

Anteproyecto - Guía 2 - Resolución Preguntas

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

Integrantes del Proyecto:

Diana Liceth Bayona
Juan Diego Pérez
Jorge Andrés Rodríguez
Santiago Gutiérrez

Parte 1. Del Qué del problema de negocio al Cómo del problema de Analytics

Fase a) Los encargados de la solución

1. ¿Qué resultados se esperan obtener tras la implementación del proyecto?

Se espera que la implementación de la solución de detección de anomalías en el consumo de gas proporcione resultados significativos en la optimización de la operación de Contugas. Esto incluye la identificación precisa de comportamientos atípicos en las variables operacionales (presión, temperatura y volumen de gas), lo que permitirá alertar a los operadores en tiempo real sobre irregularidades, mejorar la toma de decisiones basadas en datos históricos y tendencias, y **reducir los costos** asociados con la detección manual de problemas. Además, se anticipa un incremento en la seguridad de la red de distribución, una mayor eficiencia en la gestión de recursos, una mayor satisfacción de los clientes y un potencial retorno económico positivo al evitar pérdidas derivadas de fugas, fallas o abandono de clientes en el sistema.

2. ¿Quiénes participarán en la solución analítica?

La solución analítica requerirá la participación activa de un equipo multidisciplinario compuesto por el Gerente de Operaciones de Contugas, el equipo de Innovación de Contugas, la Universidad de los Andes, el equipo de analítica del Grupo Energía Bogotá (GEB), los proveedores de tecnología y el equipo de TI de Contugas. Cada uno de estos *stakeholders* aportará su experiencia y conocimientos específicos para asegurar que la solución sea efectiva y alineada con los objetivos operativos y estratégicos de la empresa.

3. ¿Qué cambios se generarían en la organización/unidad cliente como resultado del proyecto?

Como resultado del proyecto, se generarán cambios significativos en la estructura y funcionamiento de la unidad de operaciones de Contugas. La automatización en la detección de anomalías disminuirá la dependencia de la intervención manual, lo que permitirá a los equipos operativos concentrarse en tareas más estratégicas y menos

repetitivas. Esto conducirá a una mayor eficiencia operativa, mejorando la capacidad de respuesta ante situaciones anómalas. Además, la cultura organizacional evolucionará hacia un enfoque más basado en datos, promoviendo la modernización de procesos y la adopción de tecnologías avanzadas en el monitoreo y gestión del consumo de gas. Sin embargo, también se deberá gestionar la transición para aquellos empleados que realizaban la detección manual, para evitar resistencia al cambio.

Fase b) Lo que el cliente requiere

1. ¿Cuáles son los requerimientos básicos u obligatorios que se atienden con el proyecto?

Los requerimientos básicos u obligatorios del proyecto incluyen la capacidad de monitorear en tiempo real las variables operacionales clave (presión, temperatura y volumen de gas) y la generación de alertas ante la detección de anomalías. Es esencial que el sistema pueda visualizar datos históricos y proporcionar resúmenes descriptivos del comportamiento de consumo de cada cliente industrial. Además, el sistema debe ser accesible y fácil de usar para los operadores, asegurando que el personal pueda interpretar rápidamente las alertas y actuar en consecuencia.

2. ¿Cuáles son aquellos requerimientos que se consideran indiferentes en el producto del proyecto?

Los requerimientos considerados indiferentes son aquellos que no afectan significativamente la funcionalidad ni la experiencia del usuario final. Por ejemplo, la personalización estética de la interfaz de usuario o la inclusión de gráficos de visualización avanzados que no impacten directamente en la detección de anomalías pueden considerarse indiferentes. Aunque pueden mejorar la presentación de datos, no son críticos para el funcionamiento esencial del sistema ni para el cumplimiento de los objetivos del proyecto.

3. ¿Cuáles son aquellos requerimientos considerados inversos?

Los requerimientos inversos son aquellos que, si se implementan, afectarían negativamente la satisfacción del cliente o la eficiencia del sistema. En este caso, un requerimiento inverso podría ser la implementación de un proceso de validación manual antes de generar alertas automáticas. Esto podría ralentizar la respuesta ante anomalías y disminuir la efectividad del sistema, ya que la automatización es una de las principales ventajas buscadas por Contugas para optimizar la detección de irregularidades.

4. ¿Cuáles son los requerimientos unidimensionales?

Los requerimientos unidimensionales son aquellos que generan satisfacción proporcional a su nivel de cumplimiento. En este contexto, un requerimiento

unidimensional podría ser la precisión de las alertas generadas por el sistema. A mayor precisión en la detección de anomalías, mayor será la satisfacción de los usuarios, ya que podrán confiar en la información proporcionada para tomar decisiones informadas y oportunas, mejorando así la gestión del consumo de gas.

