

Caso Contugas - Detección de Anomalías en consumo de gas

Introducción

Este documento presenta la solución final desarrollada para la detección automática de valores atípicos en los datos operativos de la empresa distribuidora de gas Contugas S.A.C. en Perú. La solución ha sido implementada con modelos estadísticos y de machine learning, integrados en un pipeline ejecutable a través de notebooks y un dashboard web interactivo. El objetivo del prototipo es apoyar la supervisión técnica y operativa mediante herramientas analíticas que permiten identificar comportamientos anómalos en variables como volumen, presión y temperatura..

Requerimientos del Artefacto

Los siguientes requerimientos fueron definidos para el desarrollo del prototipo. A continuación se detalla su grado de cumplimiento en el prototipo final:

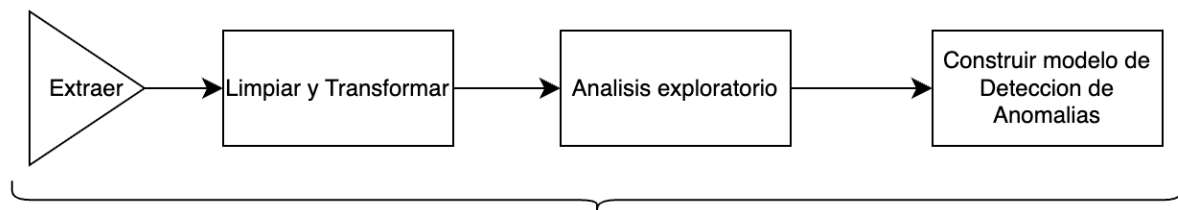
Requerimiento	Estado	Comentario
Visualización de consumo y variables operacionales	Cumplido	A través del dashboard en Streamlit
Alertas automáticas si se detectan anomalías	Parcial	Se detectan, pero no se notifican en tiempo real. Esto se puede implementar en un siguiente alcance del proyecto
Reporte descargable por cliente	Cumplido	Desde la vista de perfil en el dashboard
Filtro por fecha, tipo de cliente, ciudad, etc.	Cumplido	Filtros dinámicos en el dashboard
Funcionar en la nube	Cumplido	Desplegado en Streamlit Cloud
No requerir codificación para el usuario	Cumplido	Interfaz visual amigable
Actualizaciones automáticas cada hora	No implementado	Proceso manual, pero estructurado para automatización

Desarrollo del Artefacto

Sobre los Datos

La solución fue implementada a partir de datos históricos entregados por Contugas. Se realizaron procesos de validación de calidad de datos, limpieza, transformación y modelado automático por cliente y por métrica. Los modelos seleccionados fueron Prophet y SARIMA, según la naturaleza estacional o estacionaria de cada serie temporal.

El procesamiento se encuentra documentado y ejecutable en notebooks en Python. Los resultados finales se consolidan en un archivo CSV que alimenta un dashboard construido con Streamlit.



Herramientas

GIT: Para el versionamiento del código de cada una de las etapas.

DVC: Versionamiento de los datos preprocesados para la construcción del modelo.

MLFlow: Para gestionamiento de los experimentos de machine learning.

Python: Lenguaje de programación base para hacer uso de los paquetes de análisis de datos para todos los procesos.

Modelos y Tipos de Análisis

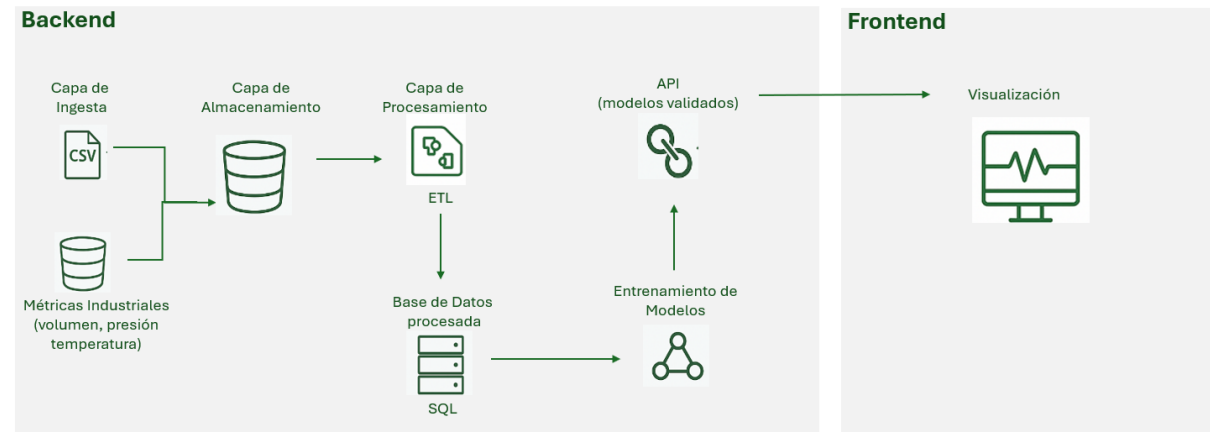
Se aplicaron modelos estadísticos (desviación estándar, IQR, percentiles) y modelos no supervisados como Isolation Forest, DBSCAN y Autoencoders, además de modelos de series temporales Prophet y SARIMA. La selección por cliente depende del comportamiento de la serie (estacional o estacionaria).

