

Guardian de fraudes



Manual de usuario

By Juan Sebastian Perdomo Becerra



1 ¿Qué es y qué hace el sistema?

Guardián de Fraudes es un sistema interactivo diseñado para ayudar a empresas distribuidoras de gas industrial a detectar comportamientos anómalos en el consumo de sus clientes. El sistema analiza los datos históricos de presión, temperatura y volumen para evaluar si un registro individual o una tendencia de consumo se sale de lo esperado.

Está dividido en dos componentes principales:

Eje 1 – Detección de anomalías

Evalúa cada registro de forma puntual. Aplica modelos de machine learning y lógica física para decir si un valor es normal o sospechoso. Se usa principalmente para revisión rápida o para cargas diarias de operación.

- Usa modelos como Isolation Forest, Autoencoder y Z-Score físico.
- También incorpora variables derivadas como PV/T, PV/ZT, etc.
- Tiene una lógica de votación ponderada entre los modelos para determinar si un dato es sospechoso.

Eje 2 – Predicción por cliente

Este eje trabaja más como un sistema de pronóstico. A partir de los datos históricos de cada cliente, genera predicciones a corto plazo (24, 48 o 72 horas) y detecta si lo que realmente sucedió se aleja de esa predicción.

- Aplica modelos por cliente: Prophet y NeuralProphet.
- Si el cliente no tiene suficiente histórico, usa un modelo general.
- Marca como sospechoso si tanto Prophet como NeuralProphet fallan por encima de cierto umbral.

Ventajas del sistema

- Análisis automático y en tiempo real al cargar nuevos datos.

- Detección adaptativa por cliente con modelos entrenados a partir de su comportamiento histórico.
- Visualización clara de resultados y variables físicas derivadas.
- Diseño web sin instalación, accesible desde cualquier navegador.

Limitaciones actuales

- Se requiere mínimo 48 registros por cliente para poder entrenar modelos personalizados (de lo contrario se usa el modelo general).
- El sistema no guarda los datos que se cargan; la información es evaluada pero no almacenada de forma persistente.
- No se puede editar información desde el dashboard. Si hay un error, se debe corregir el archivo .csv y volver a cargar.

Advertencias importantes

- Los datos deben estar completos, sin valores vacíos ni variables con nombres incorrectos.
- Para el Eje 2, los modelos predicen índices físicos, no directamente la presión o temperatura, así que deben estar calculados correctamente o el sistema los derivará automáticamente.
- Se recomienda trabajar con datos horarios y homogéneos, ya que el rendimiento del modelo puede bajar con datos irregulares.

2 ¿Qué necesita el usuario para usar el sistema?

El sistema Guardián de Fraudes está diseñado para ser ejecutado directamente desde el navegador, sin necesidad de instalación local ni configuraciones adicionales. Basta con ingresar al siguiente enlace desde cualquier equipo con conexión a internet:

<http://35.224.250.143:8502/>

Una vez dentro, el usuario se encuentra con una interfaz intuitiva dividida por secciones, cada una correspondiente a los ejes del modelo. Aunque no es un sistema difícil de usar, es importante aclarar que no está hecho para un público general, sino para personal técnico o usuarios con experiencia en análisis de datos operativos y fundamentos de física aplicada a sistemas de gas.

Para operar correctamente el sistema, el usuario debe tener conocimientos básicos en variables físicas como presión absoluta, volumen y temperatura en Kelvin, y comprender cómo se combinan estas variables en relaciones como **PV/T** (presión por volumen sobre temperatura), que proviene directamente de la ley de gases ideales. Así mismo, el sistema hace uso de otros índices físicos como **PV/ZT** (donde Z es el factor de compresibilidad), **V/T** y **P/V**, todos ellos utilizados por los modelos para detectar comportamientos anómalos o inesperados. Por eso, aunque el sistema calcula automáticamente estas variables, el usuario debe ser capaz de interpretarlas con criterio técnico.

El prototipo requiere como entrada archivos en formato .csv, los cuales deben contener las siguientes columnas con nombres exactos: `cliente`, `fecha`, `presion_abs`, `temp_k` y

volumen. El formato de la fecha debe ser claro y consistente (ejemplo: 2025-05-01 06:00:00) y los valores numéricos no deben tener vacíos ni datos fuera de rango. En caso de errores, el sistema no continuará con el análisis y simplemente notificará que hay un fallo en la entrada. No hay aún validación automatizada del contenido semántico del archivo, así que esa revisión debe hacerla el usuario antes de cargar los datos.