5. ¿Qué requerimiento son atractivos?

Los requerimientos atractivos son aquellos que, si se cumplen, superan las expectativas del cliente y brindan un valor agregado significativo. En este caso, un requerimiento atractivo podría ser la capacidad del sistema para aprender de los patrones históricos de consumo y ajustar sus alertas en función de la evolución del comportamiento del cliente. Esto no solo mejoraría la precisión en la detección de anomalías, sino que también podría ofrecer recomendaciones proactivas para optimizar el consumo, brindando un valor adicional a Contugas y sus clientes industriales.

Fase c) Los requerimientos finales

1. ¿Cuáles son los requerimientos? Enlista acorde su clasificación.

Requerimientos obligatorios

- Monitoreo en tiempo real de presión, temperatura y volumen de gas.
- Generación de alertas ante anomalías.
- Visualización de datos históricos de consumo.
- Resúmenes descriptivos del comportamiento de cada cliente.

Requerimientos indiferentes

- Personalización estética de la interfaz de usuario.
- Inclusión de gráficos avanzados que no impacten la detección.

Requerimientos inversos

- Proceso de validación manual de alertas.

Requerimientos unidimensionales

- Precisión de las alertas generadas.

Requerimientos atractivos

- Capacidad del sistema para aprender y ajustar alertas basadas en patrones históricos.

2. ¿Quiénes o qué áreas deberán participar en la solución analítica?

La solución analítica requerirá la participación de varias áreas y equipos dentro de la organización. Esto incluye al Gerente de Operaciones de Contugas, el equipo de Innovación de Contugas, la Universidad de los Andes para el soporte académico y técnico, el equipo de analítica de GEB, así como los proveedores de tecnología y el

equipo de TI de Contugas. La colaboración entre estos grupos será fundamental para asegurar el éxito del desarrollo y la implementación del sistema de detección de anomalías.

3. ¿Cómo la implementación de estos requerimientos impactaría en los objetivos del proyecto?

Requerimientos a Incrementar			
Nombre y Descripción	Escala de viabilidad de desarrollo e implementación (1-10)	Escala de impacto en el objetivo del proyecto (1-10)	Score Total
Monitoreo en tiempo real de presión, temperatura y volumen de gas.	9	10	90
Generación de alertas ante anomalías.	8	9	72
Visualización de datos históricos de consumo.	7	8	56
Resúmenes descriptivos del comportamiento de cada cliente.	6	7	42
Capacidad del sistema para aprender y ajustar alertas basadas en patrones históricos.	5	8	40
Requerimientos a aminorar			
Nombre y Descripción	Escala de viabilidad en la reducción de desarrollo e implementación (1-10)	Escala de impacto en el objetivo del proyecto (1-10)	Score Total
Proceso de validación manual de alertas.	8	4	32
Personalización estética de la interfaz de usuario.	9	3	27
Inclusión de gráficos avanzados que no impacten la detección.	7	2	14

La implementación de los requerimientos a incrementar se alinearán directamente con los objetivos del proyecto, maximizando la efectividad de la detección de anomalías y optimizando la gestión del consumo de gas en tiempo real. Por otro lado, la reducción de los requerimientos que se consideran menos críticos permitirá liberar recursos para concentrarse en aquellos elementos que realmente impulsan el valor del proyecto, garantizando un desarrollo más ágil y enfocado.

4. **Prioriza los requerimientos generando un orden o decisión de aquellos que van a ser incorporados al proyecto de acuerdo con el criterio de tu equipo. Considera que la escala de desarrollo en la implementación o su contrapartida puede también involucrar aspectos tecnológicos, monetarios o de otra índole que se deberían tener en cuenta en el desarrollo. Estos requerimientos se conocen también como los Key-Drives del proyecto. Ten presente que regularmente, estas evaluaciones se validan con el equipo y los SH. Y se ajustan de acuerdo con ellos.**

Priorización de requerimientos					
Prioridad	Nombre y Descripción	Escala de viabilidad de desarrollo e implementación (1-10)	Escala de impacto en el objetivo del proyecto (1-10)	Score Total	Justificación
1	Monitoreo en tiempo real de presión, temperatura y volumen de gas	9	10	90	Fundamental para la detección oportuna de anomalías y optimización del consumo. Se alinea con los objetivos estratégicos de Contugas.
2	Generación de alertas ante anomalías	8	9	72	Permite una respuesta rápida ante irregularidades, aumentando la seguridad y eficiencia operativa.
3	Visualización de datos históricos de consumo	7	8	56	Ayuda en el análisis de patrones y tendencias, apoyando la toma de decisiones informadas.