Diagrama de Arquitectura

Back - End

La arquitectura técnica del artefacto propuesto define los componentes fundamentales que permiten transformar los datos operativos de Contugas en información útil para la detección temprana de anomalías. El diagrama del back-end representa el flujo lógico y tecnológico del sistema, desde la entrada de datos generados por los medidores industriales, pasando por

los procesos de limpieza, transformación y análisis, hasta su visualización final en un dashboard interactivo diseñado para los usuarios operativos:



Componentes del Backend:

Métricas Industriales	Representa los datos recolectados de los medidores de volumen, presión y temperatura.
Capa de Ingesta	Proceso que transforma los datos capturados en archivos CSV para su almacenamiento inicial.
Capa de Almacenamiento	Almacenamiento intermedio de los datos en su formato bruto en la nube (AWS S3, Azure Blob).
Capa de Procesamiento (ETL)	Transformación, limpieza, normalización y estructuración de los datos para análisis.
Base de Datos Limpia (SQL)	Repositorio de datos ya depurados y estructurados, listos para ser usados por modelos.
Iteración de Modelos	Entrenamiento, evaluación y validación de modelos de detección de anomalías.

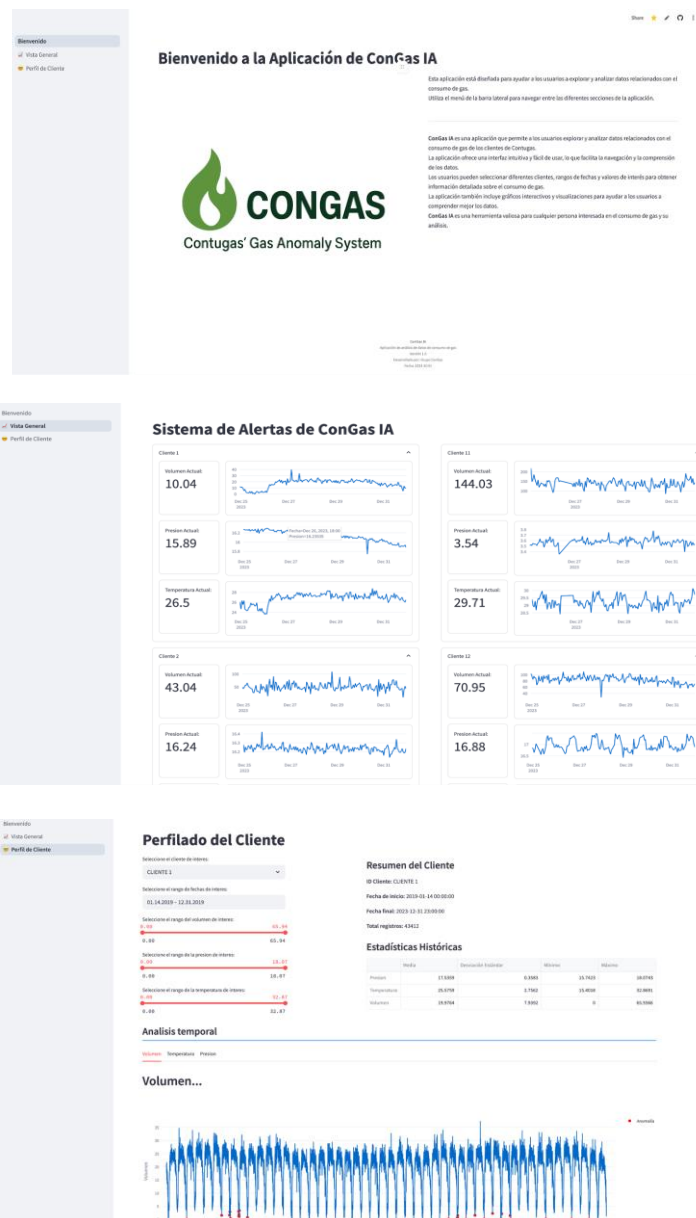
API de Modelos

Servicio de integración que expone las predicciones o anomalías detectadas al sistema web.

Visualización (Frontend)

Dashboard accesible por los usuarios vía navegador web.

Front - End



Entrega Final del Prototipo

El artefacto ConGas-IA ha sido entregado con los siguientes componentes:

- Notebooks: ProyectoFinal.ipynb y ProyectofinalExploracion.ipynb
- Dashboard en Streamlit: Visualización de anomalías, filtros por cliente y variables
- Archivo anomaliasDetectadas.csv generado por el modelo
- Repositorio GitHub con código fuente y documentación

Acceso a la app: <https://proyectouandes-kukn8x3ysysm2iybfuwgx2.streamlit.app/>

Repositorio GitHub: <https://github.com/josesu92/proyectoUAndes>