**Parte I: Planteamiento del problema de negocio en uno de Analytics. es traducir el “qué” del problema de negocio en un “cómo”, que se convertirá en el problema de Analytics.**

**Fase a) Los encargados de la solución**

**¿Qué resultados se desean obtener tras la implementación del proyecto?**

* Implementar un sistema de detección automática de anomalías que permita identificar patrones anómalos de consumo mediante la analítica de datos con una precisión mayor al 85%.
* Reducir el tiempo de respuesta en la detección de anomalías para optimizar los costos en recursos usados en este monitoreo.
* Disminución del error humano para mejorar la confiabilidad del proceso de detección de patrones anómalos.
* Visualización por medio de dashboards de los históricos y en tiempo real de los patrones de consumo para la detección automática de anomalías en el consumo de contugas

**¿Quiénes participarán en la solución analítica?**

Los stakeholders de Contugas directamente involucrados en el proyecto serán:

* Equipo de Operaciones de Contugas: para validar los datos operativos y la detección de anomalías, por tanto, son fundamentales en el proceso del desarrollo del sistema de detección automática de anomalías
* Equipo de Tecnología de la Información (TI): responsable de la infraestructura y seguridad del sistema y gestión de datos y también del soporte técnico.
* Proveedores de tecnología: para proporcionar las herramientas y tecnologías necesarias para implementar la solución.
* Gerencia de Contugas: para supervisar y tomar decisiones estratégicas en el proyecto

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

* Reducción de tiempos de detección de anomalías que inciden en reducción de costos operativos por la atención eficiente de estos patrones anómalos es decir habrá una **mejora en la eficiencia operativa**
* **Reducción u optimización de costos en recursos humanos** destinados a identificar estas anomalías en el consumo en el consumo de gas por que el sistema hará esta detección de forma automática.
* **Mayor confianza del equipo de Contugas** en los resultados del sistema de detección automático de anomalías en el consumo de gas se reduce el error humano

**Fase b) Lo que el cliente requiere**

**¿Cuáles son los requerimientos básicos u obligatorios que se atienden con este proyecto?**

Los requerimientos básicos y mínimos viables que se atenderán con el proyecto es la precisión del modelo sea 85% o superior, la interfaz gráfica debe estar es idioma español, ser sencilla, clara y comprensible para el usuario, que sea en tiempo real.

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

Utilizar una instancia de AWS para el despliegue de esta solución analítica, dadas las características de escalabilidad e integración resulta muy útil para el desarrollo, así como el versionamiento de experimentos y el tipo de modelo a utilizar no afecta la satisfacción del cliente, siempre y cuando considere la mínima métrica esperada es un **requerimiento indiferente**, que la latencia no sea superior a 1 min, ya que puede considerarse **inverso** en caso tal de interrupciones del servicio, tome más de 1 min su actualización o carga.

**¿Cuáles son los requerimientos unidimensionales?**

Si la solución implementada es sencilla, amigable con el usuario, es rápida en su carga de información menos de 1 segundo.

**¿Qué requerimientos son atractivos?**

* Rendimiento: Velocidad de funcionamiento, capacidad de procesamiento.
* Calidad: Durabilidad del producto, fiabilidad, tasa de fallos de la herramienta tecnólogica.
* Costo: Precio competitivo, costo de producción, costos de mantenimiento.
* Usabilidad: Facilidad de uso, accesibilidad, ergonomía.
* Los requerimientos que pueden ser atractivos son:
* Innovación: Funcionalidades únicas o tecnología avanzada que mejore la experiencia del usuario.
* Estética: Diseño atractivo, variedad de colores y acabados.
* Personalización: Opciones para personalizar el producto según las preferencias del usuario.

**Requerimientos para incrementar:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre y Descripción** | **Viabilidad/Desarrollo/Implementación**  **(1-10)** | **Impacto (1-10)** | **Score**  **Total** |
| Visualización de Datos Históricos | 10 | 10 | 20 |
| Capacidad para mostrar datos de presión, temperatura y volumen de forma clara y visual  Resumen Descriptivo del Comportamiento Histórico Generación automática de estadísticas descriptivas (promedio, cuartiles, mínimo, máximo). | 10 | 10 | 20 |
| Modelo de Identificación de Consumos Anómalos | 10 | 10 | 20 |
| Capacidad para detectar y resaltar variaciones atípicas en los datos de consumo. | 9 | 10 | 19 |
| Alertas de Comportamiento Anómalo | 10 | 10 | 20 |
| Notificaciones sobre comportamientos anómalos, indicando la criticidad (bajo, medio, alto). | 9 | 10 | 19 |

**Requerimientos para aminorar:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nombre y Descripción** | **Viabilidad/Desarrollo/Implementación**  **(1-10)** | **Impacto (1-10)** | **Score**  **Total** |
| Minimizar latencia y carga de datos en herramienta de visualización | 10 | 10 | 20 |
| Minimizar los falsos positivos en el modelo de predicción | 10 | 10 | 20 |

**Fase d) Los recursos:**

**Personal y perfiles requeridos.**

* **Científico de Datos**: Confecciona el modelo analítico y los algoritmos para la detección de anomalías en el consumo de gas.
* **Ingeniero de Datos**: Es el responsable de la integración y procesamiento de los datos operativos de consumo.
* **Analista de Negocios**: Ayuda a traducir los requerimientos de negocio en especificaciones funcionales, además analiza los resultados de la solución.
* **Directora de proyecto:** Coordina y dirige el proyecto para asegurar la alineación de recursos y plazos de entrega.

**b. Tiempo de implementación.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Actividades** | **Fecha Inicio** | **Fecha de Fin** | **Descripción** | **Responsables** |
| Definir | 7/10/2024 | 20/10/2024 | Semana 1 y 2: Planificación inicial, identificación del problema, propósito, objetivos y alcance | Todo el equipo del proyecto |
| Medir | 21/10/2024 | 28/10/2024 | Semana 3: Recolección de datos clave, establecimiento de métricas de rendimiento | Analista de Negocio |
| Analizar | 29/10/2024 | 11/11/2024 | Semana 4 y 5: Análisis de datos para diseñar modelos de detección de anomalías | Científica de Datos |
| Mejorar | 12/11/2024 | 18/11/2024 | Semana 6 y 7: Implementación transversal de mejoras técnicas en todo el proceso | Ingeniero de Datos |
| Controlar | 19/11/2024 | 30/11/2024 | Semana 8: Monitoreo y control de mejoras para asegurar la sostenibilidad y éxito del proyecto | Directora del Proyecto |
| **Desarrollo de Modelos** | 1/4/2025 | 31/5/2025 | Semana 9 a la 16 en curso **Proyecto Aplicado en Analítica de Datos:** Fase de diseño e implementación de modelos de detección automática durante los dos meses curso | Todo el equipo del proyecto |

**c. Otros recursos tales como tecnológicos.**

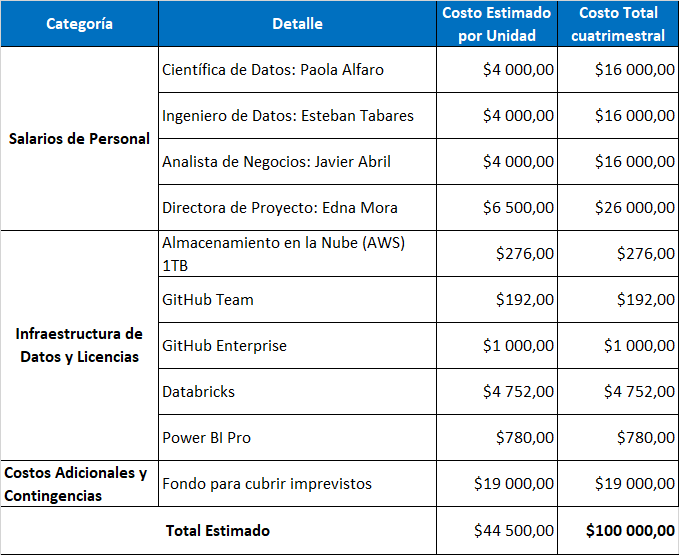
* Almacenamiento en Nube para gestionar los datos. Aws escalabilidad para gestionar en el proyecto.
* Machine Learning por ejemplo TensorFlow o PyTorch para los modelos de detección.
* Dashboards y visualización de datos: Herramientas como Power BI o Tableau para facilitar la supervisión.

**d. Presupuesto de la inversión.** Presupuesto de Inversión estimado según resultados financieros contugas:

* Salarios de personal: $75,000).
* Recursos tecnológicos y licencias: $10,000.
* Infraestructura de datos y servidores: $10,000.
* Costos adicionales y contingencias: $5,000.

