZADA)

Se presentan extractos de configuración correspondientes al exportador NetFlow/IPFIX y al sondeo SNMP utilizados para la captura de datos en la arquitectura propuesta. La información sensible (direcciones IP reales, nombres de dispositivos, credenciales y comunidades) ha sido anonimizada para preservar la confidencialidad de la infraestructura empresarial.

Exportador NetFlow/IPFIX (Ejemplo Cisco IOS)

Configuración Global:

```
flow exporter EXPORTER-ML
 destination 10.0.0.100
 source GigabitEthernet0/0
 transport udp 2055
 template data timeout 60
 option interface-table timeout 60
```

Definición del Flow Monitor:

```
flow monitor MONITOR-ML
 exporter EXPORTER-ML
 cache timeout active 60
 record netflow-original
```

Aplicación a interfaz (Uplink):

```
interface GigabitEthernet1/0/1
 description Uplink-Core
 ip flow monitor MONITOR-ML input
 ip flow monitor MONITOR-ML output
```

Nota de sanitización: La IP del colector, nombres de monitor e interfaces han sido reemplazados por valores genéricos. No se muestran rutas, gateways ni credenciales. Configuración SNMP

SNMPv2c (Ejemplo Básico):

```
snmp-server community MONITOR-RO RO
snmp-server location DataCenter-01
snmp-server contact admin@example.local
snmp-server enable traps
```

SNMPv3 (Recomendado para entorno productivo):

```
snmp-server group ML-GROUP v3 priv
snmp-server user ML-USER ML-GROUP v3 auth sha AUTH-PASS priv aes 128 PRIV-PASS
snmp-server host 10.0.0.100 version 3 priv ML-USER
```

Nota de sanitización: Comunidad, usuario y credenciales han sido sustituidos por placeholders. La dirección del servidor de monitoreo ha sido anonimizada.

Intervalo de Sondeo (Polling)

Intervalo de polling configurado: 300 segundos (5 minutos).

Principales OIDs recolectados:

- ifInOctets
- ifOutOctets
- ifInErrors
- ifOutErrors
- hrProcessorLoad

Las configuraciones presentadas son representativas del entorno implementado para la validación conceptual del modelo. Todos los identificadores reales fueron reemplazados con valores genéricos con el fin de preservar la confidencialidad de la infraestructura organizacional, sin alterar la validez técnica de la evidencia mostrada

**ANEXO D. PROMPTS UTILIZADOS PARA FACILITAR LA GENERACION DE DETALLES TECNICOS.**

Cada uno de estos prompts sirvió para generar detalles técnicos y métricas de manera coordinada y previamente validada En muchos casos no se tomó todo el contenido generado, sino que se utilizó más bien como apoyo para elaboración.

- Generación de data set ficticio (SNMP + NetFlow/IPFIX)

Actúa como ingeniero de redes y científico de datos. Genera un conjunto de datos TOTALMENTE FICTICIO (no real) para una LAN empresarial, combinando métricas tipo SNMP y NetFlow/IPFIX.

Requisitos:

Periodo: 7 días, muestreo cada 5 minutos.

Columnas mínimas: timestamp, device_id, interface, in_mbps, out_mbps, pps_in, pps_out, flows, top_src_ip, top_dst_ip, top_dst_port, protocol, app_hint, note.

IPs privadas (RFC1918) y nombres genéricos (SW-01, FW-01, SRV-DB, etc.).

Inserta 5 picos/anomalías con causa plausible (backup nocturno, actualización masiva, cámara CCTV saturando, torrent, exfiltración simulada).

Mantén rangos realistas (ej.: enlaces 100–200 Mbps, picos de 80–95% del enlace).

Entrega el dataset como tabla Markdown y además una versión “lista para Excel” separada por comas (CSV) con 200 filas de muestra representativa.

- Escenarios controlados para comparar resultados

Genera datos ficticios en 3 escenarios distintos para detección de picos en LAN:

A) “Operación normal con picos benignos”

B) “Congestión por actividad interna (backup/replicación)”

C) “Evento sospechoso (escaneo + transferencia anómala simulada)”

Para cada escenario, crea 120 registros (5 min) con las mismas columnas: timestamp, in_mbps, out_mbps, pps, flows, top_talker, port, label_ground_truth (normal/anomalia).

