

UNIVERSIDAD PERUANA UNIÓN
FACULTAD DE INGENIERÍA Y ARQUITECTURA
E.P. DE INGENIERÍA DE SISTEMAS



INFORME DE AVANCE – UNIDAD 1

CURSO:
Minería de datos

DOCENTE:
ING. ABEL ANGEL SULLON MACALUPU

ESTUDIANTE:
Rubén Mark Salazar Tocas
Uziel Sauñe Fernadez

LIMA, SEPTIEMBRE 2025

Contenido

Proyecto: Minería de Datos para la Automatización del Monitoreo de Infraestructura TI: Un Enfoque Predictivo con Zabbix	2
Fase 1 – Business Understanding	2
1.1 Introducción	2
Contexto global	2
Contexto del problema.....	2
Estado del arte internacional	3
Estado del arte nacional/local	3
Vacío en el conocimiento	3
Propósito u objeto de estudio	3
Justificación	4
1.2 Pregunta, objetivos e hipótesis	4
Pregunta de investigación	4
Objetivo general	4
Objetivos específicos	4
Hipótesis.....	5
1.3 Criterios de éxito	5
Técnico	5
De negocio	5
Social/operativo	5
Anexos:.....	6

Proyecto: Minería de Datos para la Automatización del Monitoreo de Infraestructura TI: Un Enfoque Predictivo con Zabbix

Metodología: CRISP-DM

En este proyecto aplicamos la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), que define un proceso estructurado para proyectos de minería de datos. En esta primera unidad avanzaremos con las fases:

→ Business Understanding → Data Understanding → Data Preparation

Integrantes:

- Rubén Mark Salazar Tocas
- Elías Uziel Sauñe Fernández

Fecha del informe: 09/09/2025

Fase 1 – Business Understanding

1.1 Introducción

Contexto global

En la actualidad, la dependencia de las organizaciones hacia la infraestructura tecnológica es cada vez mayor. Universidades, hospitales, bancos y empresas en general requieren que sus sistemas funcionen de manera continua, ya que una interrupción puede generar pérdidas económicas, impacto en la reputación institucional y dificultades para los usuarios. A nivel global, Se estima que cada minuto de inactividad en sistemas críticos puede generar pérdidas que oscilan entre **5.600 y 9.000 dólares**, dependiendo del sector. Esta cifra nos hace conscientes de que la continuidad de servicio ya no es un lujo, sino una necesidad estratégica.

Contexto del problema

Como grupo identificamos que la Universidad Peruana Unión (UPeU) es un buen ejemplo de una organización compleja donde la infraestructura TI es clave. La universidad cuenta con tres sedes principales (Lima, Juliaca y Tarapoto) y depende de sistemas como plataformas virtuales, bibliotecas digitales y sistemas de gestión académica. Un pico de consumo de CPU o memoria RAM en los servidores puede provocar lentitud o incluso caídas de servicio, afectando tanto a los estudiantes como a los docentes y administrativos.

Actualmente, Zabbix ya permite monitorear métricas de infraestructura en tiempo real y generar alertas cuando se superan ciertos umbrales. Sin embargo, como equipo hemos visto que este enfoque es reactivo, ya que la alerta llega cuando el problema ya está ocurriendo. La brecha está en pasar de una reacción tardía a una predicción anticipada, que nos permita actuar antes de que el sistema colapse.

Estado del arte internacional

En la literatura internacional se han explorado diferentes enfoques para predecir fallos y optimizar recursos de TI. Por ejemplo, Malhotra et al. (2015) aplicaron redes LSTM para la detección temprana de anomalías en series temporales de servidores, obteniendo un mejor rendimiento que modelos lineales como ARIMA. De manera similar, Hundman et al. (2018) utilizaron LSTM para predecir anomalías en telemetría de satélites de la NASA, lo cual demostró la robustez de este enfoque frente al ruido y valores atípicos. Estos estudios nos inspiran a utilizar técnicas de deep learning en nuestro propio contexto.

