

Guardian de fraudes: Analítica predictiva para la optimización del consumo de gas para la empresa Contugas

Juan Sebastian Perdomo Becerra js.perdomo2@uniandes.edu.co

Este documento presenta el prototipo fachado del sistema 'Guardian de Fraudes', una solución analítica orientada a la detección anticipada de consumos anómalos en clientes industriales del servicio de gas. La construcción del prototipo hace parte del proyecto final de la Maestría en Inteligencia Analítica de Datos, y tiene como objetivo generar alertas tempranas y comprensibles para apoyar la toma de decisiones en terreno y desde operaciones. A continuación, se detalla su lógica, arquitectura, funcionalidades principales, modelos implementados y herramientas tecnológicas empleadas en su desarrollo.





1 Usuario Final y Necesidad

El usuario objetivo del sistema son los analistas técnicos de Contugas encargados de monitorear el comportamiento de consumo de los clientes industriales. La necesidad que da origen al artefacto surge de una oportunidad concreta: hoy el monitoreo se hace de manera reactiva, limitando la capacidad de anticiparse a eventos anómalos que pueden generar afectaciones económicas, técnicas o de confiabilidad. Con este dashboard se busca facilitar un análisis preventivo, rápido, intuitivo y accionable del comportamiento de los clientes a partir de los datos horarios históricos que se tienen registrados.

2 Objetivo del Artefacto

Desarrollar un dashboard interactivo y visualmente amigable que permita: visualizar el comportamiento histórico de cada cliente, identificar comportamientos anómalos mediante modelos no supervisados, resumir hallazgos por cliente o globalmente, cargar datos para validaciones puntuales, ajustar la sensibilidad del modelo, y proponer recomendaciones básicas. Futuramente, se plantea incorporar datos contextuales como clima o demanda.

3 Datos Disponibles

Se cuenta con una base de datos histórica suministrada por Contugas, en formato Excel, estructurada por cliente. Cada hoja contiene datos horarios de las variables:

Fecha, Presión, Temperatura y Volumen. No se cuenta con etiquetas de anomalías, por lo cual el enfoque será no supervisado.

4 Preguntas Analíticas

- ¿Cómo ha sido el comportamiento de consumo del cliente a lo largo del tiempo?
- ¿Se han presentado comportamientos fuera de patrón y con qué frecuencia?
- ¿Qué variables presentan mayor desviación respecto al patrón normal?
- ¿Qué score tiene un dato puntual que el usuario desea validar?
- ¿Qué tipo de anomalía parece y qué acción se sugiere?

5 Producto final

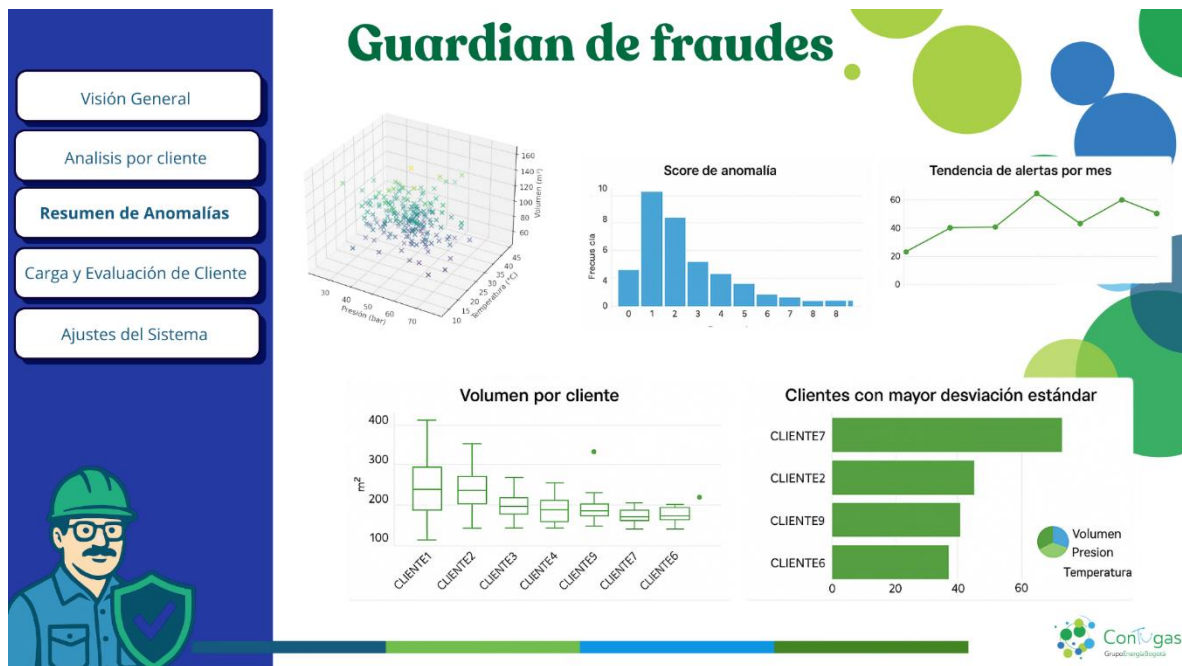
Se construirá un reporte en un dashboard que permita al cliente encontrar detecciones tempranas además KPI relevantes de la operación



Muestra indicadores clave de operación, como número de clientes monitoreados, cantidad de alertas, distribución de anomalías por variable y tendencias agregadas en el tiempo.