4	Resúmenes descriptivos del comportamiento de cada cliente.	6	7	42	Facilita la comprensión del consumo individual y la identificación de clientes con comportamiento atípico.
5	Capacidad del sistema para aprender y ajustar alertas basadas en patrones históricos.	5	8	40	Aumenta la efectividad del sistema a largo plazo, pero requiere inversión en tecnología y análisis de datos.

Observaciones sobre la Priorización

- **Inversión en Tecnología y Recursos:** Los requerimientos de mayor prioridad son aquellos que ofrecen un impacto inmediato y significativo en la operación, lo que justifica la inversión inicial necesaria en tecnología y recursos humanos.
- **Validación Continua:** Esta priorización debe ser revisada y validada con el equipo y los stakeholders para garantizar que todos los aspectos relevantes y las expectativas sean considerados antes de la implementación.
- **Adaptabilidad:** Es importante mantener flexibilidad en la priorización, ya que puede ser necesario ajustar los requerimientos en función de la retroalimentación de los stakeholders o cambios en el contexto del proyecto.

5. Valida, a partir de estos requerimientos, las métricas iniciales de éxito y fracaso identificadas en la GUIA 1. Es decir, a partir del problema de negocio y la métrica de negocio inicial establecida en la GUIA 1, descompón esa métrica en subpartes asignables a cada requerimiento seleccionado en el paso anterior. Cuando se cuenta con ellos, este paso debe ser negociado y revisado junto con los SH, para que las métricas sean posteriormente conocidas, medibles y trazables.

A partir de los requerimientos seleccionados en la priorización y las métricas iniciales de éxito identificadas en la GUIA 1, se procede a descomponer cada métrica en subpartes asignables a los requerimientos específicos. Esta descomposición permitirá

una mejor trazabilidad y negociación con los stakeholders (SH) para asegurar que las métricas sean concretas, medibles y relevantes para el éxito del proyecto.

Métricas Iniciales y su Descomposición

- 1. Reducción del tiempo de detección de anomalías en un X%**
 - a. Requerimiento Asociado:**
 - i. Monitoreo en tiempo real de presión, temperatura y volumen de gas.**
 - b. Submétrica:**
 - i. Tiempo promedio de detección de anomalías (en minutos) antes y después de la implementación del sistema.
- 2. Reducción de los costos operacionales asociados a la gestión de anomalías en un X%**
 - a. Requerimiento Asociado:**
 - i. Generación de alertas ante anomalías.**
 - b. Submétrica:**
 - i. Costo promedio por intervención manual y automatizada (reducido por la implementación de alertas automatizadas).
- 3. Mejora en la precisión de los modelos de detección en un X%**
 - a. Requerimiento Asociado:**
 - i. Capacidad del sistema para aprender y ajustar alertas basadas en patrones históricos.**
 - b. Submétrica:**
 - i. Tasa de falsos positivos y negativos en la detección de anomalías (reducción porcentual esperada).

Otras Métricas de Éxito Relacionadas con el Triángulo de Equilibrio

- 1. Tiempo: Entrega del PMV dentro del cronograma acordado**
 - a. Requerimiento Asociado:**
 - i. Visualización de datos históricos de consumo.**
 - b. Submétrica:**
 - i. Cumplimiento de los hitos en la entrega de cada funcionalidad dentro del cronograma.
- 2. Costo: Desarrollo del proyecto dentro del presupuesto establecido**
 - a. Requerimiento Asociado:**
 - i. Resúmenes descriptivos del comportamiento de cada cliente.**
 - b. Submétrica:**
 - i. Desviación financiera del presupuesto inicial para el desarrollo y mantenimiento del sistema.

- 3. Alcance: Implementación de todas las funcionalidades requeridas**
 - a. Requerimiento Asociado:**
 - i. Visualización, alertas de anomalías y criticidad.**
 - b. Submétrica:**
 - i. Porcentaje de funcionalidades implementadas y operativas al cierre del proyecto.
- 4. Calidad: Precisión y confiabilidad de los modelos de detección de anomalías**
 - a. Requerimiento Asociado:**
 - i. Monitoreo de alertas y reducción de falsos positivos.**
 - b. Submétrica:**
 - i. Evaluación de la efectividad de los modelos de detección en entornos controlados antes de la implementación.