Estado del arte nacional/local

En el Perú aún son pocos los trabajos publicados sobre minería de datos aplicada al monitoreo de TI. La mayoría de esfuerzos están centrados en la gestión de redes y ciberseguridad, mientras que la predicción de saturación de infraestructura en universidades todavía no se ha explorado con profundidad. Este vacío nos motiva a tomar el caso de la UPeU como ejemplo, dado que su tamaño organizacional y la criticidad de sus servicios la convierten en un entorno realista para evaluar un modelo predictivo con impacto tangible.

Vacío en el conocimiento

Detectamos que, aunque Zabbix ya se usa en instituciones para monitoreo, no existen implementaciones documentadas que integren predicción con LSTM dentro de la plataforma. ARIMA y otros enfoques lineales suelen dar resultados aceptables, pero presentan limitaciones frente a picos anómalos y ruido en la serie temporal, lo que puede llevar a falsos positivos o falsos negativos en las alertas. Este vacío justifica la exploración de un modelo de deep learning especializado en series temporales.

Propósito u objeto de estudio

El propósito de nuestro proyecto como grupo es implementar un modelo predictivo con LSTM, integrado a Zabbix, que anticipe el consumo de CPU y memoria en servidores críticos de la UPeU. Queremos demostrar que este enfoque permite pasar de un monitoreo reactivo a uno proactivo, generando alertas con anticipación suficiente para prevenir caídas y mejorar la gestión de los recursos tecnológicos.

Justificación

Teórica: contribuimos a la literatura nacional sobre minería de datos aplicada al monitoreo TI, un campo poco explorado en universidades peruanas.

Práctica: dotamos a la UPeU de una herramienta predictiva que permite optimizar sus recursos tecnológicos y evitar pérdidas de servicio.

Metodológica: aplicamos la metodología CRISP-DM y un modelo de deep learning (LSTM) para el tratamiento de series temporales de uso de CPU y memoria.

Social: apoyamos la continuidad de los servicios educativos y administrativos, beneficiando directamente a estudiantes, docentes y trabajadores.

1.2 Pregunta, objetivos e hipótesis

Pregunta de investigación

¿Es posible construir un modelo predictivo basado en LSTM, integrado con Zabbix, que permita anticipar el consumo crítico de CPU y memoria en los servidores de la UPeU, evitando saturaciones y caídas inesperadas?

Objetivo general

Construir un modelo predictivo basado en LSTM integrado con Zabbix, capaz de anticipar el uso de CPU y memoria RAM, reduciendo el riesgo de saturación y mejorando la administración preventiva de recursos tecnológicos.

Objetivos específicos

- Analizar los datos recolectados desde Zabbix sobre uso de CPU y memoria.
- Realizar un diagnóstico de la calidad y comportamiento de las series temporales.
- Diseñar y entrenar un modelo LSTM para la predicción de métricas de infraestructura.
- Comparar el desempeño del LSTM frente a enfoques tradicionales como ARIMA.
- Integrar el modelo predictivo con Zabbix y evaluar su utilidad práctica en la anticipación de alertas.

Hipótesis

Hipótesis general: un modelo LSTM puede anticipar con precisión los picos de consumo de CPU y memoria en los servidores de la UPeU, reduciendo el riesgo de saturación.

Hipótesis específicas:

H1: el modelo LSTM tendrá un menor error de predicción (RMSE y MAPE) frente a ARIMA.

H2: la integración de predicciones con Zabbix permitirá generar alertas preventivas con al menos 15 minutos de anticipación.

1.3 Criterios de éxito

Técnico

- Alcanzar métricas de predicción satisfactorias: RMSE bajo y $MAPE \leq 10\%$ en CPU y RAM.
- Validar que el modelo LSTM maneja mejor el ruido y los picos atípicos en comparación con ARIMA.

De negocio

- Que Zabbix reciba predicciones con al menos 15–30 minutos de anticipación antes de una saturación real.
- Reducir el número de falsos positivos y falsos negativos en comparación con el método tradicional de alertas.

Social/operativo

- Garantizar la continuidad de los servicios tecnológicos en la UPeU, evitando caídas imprevistas.
- Optimizar el uso de recursos humanos y económicos, pasando de un esquema reactivo a uno preventivo.

Anexos:

PRESENTACION CANVAS

https://www.canva.com/design/DAGv9qNOBiA/3CD70258KUxK3bnX4js9zQ/edit?utm_content=DAGv9qNOBiA&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton