

# **Anteproyecto**

## **Sistema Analítico de Monitoreo y Detección de Anomalías en el Consumo de Gas**

### **Integrantes:**

Diana Liceth Bayona  
Juan Diego Pérez  
Jorge Andrés Rodríguez  
Santiago Gutiérrez

**Resumen:** Este trabajo propone un marco para el desarrollo de un sistema analítico avanzado orientado a la detección de anomalías en el consumo de gas de clientes industriales de Contugas. Basado en técnicas de análisis de datos, el enfoque sugiere el uso de modelos de series temporales (ARIMA y LSTM), clustering no supervisado (K-Means y DBSCAN) y detección de outliers (Isolation Forest) adaptados a datos operativos registrados por hora, como presión, temperatura y volumen. El proyecto tiene como objetivo plantear soluciones que optimicen el monitoreo en tiempo real, reduzcan costos operativos, mejoren la seguridad de la red y aumenten la eficiencia en la toma de decisiones. Durante su desarrollo, se abordaron fases como la planificación, análisis y preprocesamiento de datos, así como la evaluación de riesgos mediante la metodología RBS y el planteamiento de métricas clave para medir el desempeño de las posibles soluciones. Este trabajo establece una base teórica y metodológica para futuras implementaciones en Contugas, proporcionando lineamientos sólidos para el desarrollo de sistemas que modernicen la gestión energética y fomenten la eficiencia operativa de manera sostenible y escalable.

## Contenido

|                               |    |
|-------------------------------|----|
| Introducción .....            | 3  |
| Estado del Arte.....          | 4  |
| Desarrollo del proyecto ..... | 4  |
| 1. Inicialización.....        | 5  |
| 2. Planificación.....         | 5  |
| 3. Ejecución .....            | 7  |
| 4. Monitoreo y Control.....   | 8  |
| 5. Cierre .....               | 9  |
| Referencias .....             | 10 |
| Anexos .....                  | 10 |

## Introducción

El Grupo Energía Bogotá (GEB), a través de su filial Contugas, ha identificado la necesidad de implementar soluciones avanzadas de analítica de datos para modernizar sus operaciones y mejorar la eficiencia en la detección de anomalías en el consumo de gas de sus clientes industriales. Este proyecto surge como respuesta a desafíos operativos clave, como la identificación de patrones atípicos en variables críticas de la red de distribución, incluyendo presión, temperatura y volumen de gas.

El objetivo principal es desarrollar un sistema analítico que permita detectar y gestionar anomalías en tiempo real, optimizando el monitoreo de la red de distribución, reduciendo costos operativos y garantizando un servicio confiable. Mediante la implementación de un producto mínimo viable (PMV), se busca integrar tecnologías innovadoras para visualizar datos históricos, generar resúmenes descriptivos y emitir alertas sobre posibles irregularidades, asegurando un impacto positivo tanto en la rentabilidad como en la satisfacción del cliente.

El proyecto tiene como marco inicial los elementos definidos en la carta del proyecto, incluyendo los beneficios esperados, como la reducción de falsos positivos, la disminución de tiempos de respuesta y la mejora en la precisión de las detecciones. Además, se establecen métricas clave de negocio, como la precisión de las alertas generadas, el tiempo promedio de detección de anomalías y la reducción de costos asociados con el mantenimiento y el monitoreo manual. Estas métricas guiarán el seguimiento del desempeño del proyecto y su alineación con los objetivos estratégicos de Contugas.

Este documento también detalla la estructura del proyecto, relacionando los resultados obtenidos en las Guías 1, 2 y 3, las cuales sirvieron como base para definir la problemática, analizar los riesgos y preparar los datos. En la Guía 1, se llevó a cabo un análisis inicial del problema, definiendo como objetivo principal la detección automática de anomalías para optimizar los recursos y mejorar la seguridad de la red. En este análisis se establecieron criterios de éxito, como la reducción de falsos positivos y el tiempo de respuesta en la identificación de eventos críticos. En la Guía 2, se profundizó en la calidad de los datos disponibles, identificando problemas de consistencia y brechas temporales en algunos registros, los cuales fueron corregidos mediante interpolación lineal para garantizar la continuidad temporal. Finalmente, en la Guía 3, se realizó un análisis detallado de riesgos mediante la técnica RBS, identificando factores críticos como la calidad de los datos, la rotación de personal técnico y los costos asociados a la implementación de nuevas tecnologías, y se plantearon estrategias para mitigar estos riesgos.

El proyecto se organiza en cinco fases principales: inicialización, planificación, ejecución, monitoreo y control, y cierre. Cada fase aborda aspectos clave del desarrollo, desde la definición de la problemática de negocio y la propuesta de técnicas analíticas hasta la estructuración de acciones para garantizar el éxito del sistema en el entorno operativo de Contugas. El documento detalla los hallazgos más relevantes, las decisiones estratégicas tomadas durante cada fase y cómo estas contribuyen a los objetivos generales del proyecto, posicionándolo como una solución integral para modernizar la operación y gestión del consumo de gas.

## **Estado del Arte**

Maciá Fiteni, À. (2023), en su trabajo Smart University: Detección de Anomalías en Consumo Energético, se enfoca en la identificación de patrones de uso y anomalías en el consumo energético de universidades. Utilizando redes neuronales, analiza datos como el consumo energético, conexión wifi y variables temporales (día y hora), lo que permite detectar irregularidades en tiempo real. Este proyecto no solo contribuye al ahorro energético, sino que también fomenta la sostenibilidad y la gestión responsable en instituciones educativas. La implementación se realizó mediante la metodología ágil Scrum, optimizando el proceso y garantizando adaptabilidad a los objetivos planteados.

Por otro lado, Baldacci, L., Golfarelli, M., Lombardi, D., y Sami, F. (2016), en su trabajo Natural Gas Consumption Forecasting for Anomaly Detection, abordan la previsión del consumo de gas natural y la detección de anomalías en redes de distribución. Su investigación utiliza datos históricos como flujo de gas, temperatura exterior y marcas de tiempo, combinando dos métodos predictivos: vecinos más cercanos y análisis de regresión local. Los resultados demostraron una alta capacidad para prever consumos y detectar flujos anómalos, mejorando la eficiencia en las operaciones y facilitando la toma de decisiones en tiempo real dentro de una sala de control. Este enfoque fue desarrollado en colaboración con HERA S.p.A., un proveedor líder de gas en Italia, y destaca por su contribución a la seguridad y optimización de la red de distribución.

En este punto, ambos trabajos aportan herramientas valiosas al proyecto de Contugas. El enfoque de Maciá Fiteni en el uso de redes neuronales para analizar grandes volúmenes de datos operativos y detectar anomalías de forma automática puede ser adaptado para optimizar la gestión de la red de gas. Su metodología de implementación ágil también ofrece una guía útil para estructurar las fases del proyecto. Por su parte, el trabajo de Baldacci et al. proporciona métodos concretos para la previsión del consumo de gas y la identificación de flujos anómalos, los cuales se alinean con los objetivos de Contugas de mejorar la seguridad y la eficiencia operativa en tiempo real. La combinación de estos enfoques permite construir un sistema robusto que aprovecha lo mejor de las redes neuronales y los modelos predictivos tradicionales.

En conclusión, las investigaciones de Maciá Fiteni y Baldacci et al. destacan por sus contribuciones complementarias al desarrollo de un sistema analítico de monitoreo para Contugas. Mientras que el primero aporta innovación tecnológica en la automatización y sostenibilidad, el segundo refuerza el enfoque predictivo basado en datos históricos, ambos esenciales para la optimización del monitoreo y detección de anomalías en el consumo de gas. Este marco integral no solo garantiza un servicio más eficiente y confiable, sino que también posiciona a Contugas como líder en innovación dentro del sector energético.

## **Desarrollo del proyecto**

El proyecto de monitoreo y detección de anomalías en el consumo de gas en Contugas tiene como objetivo principal optimizar la operación de la red de distribución de gas mediante el uso de analítica avanzada. Este sistema busca identificar patrones anómalos en las variables operacionales (presión, temperatura y volumen) en tiempo real, alertando a los operadores sobre posibles irregularidades. Con un presupuesto estimado de \$50,000 USD y un periodo de implementación de seis meses, el proyecto se estructura bajo principios de sostenibilidad

y eficiencia operativa, integrando herramientas tecnológicas innovadoras y una gestión efectiva de los datos

## **1. Inicialización**

La fase de inicialización del proyecto establece las bases necesarias para garantizar que todas las acciones y decisiones estén alineadas con los objetivos estratégicos de Contugas. En esta etapa, se definieron los criterios de éxito en las áreas tecnológicas y financieras, asegurando que los resultados esperados respondan tanto a las necesidades operativas como a las restricciones presupuestarias de la empresa.

Desde el punto de vista tecnológico, el éxito del proyecto se mide en función de la capacidad del sistema para monitorear en tiempo real variables críticas como presión, temperatura y volumen de gas, generar alertas automáticas con alta precisión y optimizar las operaciones al reducir la dependencia de procesos manuales. En el ámbito financiero, el proyecto busca minimizar los costos operativos asociados a la identificación manual de anomalías y maximizar el retorno de la inversión mediante una solución escalable y sostenible.

Un elemento clave de esta fase fue el desarrollo de un primer borrador de planificación, que establece las actividades principales, los recursos necesarios y los tiempos estimados. Este documento preliminar sirve como punto de partida para las fases posteriores, proporcionando una visión inicial de cómo se estructurará el proyecto y permitiendo la identificación temprana de posibles ajustes.

La carta del proyecto también fue elaborada durante esta etapa, definiendo el objetivo principal de diseñar un sistema analítico que permita detectar anomalías en el consumo de gas de manera eficiente. En ella, se estableció un alcance claro enfocado en los datos operativos registrados por Contugas, un presupuesto inicial aproximado de \$50,000 USD y una duración estimada de seis meses. Este monto incluye los costos asociados al Product Owner (PO), responsable de la gestión del proyecto; el equipo de analítica encargado del preprocesamiento y transformación de datos (ETL); el científico de datos que desarrollará los modelos analíticos, y el ingeniero que eventualmente desplegará la solución en el entorno operativo. Este documento sirve como guía para todas las actividades y asegura que las expectativas estén claramente definidas.

Los resultados de la Guía 1 jugaron un papel fundamental en esta fase al proporcionar el contexto inicial de la problemática. Esta guía permitió identificar necesidades clave de negocio, como la reducción de falsos positivos en las alertas y la mejora en el tiempo de detección de anomalías. Además, ayudó a establecer las primeras métricas de éxito, que serán esenciales para medir el desempeño del sistema a medida que avance el proyecto.

La fase de inicialización logró alinear los objetivos del proyecto con las necesidades y expectativas de Contugas. Al definir claramente los criterios de éxito, los recursos y el alcance, esta etapa proporciona una base sólida que permite avanzar hacia las siguientes fases con claridad y confianza.

## **2. Planificación**

En esta fase, se definió cómo abordar la problemática de negocio desde un enfoque de analítica, estableciendo las bases para la implementación del sistema de detección de anomalías en Contugas. El objetivo fue traducir los retos operativos de la empresa en un

problema de analítica bien estructurado, identificar los recursos necesarios y planificar los pasos a seguir para garantizar el éxito del proyecto.

El primer paso fue comprender la problemática de negocio, centrada en la necesidad de detectar anomalías en el consumo de gas de los clientes industriales de manera automatizada, mejorando la eficiencia operativa y reduciendo la dependencia de intervenciones manuales. Este desafío se transformó en un problema de analítica que implica modelar series temporales de las variables críticas (presión, temperatura y volumen) y detectar patrones atípicos mediante técnicas avanzadas. La solución propuesta también debía incluir la capacidad de predecir comportamientos futuros para anticipar problemas y tomar medidas preventivas.

Un aspecto clave de la planificación fue el análisis de riesgos, llevado a cabo mediante la metodología RBS (Risk Breakdown Structure). Este análisis identificó posibles desafíos en áreas como la calidad de los datos, la disponibilidad de recursos tecnológicos y la gestión del cambio en la organización. Entre los riesgos destacados se incluyeron la baja calidad de los datos históricos, la falta de personal capacitado y las limitaciones presupuestarias. Para mitigar estos riesgos, se diseñaron estrategias específicas, como la capacitación del equipo, la validación de los datos y la asignación de un presupuesto flexible para cubrir contingencias.

El cronograma del proyecto se estructuró en cuatro fases principales: análisis de requisitos (1 mes), diseño y desarrollo del sistema (3 meses), pruebas y validación (2 meses) y capacitación e implementación (1 mes). Este plan detallado permite una gestión eficiente del tiempo y asegura que las actividades clave se completen dentro del plazo establecido. Además, se definieron hitos específicos para cada etapa, facilitando el monitoreo del progreso del proyecto.

El análisis de datos desempeñó un papel central en la planificación, comenzando con una evaluación de la calidad del conjunto de datos disponibles. Se revisaron más de 847,000 registros históricos que incluyen mediciones de presión, temperatura y volumen de gas. Durante este proceso, se identificaron problemas de consistencia y secuencialidad en algunos clientes, los cuales se corrigieron mediante interpolación lineal para garantizar la continuidad temporal. Este preprocesamiento aseguró que los datos fueran adecuados para los análisis avanzados requeridos por el proyecto.

En el análisis inicial, la base de datos mostró un alto nivel de calidad, con la mayoría de las variables libres de valores nulos y duplicados. Sin embargo, durante la validación, se detectaron brechas temporales significativas en los registros de dos clientes específicos, correspondientes a fechas faltantes en las mediciones. Estas brechas comprometían la completitud y consistencia temporal de los datos, afectando la continuidad necesaria para análisis confiables.

Por otro lado, se verificó la consistencia de las relaciones entre variables, observándose una correlación negativa moderada entre el volumen y la presión (-0.30), lo cual es coherente con las leyes de los gases. No se identificaron relaciones incoherentes entre las variables. En cuanto a la temporalidad, se detectó que la última fecha registrada corresponde al 31 de diciembre de 2023, lo que evidencia una falta de actualización que podría impactar futuros análisis. Finalmente, el formato de los datos es adecuado, con tipos correctamente definidos y una estructura válida en la columna de fecha.

En general, los datos son fiables y están bien estructurados. Sin embargo, será necesario actualizar los registros temporales si se requieren análisis prospectivos y atender las brechas identificadas en los datos de ciertos clientes para mejorar su calidad y consistencia. Por lo cual, estos hallazgos son relevantes para el negocio, ya que los datos procesados son adecuados para construir modelos predictivos que detecten anomalías y optimicen el monitoreo del sistema de distribución de gas. Sin embargo, la desactualización de los registros (hasta diciembre de 2023) destaca la necesidad de integrar datos recientes para reflejar condiciones operativas actuales y mejorar la precisión del análisis y la toma de decisiones.

Por último, se establecieron las fuentes de datos adicionales que podrían enriquecer el análisis, como registros de mantenimiento de equipos y datos climáticos históricos. Estas fuentes complementarias permiten contextualizar mejor las anomalías detectadas, mejorando la precisión de los modelos analíticos. También se identificaron las transformaciones necesarias en los datos, como la normalización y sincronización temporal, para garantizar que sean útiles y coherentes con la problemática a resolver.

La fase de planificación permitió establecer un puente claro entre los desafíos operativos de Contugas y las soluciones analíticas propuestas. Al abordar la problemática desde un enfoque estructurado, identificar los riesgos y definir un plan detallado de actividades, esta etapa asegura que el proyecto avance sobre una base sólida, minimizando incertidumbres y maximizando las posibilidades de éxito.

### **3. Ejecución**

En esta fase, se establecen las técnicas de analítica más adecuadas para abordar el problema planteado, teniendo en cuenta las características de los datos y las necesidades operativas de Contugas. Aunque los modelos no se implementan ni despliegan aún, se identifican y justifican las metodologías que servirán como base para el desarrollo de la solución.

Se seleccionaron técnicas avanzadas de modelado de series temporales para capturar patrones en las variables clave como presión, temperatura y volumen. Entre ellas, se destaca el modelo ARIMA, que es adecuado para identificar tendencias lineales y estacionales en datos históricos, permitiendo realizar predicciones basadas en comportamientos pasados. Adicionalmente, se identificó el modelo LSTM, una red neuronal recurrente especializada en datos secuenciales que puede captar relaciones no lineales y realizar predicciones a largo plazo. Estas técnicas son esenciales para anticipar posibles anomalías y generar alertas en tiempo real.

Para complementar el análisis, se incluyeron técnicas de clustering no supervisado que ayudan a descubrir patrones comunes entre los clientes y a identificar comportamientos atípicos. El método K-Means permite agrupar clientes según similitudes en su consumo, mientras que DBSCAN se enfoca en detectar puntos aislados y comportamientos que se desvíen de los clusters principales. Estas técnicas ofrecen una visión más profunda del comportamiento de los usuarios y permiten personalizar estrategias de monitoreo para diferentes perfiles de consumo.

Además, se seleccionaron modelos específicos para la detección de outliers, con el objetivo de identificar anomalías en datos multidimensionales. Isolation Forest es una técnica eficiente que identifica valores atípicos al aislar puntos de datos inusuales en el conjunto, mientras que

el análisis de valores atípicos multivariados permite detectar desviaciones significativas en la interacción de las variables clave. Estas técnicas son fundamentales para asegurar que se detecten tanto anomalías evidentes como aquellas más sutiles que puedan comprometer la operación de la red.

Las métricas definidas para evaluar el desempeño de los modelos jugarán un papel crucial en las etapas posteriores del proyecto. Entre las métricas clave se encuentran la precisión de las alertas, que mide la proporción de anomalías correctamente identificadas, y el tiempo promedio de detección, que refleja la velocidad de respuesta del sistema. Asimismo, se consideran la tasa de falsos positivos y negativos, indicadores que evalúan la confiabilidad del sistema al minimizar errores en la detección de anomalías. Estas métricas serán utilizadas para monitorear y ajustar continuamente los modelos, asegurando que cumplan con los estándares de calidad esperados.

La fase de ejecución sienta las bases técnicas para la implementación de la solución analítica, asegurando que las metodologías seleccionadas se alineen con los objetivos de negocio y las características de los datos. Este enfoque garantiza que el sistema de monitoreo de anomalías sea preciso, confiable y adaptable a las necesidades de Contugas, maximizando su impacto en la operación de la red de distribución.

## **4. Monitoreo y Control**

En esta fase, se diseñan acciones específicas para garantizar el seguimiento adecuado del proyecto, identificando y corrigiendo posibles desviaciones respecto a la planeación inicial. Además, se establecen mecanismos para evaluar los beneficios obtenidos mediante las técnicas implementadas y validar la solución final conforme a los objetivos planteados.

El monitoreo se centra en evaluar continuamente el desempeño de las actividades y los resultados generados por los modelos analíticos. Para ello, se implementarán indicadores clave de rendimiento (KPI) previamente definidos, como la precisión de las alertas, el tiempo de detección de anomalías y la tasa de falsos positivos y negativos. Estos indicadores permitirán medir la efectividad de las técnicas utilizadas, asegurando que el sistema funcione de manera óptima y se mantenga alineado con los objetivos operativos de Contugas.

Las acciones para identificar y controlar desviaciones incluyen la revisión periódica del cronograma y los entregables, garantizando que las actividades se completen dentro de los plazos establecidos. Si se detectan retrasos o inconsistencias, se tomarán medidas correctivas, como la reasignación de recursos, el ajuste de prioridades o la modificación de los modelos utilizados. Este enfoque iterativo permitirá ajustar el proyecto en tiempo real, minimizando el impacto de las desviaciones en el alcance y los resultados esperados.

Un componente clave de esta fase es la evaluación de los beneficios obtenidos mediante las técnicas analíticas. Para ello, se compararán las métricas generadas por los modelos con los objetivos definidos en la fase de planificación, como la reducción de costos operativos, la mejora en la seguridad de la red y el aumento en la satisfacción de los clientes. Esta evaluación permitirá validar si la solución analítica aporta valor tangible a la operación de Contugas y si cumple con los criterios de éxito previamente establecidos.

Finalmente, se plantean acciones específicas para validar el modelo y las técnicas implementadas. Esto incluye pruebas iterativas para ajustar los parámetros de los modelos, asegurando que sean robustos frente a diferentes escenarios operativos. Además, se



realizará una revisión exhaustiva de los resultados obtenidos en ambientes de prueba y producción, garantizando que no existan discrepancias significativas que puedan afectar la confiabilidad del sistema. Esta validación también incluirá la participación de los stakeholders clave, quienes verificarán que la solución cumple con sus expectativas y necesidades.

La fase de monitoreo y control es esencial para garantizar que el proyecto se mantenga en el camino correcto y entregue los resultados esperados. Al implementar un sistema de seguimiento continuo, evaluar los beneficios y validar los modelos de manera rigurosa, Contugas puede asegurar que la solución analítica no solo resuelva la problemática planteada, sino que también aporte valor significativo a su operación diaria.

## 5. Cierre

La fase de cierre del proyecto se enfoca en consolidar las actividades realizadas y presentar los resultados alcanzados hasta este punto, dejando sentadas las bases para la implementación futura de la solución analítica. En esta etapa, se organiza un resumen de los avances logrados, las técnicas propuestas y las acciones necesarias para continuar con la implementación del sistema de detección de anomalías en Contugas.

En resumen, el documento incluye una estructura clara con los siguientes rubros:

- **Propuesta técnica:** Se presenta un resumen detallado de los modelos analíticos propuestos, como ARIMA, LSTM, K-Means, DBSCAN e Isolation Forest. Este apartado explica cómo estas técnicas pueden abordar la problemática identificada, permitiendo detectar patrones atípicos y prever comportamientos en las variables operativas.
- **Resultados del análisis de datos:** Se detallan los hallazgos obtenidos durante la etapa de análisis y preprocesamiento de los datos, como la identificación de patrones iniciales, la calidad de los datos y las transformaciones realizadas para garantizar su consistencia.
- **Evaluación del potencial de las técnicas propuestas:** Se incluye un análisis teórico del impacto esperado de cada técnica en la resolución del problema, considerando métricas clave como precisión esperada, sensibilidad y tiempo de detección.

Adicionalmente, en esta fase se identifican las acciones necesarias para la implementación futura del sistema. Estas acciones incluyen la planificación de pruebas adicionales en ambientes controlados, la preparación de infraestructura tecnológica y la capacitación del equipo operativo. Aunque los modelos aún no han sido desplegados, esta etapa asegura que las condiciones para su implementación estén definidas y alineadas con los objetivos de Contugas. Finalmente, los resultados de esta fase se presentan a los stakeholders mediante un video.

La fase de cierre documenta los progresos realizados, estructura las opciones analíticas propuestas y establece las bases para la implementación futura del sistema. Este enfoque asegura que Contugas pueda avanzar de manera estratégica y organizada hacia la integración de soluciones avanzadas en su operación diaria.

## Referencias

- Baldacci, L., Golfarelli, M., Lombardi, D., & Sami, F. (2016). Natural gas consumption forecasting for anomaly detection. Expert Systems with Applications, 62, 190-201. <https://doi.org/10.1016/j.eswa.2016.06.013>
- Maciá Fiteni, À. (2023). Smart University: Detección de anomalías en consumo energético (Trabajo de grado). Repositorio de la Universidad de Alicante. [https://rua.ua.es/dspace/bitstream/10045/137670/1/Smart\\_University\\_Deteccion\\_de\\_Anomalias\\_en\\_Consumo\\_Energet\\_Macia\\_Fiteni\\_Alex.pdf](https://rua.ua.es/dspace/bitstream/10045/137670/1/Smart_University_Deteccion_de_Anomalias_en_Consumo_Energet_Macia_Fiteni_Alex.pdf)

## Anexos

- Repositorio Auditoria de Datos: [https://github.com/dianabayonap/Proyecto\\_G14.git](https://github.com/dianabayonap/Proyecto_G14.git)
- Guías 1, 2, 3, 4 y 5