Proceso de Negociación y Revisión con Stakeholders

Estas métricas deben ser revisadas y negociadas con los stakeholders involucrados para asegurar que reflejan adecuadamente las expectativas y necesidades del proyecto. Esto incluye discusiones sobre la viabilidad de alcanzar los porcentajes esperados, la relevancia de cada métrica para los objetivos estratégicos de Contugas y la manera en que serán monitoreadas durante el ciclo de vida del proyecto. La validación y ajuste de estas métricas garantizarán que el progreso pueda ser medido de manera efectiva y que el proyecto se mantenga alineado con sus objetivos de negocio.

Fase d) Los recursos

a. Personal y perfiles requeridos

Para llevar a cabo el proyecto, se requerirá un equipo multidisciplinario que incluya un Gerente de Proyecto responsable de la planificación y ejecución, analistas de datos para la preparación y análisis de los datos, ingenieros de software para la implementación técnica, especialistas en machine learning para el desarrollo de modelos de detección de anomalías, personal de TI para el soporte técnico, así como los stakeholders de Contugas, como el Gerente de Operaciones y el equipo de Innovación, quienes proporcionarán el contexto del negocio y validarán los resultados.

b. Tiempo de implementación

La duración total del proyecto se estima en aproximadamente 6 a 8 meses, distribuidos en cuatro fases: la recolección de requisitos y análisis tomará 1 mes, el diseño y desarrollo del sistema requerirá 3 meses, seguido de 2 meses dedicados a pruebas y validación, y finalmente, 1 mes para la implementación y capacitación del personal involucrado.

c. Otros recursos tales como tecnológicos

El proyecto necesitará recursos tecnológicos que incluyan hardware como servidores y equipos de cómputo para el personal; software para el análisis de datos, como Python y herramientas de BI como Tableau; una plataforma de machine learning en la nube como AWS o Azure; y bases de datos para almacenar datos históricos y operacionales, así como licencias y suscripciones a los servicios requeridos.

d. Presupuesto de la inversión (Supongan un aproximado de acuerdo con la información con la que cuentan).

Para la implementación del proyecto de detección de anomalías en el consumo de gas, se requerirá un equipo multidisciplinario que incluya un analista de datos, un ingeniero de software, un especialista en sistemas de información geográfica (SIG), un experto en operaciones del sector energético, un gestor de proyectos y un especialista en capacitación. Este equipo deberá colaborar estrechamente durante un periodo estimado de seis meses, distribuidos en fases que abarcarán desde la recopilación de datos y desarrollo del sistema hasta pruebas de funcionalidad y capacitación del personal operativo en Contugas. En cuanto a los recursos tecnológicos, se necesitarán herramientas de software para análisis y visualización de datos, servidores para almacenamiento y procesamiento de información, así como dispositivos de medición y monitoreo en campo. El presupuesto de inversión se ha establecido en aproximadamente **\$50,000 USD**, que cubrirá costos de personal, tecnologías necesarias, licencias de software y otros recursos adicionales, y deberá ser gestionado cuidadosamente para asegurar que se cumplan todos los objetivos del proyecto sin comprometer la calidad de los resultados.

Fase e) El planteamiento inicial del Problema de Analytics

Escribe una descripción articulada del problema a abordar empleando desde los requerimientos del primer punto hasta los resultados del punto anterior. Esta descripción deberá considerar los siguientes aspectos:

a. Objetivos a abordar con base en los requerimientos validados en la solución.

El objetivo principal del proyecto es desarrollar una solución analítica que permita a Contugas detectar anomalías en el consumo de gas industrial, especialmente a través de variables críticas como presión, temperatura y volumen. Este sistema deberá identificar patrones atípicos en tiempo real y emitir alertas automáticas para que el equipo operativo pueda intervenir rápidamente. Además, se busca reducir los costos asociados a la gestión manual de anomalías y optimizar la precisión del sistema, minimizando falsos positivos y garantizando una respuesta eficiente ante posibles incidentes.

b. Presupuesto de recursos, tales como económico, humano, etc.

Para llevar a cabo este proyecto, se cuenta con un presupuesto total de \$50,000 USD, destinado a cubrir licencias de software analítico, adquisición de herramientas tecnológicas, y salarios de especialistas en analítica de datos, desarrollo y gestión de proyectos. Este monto también contempla la capacitación del equipo operativo en el uso del sistema. Adicionalmente, se requiere un equipo multidisciplinario que incluya ingenieros de datos, científicos de datos, expertos en TI y personal del área operativa de Contugas para asegurar el éxito de la implementación.

c. Tiempo disponible de solución.

El tiempo estimado para completar el proyecto es de seis meses. Este plazo considera cada una de las fases del proyecto: desde el diseño y desarrollo inicial del sistema, pasando por las pruebas y validaciones necesarias, hasta la capacitación del personal operativo. A lo largo de este período, se realizarán revisiones periódicas y ajustes necesarios para asegurar que la solución cumpla con los estándares de precisión y efectividad esperados.

d. Resultado o producto esperado.

El producto final de este proyecto será un sistema de monitoreo de consumo de gas que ofrezca visualización de datos históricos, identificación y alerta en tiempo real de consumos anómalos, y una interfaz intuitiva para los usuarios. Este sistema permitirá a Contugas tomar decisiones de manera informada y oportuna, mejorando la eficiencia operativa y contribuyendo a la seguridad en la red de distribución, con un impacto positivo en la calidad del servicio brindado a sus clientes industriales.

Parte 2. Establecimiento de las fuentes y objetivos de análisis de los datos.

Fase a) Establecer las fuentes

1. Acorde el planteamiento analítico y avance al momento ¿identificas que se requieren más fuentes de datos que las planteadas en la GUÍA 1? En caso afirmativo, describe cuales serían estas fuentes de información.

Para robustecer el análisis de detección de anomalías, es recomendable incluir fuentes adicionales, además de las ya establecidas en la Guía 1. Estas fuentes deberían incluir el historial de mantenimiento de equipos de medición, registros de eventos operativos que puedan afectar el suministro (como interrupciones programadas o fallos), y condiciones climáticas, ya que factores ambientales pueden influir en las variaciones de presión y temperatura del gas. Incluir estos datos permitiría mejorar la precisión de las detecciones al contextualizar los eventos atípicos dentro del sistema de monitoreo.

2. ¿Los datos identificados son coherentes para abordar el problema de negocio planteado o se tienen que replantear?

La coherencia de los datos actuales es adecuada para abordar el problema de negocio, ya que incluyen las variables clave de presión, temperatura y volumen que afectan directamente el consumo de gas. Sin embargo, la adición de datos sobre el estado del equipo y las condiciones externas permitiría una visión más completa y precisa, aumentando la confiabilidad en la detección de anomalías y facilitando la identificación de causas subyacentes.

3. ¿las fuentes adicionales de información requieren el desarrollo y aplicación de algún instrumento (ej. Encuesta)? Si sí, describe. Nota: considera que requerirás tiempo para obtenerlo.

En este caso, no es necesario desarrollar instrumentos como encuestas. Sin embargo, la integración de datos adicionales sobre mantenimiento y condiciones ambientales sí requeriría la colaboración de otros departamentos de Contugas. La información de mantenimiento y eventos operativos podría recopilarse de bases de datos internas ya existentes, mientras que los datos climáticos podrían provenir de fuentes externas públicas o privadas, pero sin requerir tiempo adicional para desarrollar nuevos instrumentos.

4. Para los datos que existen y los que se tengan que generar, indica qué tipo de datos son (ej., series de tiempo, datos no estructurados, datos panel, entre otros).

Los datos actuales son series de tiempo registradas por cliente, esenciales para detectar variaciones horarias en las variables de consumo. Los datos adicionales, como el historial de mantenimiento, también pueden estructurarse como series de tiempo, y los datos climáticos podrían organizarse de la misma forma. Esta estructuración en series de tiempo facilita su integración con el sistema actual y permite analizar las variaciones en contextos específicos.

5. Con base en los requerimientos establecidos y tu respuesta anterior, especifica si es necesaria alguna transformación que genere datos útiles para el análisis. Ejemplo, si consideras datos basados en observaciones humanas, cuestionarios, opiniones de redes sociales.

Para mejorar la utilidad en el análisis, sería necesario aplicar transformaciones específicas, como la normalización y sincronización temporal entre todas las fuentes de datos. Las series de tiempo deberían estandarizarse en intervalos horarios y alinearse cronológicamente para facilitar la detección de patrones y reducir posibles distorsiones en el análisis.

6. Si es posible, identifica qué departamentos o áreas de la organización están involucradas con las fuentes de datos que dan respuesta a los requerimientos antes establecidos.

La colaboración de los departamentos de TI, mantenimiento y operaciones de Contugas es fundamental para la integración de estas fuentes de datos adicionales. TI sería responsable de estructurar y asegurar la disponibilidad de datos en la plataforma de análisis, mientras que mantenimiento y operaciones pueden proporcionar y actualizar la información de los eventos operativos y los registros de mantenimiento.

Fase b) Determinar los objetivos del análisis de los datos

1. Acorde al planteamiento de negocio y requerimientos previamente establecidos, ¿cuál(es) sería el objetivo de análisis de los datos para responder a este problema?, redáctalo(s).

El objetivo de análisis de los datos para este proyecto es detectar patrones y anomalías en el consumo de gas de clientes industriales de Contugas mediante la utilización de series temporales de presión, temperatura y volumen. Esto implica desarrollar un sistema de monitoreo que identifique desviaciones inusuales en las variables operativas clave, para así optimizar el mantenimiento de la red, reducir costos y mejorar la eficiencia operativa. Además, el análisis debe priorizar la precisión en la detección de anomalías, minimizando tanto los falsos positivos como los falsos negativos, lo que asegurará la confiabilidad en el sistema de alertas y reducirá las interrupciones de servicio.

2. ¿Qué nivel de transformación de los datos generaría una respuesta a tu problema de negocio?

La solución requiere transformar los datos mediante procesos de limpieza, estandarización y sincronización temporal para asegurar la precisión de los análisis. Las series temporales de presión, temperatura y volumen deben normalizarse a intervalos específicos (por ejemplo, cada hora) y alinearse cronológicamente. También sería útil aplicar técnicas de suavizado y filtrado para reducir el ruido y facilitar la identificación de patrones y tendencias en el consumo. Adicionalmente, la creación de indicadores derivados, como la relación entre presión y volumen en condiciones de temperatura constante, puede facilitar la detección de comportamientos anómalos en situaciones operativas específicas.

3. ¿Qué modelos y técnicas podrían dar solución al problema?

El proyecto puede beneficiarse de modelos de detección de anomalías basados en métodos de aprendizaje supervisado y no supervisado. Entre las técnicas recomendadas están los modelos de series temporales como ARIMA o LSTM (redes neuronales de memoria a largo plazo) para capturar patrones de consumo. También se podrían emplear técnicas de clustering, como DBSCAN o K-Means, para identificar

comportamientos fuera de lo común en el consumo de gas de los clientes. Alternativamente, los modelos de bosques de aislamiento (Isolation Forest) y el análisis de valores atípicos multivariados pueden ser útiles para detectar anomalías en varias dimensiones (presión, temperatura y volumen) de manera simultánea.

4. Posteriormente, ¿qué actividades implementarías para dar respuesta al problema de negocio?

Las actividades clave para responder al problema de negocio incluyen: (1) consolidación y preprocesamiento de los datos de las variables operativas clave, asegurando su calidad y consistencia; (2) desarrollo y ajuste de modelos de detección de anomalías utilizando las series temporales preprocesadas; (3) validación de los modelos y pruebas en entornos controlados para garantizar su precisión y confiabilidad; (4) implementación de un sistema de alertas en tiempo real para notificar eventos atípicos y facilitar la respuesta inmediata; y (5) diseño de una interfaz amigable que permita a los usuarios visualizar las anomalías y realizar análisis detallados, promoviendo una toma de decisiones informada para la gestión eficiente del servicio de gas.

Fase c) Determinar criterios de éxito en el análisis

1. Describe las métricas para la evaluación de las técnicas identificadas.

Para evaluar la efectividad de las técnicas de detección de anomalías en el consumo de gas, se utilizarán métricas de precisión, como la tasa de detección de anomalías (recall) y la tasa de falsos positivos (FPR), que indicarán qué tan bien los modelos identifican anomalías sin emitir alertas innecesarias. Además, se evaluará la precisión general (accuracy) para medir el rendimiento del modelo en detectar tanto anomalías como casos normales correctamente. El tiempo de procesamiento será otra métrica clave, dado que la detección en tiempo real es esencial para la respuesta oportuna ante posibles incidentes.

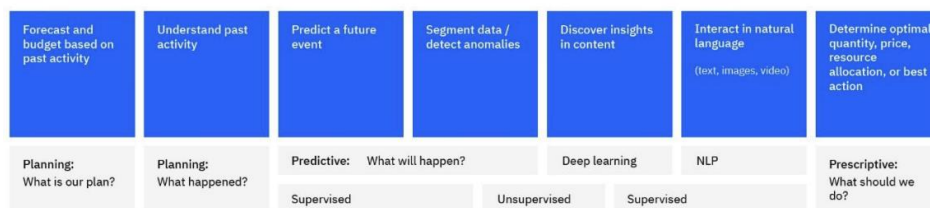
2. Define el benchmark para evaluar el éxito de las métricas específicas.

El benchmark para el éxito de estas métricas se definirá en función de umbrales específicos: un recall mínimo del 90% para asegurar que la mayoría de las anomalías sean detectadas, una tasa de falsos positivos inferior al 5% para evitar alertas incorrectas que puedan incrementar los costos operativos, y un tiempo de procesamiento menor a un minuto por lote de datos para mantener la detección en tiempo real. Estos valores serán establecidos con base en análisis históricos y validaciones en datos previos para equilibrar precisión y eficiencia en la operación.