Permite seleccionar un cliente específico y visualizar sus variables (presión, temperatura, volumen) a lo largo del tiempo, resaltando puntos anómalos y su score.



Brinda visualizaciones transversales como histogramas de scores, boxplots por cliente y diagramas de dispersión 3D (PVT) para entender mejor la naturaleza de los comportamientos atípicos.

Permite subir un archivo puntual para analizar si los datos de un nuevo cliente, o de un periodo no contemplado, presentan comportamientos anómalos. Incluye

clasificación tentativa del tipo de irregularidad.



Desde esta sección se pueden modificar parámetros clave del modelo: sensibilidad por variable, selección de algoritmo (Isolation Forest, Prophet, Z-score), y ver el efecto en la generación de alertas.



6 Modelos y Lógica Analítica

Se emplean Isolation Forest, Z-score, Prophet, Mahalanobis y lógica basada en reglas o árboles de decisión para clasificación del tipo de anomalía.

7 Backend y Arquitectura Tecnológica

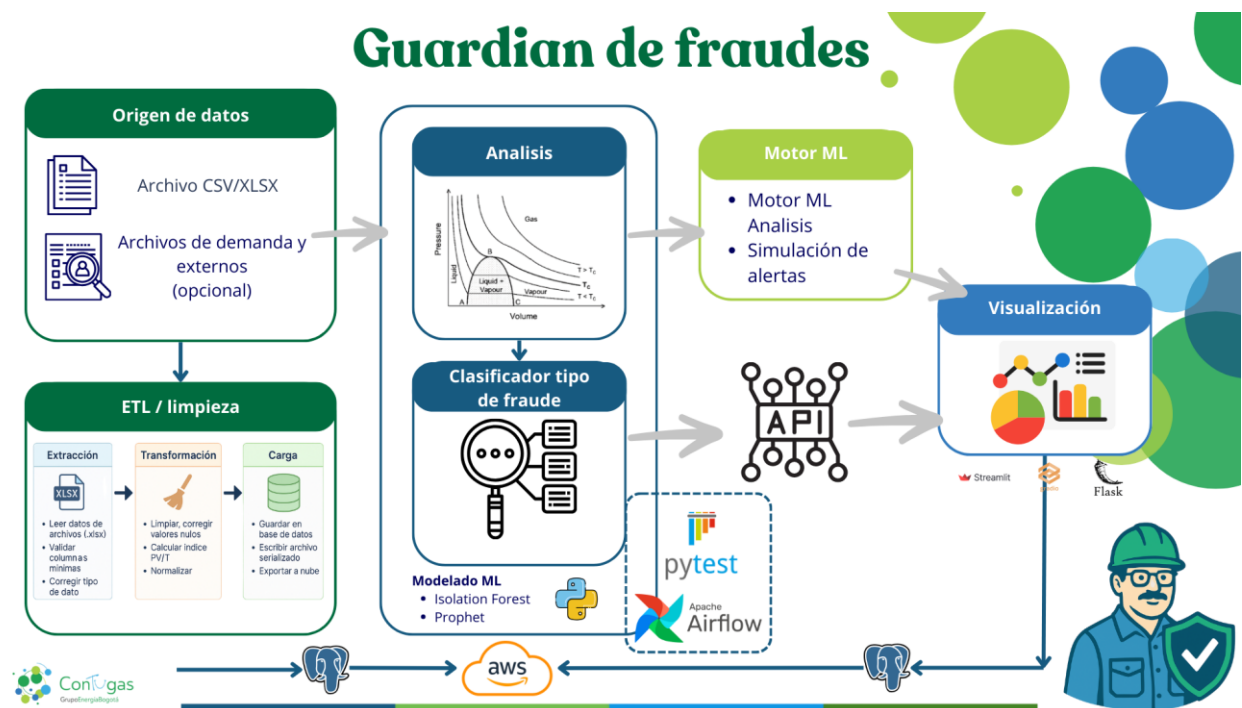
Tecnologías utilizadas: pandas, sklearn, streamlit, flask, postgresSQL, pytest, airflow. El sistema operará sobre AWS (EC2, S3). Toda la ejecución se orquesta en 8 semanas con control de versiones en Git y pruebas automatizadas. La arquitectura es modular, escalable y permite ejecución local o como API.

8 Diagrama de Arquitectura

A continuación, se muestra la arquitectura del sistema Guardian de Fraudes, desde la carga de datos hasta la visualización final del dashboard.

9 Alcance y Consideraciones

Este prototipo tiene como eje la simplicidad y la capacidad de detección preventiva. Se contempla una milla extra para clasificación del tipo de anomalía y recomendaciones. Es un producto escalable y con proyección a integración futura con fuentes externas.



10 Anexos

Adjunto el detalle del mock up del Dashboard



Dash board.pdf



Imagen prototipo
fachada.pdf