Desde el punto de vista del conocimiento funcional, se espera que quien use el sistema pueda distinguir entre una predicción y un valor real, interpretar una gráfica básica de comportamiento (líneas y puntos), y entender el concepto de desviación respecto a un modelo esperado. Esto es clave especialmente en el Eje 2, donde los modelos Prophet y NeuralProphet generan predicciones de los índices físicos para las siguientes 24, 48 o 72 horas, y si el comportamiento real se aleja mucho de esa predicción, el sistema genera una alerta.

Finalmente, no se necesita ningún conocimiento de programación, pero sí cierta familiaridad con el análisis de registros operativos y con el uso de herramientas como Excel, Google Sheets o editores de texto para revisar y preparar archivos .csv.

3 Casos de uso y paso a paso

El sistema Guardián de Fraudes tiene una estructura clara de navegación: en el menú lateral izquierdo se encuentran las distintas páginas disponibles. Antes de entrar a los modelos como tal, el usuario puede acceder a dos pestañas enfocadas en **análisis descriptivo de las variables**, que ayudan a entender el comportamiento general del sistema y de los clientes antes de aplicar los modelos. Estas son:

- **Resumen operativo:** muestra un panorama general del comportamiento histórico de los clientes. Incluye métricas agregadas, gráficos de dispersión y evolución por cliente, así como los principales patrones detectados en el sistema.
- **Análisis por cliente:** permite seleccionar un cliente específico y ver su evolución histórica de consumo, sus variables físicas derivadas y los puntos en que se han registrado posibles comportamientos atípicos. Esta sección es útil para el diagnóstico previo antes de pasar al análisis con modelos predictivos o de anomalía.

Una vez entendido el contexto general con estas dos pestañas iniciales, el usuario puede pasar a los dos ejes principales del sistema:

1. Análisis de anomalías individuales (Eje 1)

Este caso de uso aplica cuando se quiere revisar si un conjunto de registros (ya sean recientes o históricos) tienen comportamientos que podrían indicar un posible error de lectura, una alteración en la operación o un comportamiento sospechoso por parte del cliente.

El usuario debe ingresar al sistema, ir a la pestaña del menú lateral llamada **“Eje 1 – Detección de Anomalías”** y seleccionar la forma de carga de los datos. Puede hacerlo de dos formas:

2. **Ingreso manual:** útil para revisar casos puntuales. Se digitan a mano los valores de presión absoluta, temperatura en Kelvin y volumen consumido. El sistema deriva automáticamente las variables físicas e informa si el registro es sospechoso o no, mostrando también los resultados de cada modelo por separado.
3. **Carga masiva mediante archivo .csv:** recomendada para análisis por lotes. El archivo debe tener las columnas cliente, fecha, presion_abs, temp_k y volumen. Tras la carga, el sistema evalúa cada fila, muestra una tabla con los resultados y genera visualizaciones asociadas como gráficos de dispersión, evolución temporal y top de clientes con más anomalías.

En ambos casos, se puede especificar si el cliente ya tiene un modelo personalizado de aislamiento entrenado o si es un cliente nuevo (para usar el modelo general). El resultado final siempre se resume como “Normal” o “Sospechoso”, acompañado de gráficos e indicadores que ayudan a interpretar por qué fue clasificado así.

4. Predicción por cliente y alertas por desviación (Eje 2)

Este caso de uso aplica cuando se desea evaluar el comportamiento esperado de un cliente a partir de sus propios datos históricos. Es útil para anticipar irregularidades, validar tendencias o generar una alerta si el cliente empieza a salirse de su patrón usual.

El usuario debe entrar a la sección **“Eje 2 – Predicción por cliente”**. Allí se le solicitará un archivo .csv con estructura similar al del eje anterior (fecha, cliente, presion_abs, temp_k, volumen). El sistema derivará internamente las variables físicas que usa como objetivo.

Después de cargar el archivo, el sistema solicita tres decisiones:

- Cliente a evaluar: si hay varios en el archivo, se puede seleccionar cuál analizar.
- Variable física a predecir: puede ser pv_t, pv_zt, v_t, p_v, o todas al tiempo.
- Horizonte de predicción: se puede elegir entre 24, 48 o 72 pasos hacia adelante, dependiendo de la profundidad deseada.

El sistema usa modelos previamente entrenados para ese cliente (Prophet y NeuralProphet). Si el cliente tiene suficiente histórico, se usa su modelo personalizado. Si no, se recurre al modelo general. Una vez hecha la predicción, se muestra:

- La curva real vs. predicha por cada modelo.
- Métricas de error: MAE y RMSE para Prophet y NeuralProphet.
- Y lo más importante: si ambos modelos se equivocan más allá de un umbral, el sistema marca esa predicción como **“Sospechosa por desviación”**.

Este enfoque permite que el usuario no solo vea lo que el modelo predijo, sino que también reciba una alerta automatizada si el cliente está teniendo un comportamiento inesperado según ambos modelos.

4 Advertencias, recomendaciones y limitaciones

Aunque el sistema está diseñado para funcionar de forma estable y ofrecer resultados interpretables por parte del usuario, es importante dejar claras algunas condiciones y recomendaciones para evitar errores comunes y aprovechar bien el análisis.

El primer punto clave es que los datos de entrada deben estar limpios, completos y correctamente estructurados. Tanto en el Eje 1 como en el Eje 2, el sistema espera archivos .csv con columnas específicas: cliente, fecha, presion_abs, temp_k, volumen. Si alguno de estos campos está mal nombrado, contiene valores faltantes o formatos incorrectos (por ejemplo, fechas sin hora o texto en columnas numéricas), el sistema no podrá procesarlo y mostrará un mensaje de error genérico. En este momento no hay validación automática del contenido ni corrección de errores, por lo que es responsabilidad del usuario verificar los datos antes de cargarlos.

También es importante mencionar que el sistema no guarda ni almacena los datos cargados. Cada análisis es independiente, y si se desea revisar un resultado anterior, se debe volver a cargar el archivo. Esta decisión se tomó para garantizar simplicidad en el prototipo, pero implica que el usuario debe tener un control local de sus datos de entrada y salida.

En cuanto al Eje 2, hay que entender que los modelos predictivos trabajan con variables derivadas de la ley de los gases, no con las variables originales (presión, temperatura, volumen). Esto significa que, aunque el usuario cargue datos crudos, el sistema predice cosas como PV/T o P/V , lo cual puede no ser tan intuitivo al principio si no se está familiarizado con este tipo de indicadores. Es por eso por lo que se recomienda que los usuarios tengan un mínimo conocimiento de cómo se interpretan estas relaciones físicas.

Por último, vale la pena recordar que este es un prototipo funcional, aún en fase de validación. No está conectado a una base de datos productiva, no tiene módulos de edición o auditoría de usuarios, y su interfaz puede seguir evolucionando. Aun así, es una herramienta útil y poderosa para análisis técnico-operativo, siempre y cuando se use con criterio y se respeten los lineamientos descritos aquí.

5 Conclusiones y consideraciones

El sistema Guardián de Fraudes permite realizar análisis de datos operativos de gas desde dos frentes: detección puntual de registros anómalos y predicción del comportamiento esperado por cliente. Su enfoque combina modelos de machine learning con variables físicas derivadas, lo que le da un equilibrio entre rigor técnico y aplicabilidad práctica en campo.

En la práctica, el sistema ha demostrado que es capaz de identificar desviaciones importantes tanto en registros individuales como en comportamientos que se salen del patrón histórico. El uso combinado de modelos por cliente y modelos generales permite adaptarse a distintos niveles de disponibilidad de datos, y la inclusión de un mecanismo de alerta por error conjunto fortalece su capacidad de anticiparse a eventos inusuales.

En términos de experiencia de uso, el sistema no requiere instalación, pero sí exige cuidado con la calidad de los datos y un mínimo de comprensión sobre las variables físicas involucradas. En su estado actual como prototipo funcional, ha sido probado con distintos escenarios y ha entregado resultados consistentes, tanto en entornos de prueba como con datos reales de clientes simulados.

A medida que se continúe con el desarrollo, hay espacio para integrar nuevas funcionalidades (como validación automática, almacenamiento o trazabilidad), pero incluso en su estado actual, el Guardián de Fraudes ya ofrece un soporte valioso para decisiones técnicas y operativas.

Manual de usuario

By Juan Sebastian Perdomo Becerra

