



Grupo 25

GUIA DE USUARIO PARA INSTALACION DE PROTOTIPO Y SU DESCRIPCION

Jhon Farley Adarve Diaz

j.adarve@uniandes.edu.co

Giovvany Andres Cardona Rojo

g.cardonar@uniandes.edu.co

Angel Benito-Rebollo Sierra

a.benitorebollo@uniandes.edu.co

1. ¿Qué es y qué hace el artefacto propuesto?

- ¿Qué es?

El "Sistema de Detección de Anomalías - Contugas" es un prototipo de herramienta analítica avanzada, desarrollado como respuesta a los desafíos críticos que enfrenta Contugas en la detección oportuna de irregularidades en el consumo de gas (de sus clientes industriales en el departamento de Ica, Perú). Este sistema busca modernizar los procesos de monitoreo, alineándose con la estrategia de innovación y eficiencia operativa del Grupo Energía Bogotá (GEB).

- ¿Qué hace?

- **Aborda un problema crítico:** Contugas enfrenta el reto de identificar patrones de consumo inusuales que pueden indicar fugas, fallas en medidores, daños estructurales o manipulaciones externas. Estos eventos, si no se detectan a tiempo, generan pérdidas económicas, riesgos operacionales y pueden impactar la seguridad y continuidad del suministro.
- **Procesa datos operativos:** Carga y procesa mediciones horarias de consumo (presión, temperatura, volumen) de los 20 clientes industriales monitoreados.
- **Detecta anomalías inteligentemente:** Utiliza un modelo de Machine Learning no supervisado (Isolation Forest) para identificar automáticamente puntos de datos que se desvían significativamente del comportamiento normal histórico de cada cliente, sin necesidad de datos etiquetados previamente.
- **Visualiza información para la acción:** Presenta los datos de consumo, las anomalías detectadas y estadísticas relevantes en un dashboard interactivo y amigable.
- **Clasifica y Alerta para priorizar:** Asigna un nivel de severidad (Baja, Media, Alta) a las anomalías detectadas basado en la magnitud de la desviación, permitiendo al equipo operativo de Contugas enfocar sus esfuerzos de manera más efectiva.
- **Facilita el análisis proactivo:** Permite a los usuarios filtrar datos por cliente, rango de fechas, variable de consumo y severidad de la anomalía, facilitando la investigación y la toma de decisiones informada.

- **Ventajas (Impacto en el Negocio):**

- **Mejora la Detección Temprana:** Reduce drásticamente el tiempo de detección de anomalías (objetivo: de 12 horas a 1 hora), permitiendo una respuesta más rápida ante posibles fugas o fallas.
- **Reduce Pérdidas No Técnicas:** Contribuye a la disminución de pérdidas de gas no facturado (objetivo: reducción del 10% al 30%), impactando positivamente la rentabilidad.
- **Optimiza la Eficiencia Operativa:** Al automatizar la detección y ayudar en la priorización, se optimizan los procesos de mantenimiento preventivo y correctivo.

- **Aumenta la Precisión de Detección:** Busca alcanzar una alta precisión en la identificación de eventos inusuales (objetivo: del 85% al 95%).
- **Mejora la Satisfacción del Usuario Operativo:** Proporciona una herramienta intuitiva que facilita el monitoreo y la toma de decisiones (objetivo: tasa de satisfacción del 80% al 95%).
- **Contribuye a la Sostenibilidad:** Al reducir fugas, se disminuye el impacto ambiental y se promueve el uso eficiente de recursos energéticos.
- **Fortalece la Innovación:** Posiciona a Contugas como una empresa que utiliza analítica avanzada para mejorar sus operaciones, alineándose con los objetivos estratégicos del GEB.
- **Limitaciones:**
 - **Calidad de los datos:** Fundamental para la efectividad. Ya mencionamos en el anteproyecto la necesidad de limpieza y validación. Esto básicamente por el gran número de valores faltantes, para esto se propuso trabajar generando por interpolación algunos de estos datos solo cuando es posibles, en espacios grandes de tiempo es necesario obviar esta información y es un punto a trabajar a futuro con los expertos, como capturar la información.
 - **Calibración del modelo:** El parámetro contamination de Isolation Forest y los umbrales de severidad requieren ajuste fino, idealmente con feedback de expertos de Contugas, para balancear la detección de anomalías reales versus falsas alarmas (se establecio un benchmark de Tasa de Falsos Positivos y Negativos combinada no superior al 10%).
 - **Interpretación de severidad:** Requiere validación experta.
 - **Naturaleza no supervisada:** Puede marcar cambios legítimos como anómalos inicialmente.
- **Advertencias Pertinentes:**
 - Este es un prototipo y no debe usarse para la toma de decisiones críticas en un entorno de producción sin una validación exhaustiva y pruebas piloto.
 - Los umbrales de severidad (Alta: 0.75, Media: 0.60 sobre el score de anomalía) son configurables pero requieren validación experta para asegurar su adecuación al contexto de negocio de Contugas.

2. Acciones para ponerlo en funcionamiento y conocimientos/habilidades del usuario:

- **Acciones para ponerlo en funcionamiento (Entorno de Desarrollo/Prueba):**
 1. **Requisitos:** Python 3 instalado.
 2. **Obtener el código:** Clonar el repositorio del proyecto o descargar y descomprimir el archivo ZIP del proyecto en una carpeta local.
 3. **Navegar a la carpeta del proyecto:** Abrir una terminal o línea de comandos y moverse a la carpeta raíz del proyecto (ej. `cd ruta/a/Proyecto_Contugas`).
 4. **Crear y activar un entorno virtual (recomendado).**
 5. **Instalar dependencias.**

6. **Preparar datos crudos:** Colocar los archivos Excel (.xlsx) con los datos de consumo de los clientes en la subcarpeta `data/raw/`. Cada hoja del Excel debe representar un cliente y contener las columnas `Fecha`, `Presion`, `Temperatura`, `Volumen`.
7. **Ejecutar el pipeline ETL:** Esto generará el archivo `data/preprocessed.csv`.
8. **Ejecutar la detección de anomalías:** Esto generará el archivo `data/anomalies.csv` y, si no existe, el modelo entrenado en `models/trained/`.
9. **Lanzar el dashboard.**
10. **Acceder al dashboard:** Abrir un navegador web e ir a <http://127.0.0.1:8050/>.

- **Conocimientos o habilidades que demanda del usuario (operador del dashboard):**

- Comprensión básica de los datos de consumo de gas (presión, temperatura, volumen) y su importancia operativa en Contugas.
- Familiaridad con el uso de interfaces web.
- Capacidad para interpretar gráficos de series temporales y tablas de datos.
- **Conocimiento del contexto operativo de Contugas:** Entender qué podría significar una anomalía de "Presión" para el "CLIENTE12" en un día específico (ej. ¿estaba programado un mantenimiento? ¿hubo algún evento reportado?). El dashboard es una herramienta de apoyo a la decisión, no un reemplazo completo del juicio experto.

3. Casos de uso que soporta y paso-a-paso para el uso esperado:

- **Caso de Uso 1: Monitoreo general del consumo de un cliente específico para identificar patrones inusuales y posibles fugas o fallas (Objetivo: Mejorar eficiencia operativa, reducir pérdidas).**

- **Acceder al Dashboard:** Abrir <http://127.0.0.1:8050/>.
- **Seleccionar Cliente:** En el panel de "Filtros", desplegar la lista "Cliente" y seleccionar el cliente de interés (ej. CLIENTE12).
- **Seleccionar Periodo:** Usar el selector de "Periodo" para definir el rango de fechas a visualizar.
- **Seleccionar Variable:** En "Variable", elegir la magnitud a analizar (ej. Presion, Volumen o Temperatura).
- **Aplicar Filtros:** Hacer clic en el botón "Aplicar Filtros".
- **Analizar Visualización:**

Observar el gráfico principal para ver la evolución de la variable seleccionada y los puntos marcados como anomalías (en rojo).

Revisar los "KPIs" (Total Clientes, Con Anomalías, Alta Severidad, Alertas Activas) que se actualizan según los filtros.

Consultar las "Estadísticas clave" (Promedio, Desv. Estándar, etc.) para la vista actual.

Ver la "Tabla de Anomalías" en la parte inferior para obtener detalles de cada anomalía detectada en el rango filtrado.