Aclara en una nota que los datos son simulados y que label_ground_truth se usa solo para evaluación conceptual.

Devuelve 3 tablas separadas en Markdown.

- Tabla extendida de eventos detectados (más filas que en el capítulo)

A partir de un dataset simulado de tráfico LAN, crea una “Tabla extendida de eventos detectados” con 30 filas.

Columnas: evento_id, timestamp_inicio, timestamp_fin, duracion_min, tipo_pico (corto/medio/prolongado), in_mbps_peak, out_mbps_peak, pps_peak, flows_peak, top_src_ip, top_dst_ip, top_dst_port, probable_causa, impacto_estimado (bajo/medio/alto), accion_sugerida.

Reglas: IPs privadas, causas plausibles, y coherencia interna (si el puerto es 443, sugiere HTTPS; si 445, SMB; si 53, DNS, etc.).

Entrega en formato de tabla Markdown con título “Anexo: Eventos detectados (datos simulados)”.

- Resultados ficticios por algoritmo (Isolation Forest / One-Class SVM / DBSCAN)

Actúa como analista de datos. Con base en un dataset ficticio de tráfico LAN con picos insertados, genera resultados comparativos simulados para 3 algoritmos no supervisados: Isolation Forest, One-Class SVM y DBSCAN.

Entrega:

Una tabla con columnas: algoritmo, TPR, FPR, precision, recall, F1, tiempo_entrenamiento_ms, tiempo_inferencia_ms, comentario.

Asegura coherencia: DBSCAN tiende a marcar ruido si parámetros no ajustan; One-Class SVM sensible a kernel/nu; Isolation Forest estable con contaminación baja-moderada.

Indica explícitamente que los valores son referenciales y simulados (no medidos en producción).

- Tabla de hiperparámetros probados + criterio de selección

Construye una tabla ficticia de hiperparámetros probados para:

Isolation Forest (n_estimators, max_samples, contamination, max_features)

One-Class SVM (kernel, nu, gamma)

DBSCAN (eps, min_samples, metric)

Incluye: valores_probados, valor_elegido, justificacion_breve.

El criterio de selección debe mencionar balance entre falsos positivos, estabilidad y costo computacional, dentro de un enfoque conceptual.

Devuélvelo en una tabla Markdown titulada “Tabla: Hiperparámetros evaluados (simulado)”.

- Pipeline de preprocesamiento (pseudocódigo)

Redacta un pipeline de preprocesamiento para tráfico de red (SNMP + NetFlow/IPFIX) en pseudocódigo claro y breve.

Debe incluir: sincronización temporal, manejo de faltantes, agregación por ventana (5 min), normalización Min-Max, creación de features (rolling mean, rolling std, ratio in/out, z-score local), y salida lista para modelos no supervisados.

Indica que es un pipeline conceptual.

Entrega solo el pseudocódigo.

- Evidencia de configuración sanitizada (NetFlow/IPFIX + SNMP)

Genera extractos de configuración SANITIZADOS y FICTICIOS que evidencien:

A) Exportación NetFlow/IPFIX (con IPs privadas, puertos ejemplo, sin credenciales reales).

B) Configuración SNMP (solo versión, comunidad enmascarada, usuarios genéricos, ACL con rangos privados).

Requisitos: no incluir datos sensibles, reemplazar secretos por <REDACTED>, y añadir un comentario al inicio: “Configuración simulada para fines académicos”.

Entrega los extractos en bloques de texto monoespaciado.

- Comparación de falsos positivos por escenario

Con base en 3 escenarios simulados (normal, backup interno, evento sospechoso), genera una tabla comparativa de falsos positivos por algoritmo.

Columnas: escenario, IsolationForest_FPR, OCSVM_FPR, DBSCAN_FPR, nota_interpretativa.

Los valores deben ser coherentes y explícitamente simulados.

Devuelve en tabla Markdown.

- Tiempos de procesamiento (benchmark ficticio)

Crea un benchmark simulado de tiempos de procesamiento para el pipeline completo (ingesta → features → detección → consolidación).

Incluye 3 tamaños de carga: 10k, 100k y 1M registros.

Tabla con: registros, tiempo_ingesta_s, tiempo_features_s, tiempo_modelo_s, tiempo_total_s, uso_ram_gb_estimado.

Indica que es una estimación conceptual, no medición real.