3. Define si hay algunas métricas subjetivas que impacte en la decisión de si el modelo se considerará o no exitoso.

Como métrica subjetiva, se considerará la percepción de los equipos operativos y de mantenimiento sobre la utilidad y facilidad de uso de las alertas generadas por el modelo. Si el sistema es percibido como excesivamente complejo o produce muchas alertas irrelevantes, esto podría impactar su aceptación y uso. Además, la confianza de los stakeholders en la capacidad del sistema para identificar anomalías críticas y en su consistencia a lo largo del tiempo será clave para determinar el éxito del modelo, ya que un sistema confiable y fácil de interpretar impulsará su integración en las operaciones diarias de Contugas.

Fase d) Tipos de Problemas de Analytics



1. Acorde los objetivos de análisis de los datos y criterios de éxitos en el análisis, ¿cuál es o son los tipos de enfoques y técnicas que identificas son candidatos para dar respuesta al problema de negocio?

Para el problema de detección de anomalías en el consumo de gas en clientes industriales de Contugas, los enfoques principales serían los de segmentación y detección de anomalías mediante técnicas no supervisadas, como el clustering y métodos de outlier detection. Adicionalmente, técnicas predictivas basadas en series de tiempo, como ARIMA o LSTM, podrían usarse para modelar patrones de consumo y prever comportamientos esperados, de manera que se detecten desviaciones significativas. Estos enfoques analíticos pueden combinarse en un pipeline de análisis para optimizar la precisión y confiabilidad en la detección de comportamientos anómalos en tiempo real.

2. ¿De qué manera estas técnicas dan respuesta a los objetivos de análisis antes establecidos?

Las técnicas de clustering y detección de anomalías permiten segmentar los datos históricos de consumo para definir patrones normales de comportamiento, identificando cualquier desviación significativa como una anomalía. A su vez, los modelos predictivos de series de tiempo ayudan a anticipar patrones futuros de consumo bajo condiciones normales, permitiendo detectar desviaciones en el tiempo en el que se producen, lo que optimiza la rapidez y precisión en la respuesta. Este

enfoque asegura una alta precisión en la identificación de anomalías y minimiza falsos positivos, alineándose con los objetivos de reducir costos y mejorar la eficiencia operativa.

3. ¿Cómo se alinearían los resultados de la aplicación de estas técnicas con el problema de negocio?

La aplicación de estas técnicas permitirá detectar de manera oportuna y precisa las anomalías en el consumo de gas, lo que contribuye a optimizar el monitoreo de la red, reducir costos de mantenimiento e inspección, y evitar incidentes mayores. Los resultados serán indicadores confiables de las condiciones operativas, permitiendo a Contugas tomar decisiones de intervención temprana y minimizar interrupciones en el servicio, maximizando la satisfacción del cliente y la rentabilidad de la compañía al evitar fugas o comportamientos inusuales que puedan afectar la continuidad del servicio.

4. ¿Qué acciones implementarías en cada caso de la aplicación de técnicas para dar respuesta al problema de negocio?, ¿se requeriría la aplicación de otro nivel más de técnicas? (lo que comentamos del sentido de generar patrones de técnicas a los datos) Si sí, describir.

Implementaría un pipeline iterativo que inicie con la preparación de datos y el uso de técnicas de clustering para establecer patrones normales de consumo, seguido de la aplicación de modelos predictivos para detectar anomalías en tiempo real. En caso de detecciones, se generarían alertas automáticas para el equipo operativo. Podría ser necesario integrar técnicas de validación continua, como el ajuste de umbrales de alerta en función de la retroalimentación del equipo de operaciones y el monitoreo de la tasa de falsos positivos. Un paso adicional sería incorporar una capa de aprendizaje profundo (como LSTM) si las anomalías resultan complejas y no siguen patrones evidentes en las series de tiempo, para asegurar un sistema que evolucione con los cambios de comportamiento del consumo de gas en el tiempo.