- **Caso de Uso 2: Investigar anomalías de una severidad específica para un cliente para priorizar acciones correctivas (Objetivo: Optimizar procesos de mantenimiento, mejorar satisfacción de usuarios operativos).**
 - Seguir los pasos 1-5 del Caso de Uso 1.
 - **Filtrar por Severidad:** En el panel de "Filtros", desplegar la lista "Severidad" y seleccionar el nivel deseado (ej. Media, Baja; o Alta si hubiera).
 - **Aplicar Filtros:** Hacer clic nuevamente en "Aplicar Filtros".
 - **Analizar:** El gráfico y la tabla mostrarán solo las anomalías que coincidan con la severidad seleccionada.
- **Caso de Uso 3: Obtener una lista de clientes con anomalías recientes para una visión general del estado de la red (Objetivo: Detección temprana, apoyo a objetivos estratégicos).**
 - **Acceder al Dashboard.**
 - **Seleccionar Periodo Reciente:** Definir un rango de fechas reciente (ej. última semana).
 - Dejar el filtro "Cliente" sin una selección específica (o si no permite "todos", seleccionar uno y luego observar el panel lateral).
 - **Aplicar Filtros.**
 - **Consultar Panel Lateral:** Observar la lista "Clientes con Anomalías" en el panel lateral izquierdo, que muestra los clientes que presentaron anomalías en el periodo y con la variable/severidad seleccionada (si aplica).
- **Caso de Uso 4: Descargar datos de anomalías filtradas para análisis posterior o reporte (Objetivo: Facilitar documentación y análisis detallado).**
 - Una vez aplicados los filtros deseados (cliente, periodo, variable, severidad).
 - Hacer clic en el botón "Descargar CSV" ubicado debajo de la "Tabla de Anomalías".
 - El navegador descargará un archivo `anomalias_filtradas.csv` con los datos de la tabla visible.

4. Posibles mejoras en versiones futuras

- **Comparar clientes:** El dashboard actual no permite superponer el consumo de múltiples clientes en el mismo gráfico principal.
- **Ver todas las variables de un cliente:** El filtro de "Variable" permite seleccionar una a la vez para el gráfico principal. Para ver todas, el usuario tendría que cambiar el filtro. La tabla de anomalías sí puede mostrar diferentes variables si una anomalía se detectó en múltiples a la vez para el mismo timestamp. Una visualización que muestre las 3 variables simultáneamente para un cliente .

Preguntas Abiertas que Persisten (y que son importantes para el diligenciamiento de la rúbrica y el reporte final):

1. **Línea Base del "Antes":** Para medir el impacto real en la reducción del tiempo de detección (N1) y la reducción de pérdidas (N2, y "Reducción de Pérdidas No Técnicas" del Anteproyecto), ¿existe alguna estimación o dato histórico del desempeño de Contugas *antes* de este prototipo? Si no, ¿cómo se planea medir este impacto en un futuro piloto?
2. **Validación de "Ground Truth" para Anomalías:** Para validar la precisión (N2, D1, "Precisión de Detección de Anomalías" del Anteproyecto) y la Tasa de Falsos Positivos/Negativos, ¿hay algún conjunto de datos históricos donde se hayan identificado y confirmado anomalías reales (fugas, fallas de medidor, etc.) con las que se pueda comparar la salida del modelo? Si no, ¿cómo se planea abordar esta validación? (¿Quizás con simulación o con la validación de expertos sobre las alertas generadas por el prototipo durante un periodo de prueba?)
3. **Calibración Fina de contamination y Umbrales de Severidad:** ¿Cómo se determinará el valor óptimo de contamination para Isolation Forest en el contexto de Contugas? ¿Cómo se validará que los umbrales de score para "Alta", "Media", "Baja" severidad realmente reflejan la criticidad operativa para Contugas? (Esto usualmente requiere un proceso iterativo con los expertos del negocio).
4. **Impacto Financiero Específico:** El Anteproyecto menciona la reducción de costos operativos. Para la presentación final, ¿hay alguna estimación, aunque sea aproximada, del ahorro que podría generar el sistema (ej. por reducción de X% de gas no facturado)?