Fase e) Planteamiento del Problema de Analytics

*A partir de los resultados de las secciones Parte I y Parte II, **redacta un enunciado** que describa el problema a resolver mediante la descripción de los objetivos del análisis de los datos empleando modelos y técnicas de Analytics, el tipo de datos requeridos, sus características, las métricas de éxito de la aplicación de tales técnicas y cómo los resultados potenciales de estos análisis apoyan la pregunta de negocio. Igualmente, incluye una descripción de los recursos tales como el tiempo base considerado en la solución y los beneficios esperados.*

Cabe mencionar que una vez identificados todos los puntos de la Parte II, debes proceder a obtener los datos, ya sea mediante instrumentos de recolección o adquisición a las fuentes de la organización. Considera que los datos deben:

- *Capturar y mantener la descripción del problema a resolver.*
- *Ayudar a especificar todas las preguntas, hitos de negocio de la manera más precisa posible.*
- *Alinearse con los requerimientos de negocio.*
- *Permitir derivar los beneficios esperados en términos de negocio.*

El problema a resolver consiste en implementar una solución de analítica avanzada para detectar anomalías en el consumo de gas de los clientes industriales de Contugas, empleando datos operativos de presión, temperatura y volumen registrados de manera horaria en bases de datos por cliente. Los objetivos del análisis incluyen establecer patrones normales de consumo a través de técnicas de clustering y detección de *outliers*, y modelar series de tiempo para predecir comportamientos futuros y detectar desviaciones anómalas en tiempo real. Las métricas de éxito comprenden la tasa de detección de anomalías con bajo margen de falsos positivos, optimización de costos y reducción de tiempos de respuesta operativa. Se espera que los resultados de estos análisis proporcionen alertas tempranas y precisas, mejorando la eficiencia operativa y la calidad del servicio, al tiempo que se maximiza la rentabilidad mediante la reducción de costos por mantenimiento preventivo y monitoreo constante de la red. La solución, con un presupuesto base de \$50,000 USD y un tiempo de implementación estimado en seis meses, permitirá la toma de decisiones informadas y oportunas sobre el estado de la red de distribución de gas, asegurando que los datos capturados se alineen con las necesidades del negocio, respondan a preguntas clave sobre el consumo y generen beneficios tangibles en la operación y satisfacción del cliente.

Parte 3 Estado del arte: referencias de investigaciones, aplicaciones en el área.

- **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

El proyecto “Smart University: Detección de Anomalías en Consumo Energético” se centra en mejorar la eficiencia energética y sostenibilidad en universidades mediante redes neuronales. Utilizando datos de consumo, wifi, día y hora, la red detecta patrones de uso y anomalías, permitiendo a la universidad mitigar desperdicios y reducir costos en tiempo real. Con enfoque en la responsabilidad ambiental y tecnología avanzada, busca no solo ahorrar recursos, sino también establecer un modelo de gestión responsable en instituciones educativas. La implementación sigue la metodología ágil Scrum para optimizar resultados y adaptarse rápidamente a los objetivos del proyecto.

Este trabajo puede aportar enfoques y tecnologías innovadoras al proyecto de Contugas, especialmente en cuanto al uso de redes neuronales para detectar patrones y anomalías automáticamente en grandes volúmenes de datos. La experiencia en la gestión eficiente de consumo energético y la implementación de una solución de monitoreo en tiempo real ayudarían a Contugas a optimizar su capacidad de respuesta ante anomalías, reducir costos y mejorar la eficiencia operativa.

- **Baldacci, Lorenzo & Golfarelli, Matteo & Lombardi, Davide & Sami, Franco. (2016). Natural Gas Consumption Forecasting for Anomaly Detection. Expert Systems with Applications. 62. 10.1016/j.eswa.2016.06.013.**

Predecir el consumo de gas natural es crucial para las actividades de muchas empresas suministradoras de gas, como la optimización en la compra de gas, el monitoreo, la gestión y la seguridad de la red de tuberías. Este estudio documenta la colaboración realizada con HERA S.p.A., un proveedor líder de gas en Italia, con el fin de prever el consumo de gas en una red específica y detectar flujos anómalos basándose en datos históricos, facilitando así los procesos de monitoreo y seguridad en su sala de control central. Cada 15 minutos se toman muestras de las condiciones históricas de la red, incluyendo el flujo de gas, la temperatura exterior y la marca de tiempo de cada registro. Para lograr esto, se realizaron análisis descriptivos de datos históricos de un pueblo y una pequeña ciudad, y se implementaron dos enfoques de pronóstico: uno mediante el método de vecinos más cercanos y otro basado en análisis de regresión local. Los resultados experimentales indican que los datos históricos recopilados permiten una previsión confiable del consumo de gas. Se presenta un análisis detallado, tanto cuantitativo como cualitativo, de ambos métodos, resaltando sus fortalezas y limitaciones. Además, en colaboración con expertos en el tema, evaluamos la capacidad de estas técnicas para identificar consumos anómalos de gas, lo que facilita a los operadores en la sala de control la detección de anomalías en el consumo. Esta función de previsión constituye un paso clave hacia el desarrollo de un sistema experto completo, que permita a los operadores avanzados interpretar el comportamiento de la red de gas y sugiera a operadores con menos experiencia las reacciones adecuadas ante eventos inusuales.