Total, estimado: $100,000. Esto como un porcentaje mínimo y precavido de uso económico para el proyecto

Detalle del presupuesto:



**Fase e) El planteamiento inicial del Problema de Analytics**

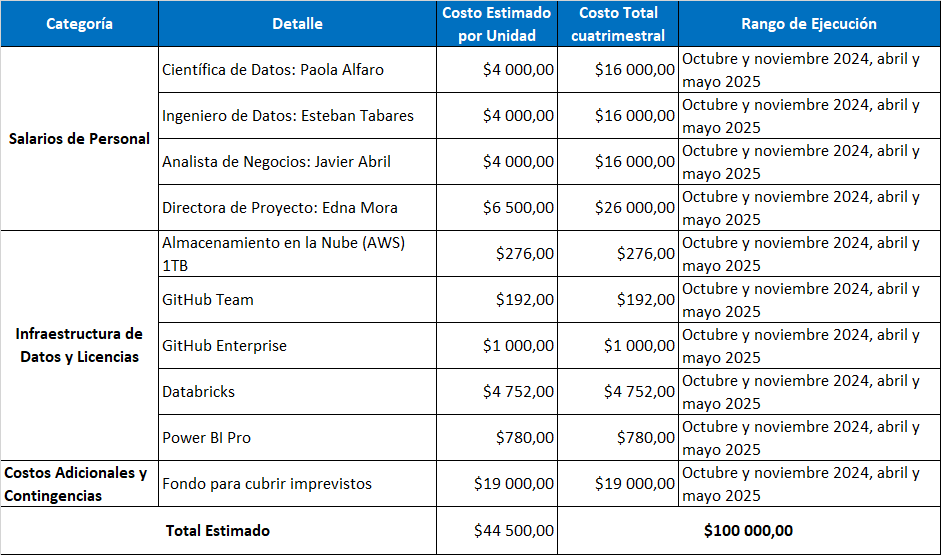
La detección automática de anomalías en el consumo de gas en los clientes industriales de Contugas para una mejor capacidad de monitoreo y respuesta ante eventos atípicos críticos.

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

1. **Desarrollo de un sistema de detección de anomalías** para identificar patrones de consumo anómalos con una precisión superior al 85%.
2. **Optimización de tiempo de respuesta** en la identificación de anomalías, para reducir costos en monitoreo y costos y errores manuales.
3. **Implementación de una plataforma de visualización de datos** que facilite la detección de anomalías en tiempo real para toma de decisiones basadas en datos confiables.

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

Total, estimado: $100,000. Se adjunta detalle de recursos económicos detallando el tiempo de ejecución:



* **Científico de Datos**: Paola Alfaro: Confecciona el modelo analítico y los algoritmos para la detección de anomalías en el consumo de gas.
* **Ingeniero de Datos**: Esteban Tabares: Es el responsable de la integración y procesamiento de los datos operativos de consumo.
* **Analista de Negocios**: Javier Abril: Ayuda a traducir los requerimientos de negocio en especificaciones funcionales, además analiza los resultados de la solución.
* **Directora de proyecto:**  Edna Mora:Coordina y dirige el proyecto para asegurar la alineación de recursos y plazos de entrega.

**c. Tiempo disponible de solución.**

La entrega final del proyecto se estima para mayo de 2025, considerando un segundo curso de desarrollo del modelo en abril y mayo. Se adjunta el cronograma en la fase d. Tiempos de implementación

**d. Resultado o producto esperado.**

Sistema automatizado de detección de anomalías en el consumo de gas para clientes industriales de Contugas, que permita identificar patrones de consumo atípicos con una precisión superior al 85%.

**Parte II: Establecimiento de fuentes y objetivos de análisis de los datos**

**Fase a) Establecer las fuentes:**

1. **Coherencia de los datos identificados:** Los datos de presión, volumen y temperatura son coherentes y fundamentales para abordar el problema de detección de anomalías en el consumo de gas. Sin embargo, se deben incorporar fuentes adicionales, como datos meteorológicos y de mantenimiento, para enriquecer el análisis y aumentar la precisión en la identificación de anomalías.
2. **Fuentes adicionales de datos:** Se identifican varias fuentes adicionales necesarias para mejorar el análisis y reducir la incidencia de falsos positivos:

Las siguientes serían posibles fuentes de datos que nos ayudarían a enriquecer el proyecto, sin embargo, por el factor tiempo o alcance del proyecto **no se van a implementar,** ejemplos: **Datos meteorológicos:** Variables como temperatura y humedad para ajustar modelos de consumo de gas. Se puede obtener del servicio meteorológico local o de servicios en línea como AccuWeather u otros. **Historial de mantenimiento & registros de intervenciones externas:** Usar registros de estado y mantenimiento de medidores y tuberías para contextualizar posibles causas de anomalías. Se pueden obtener de sistemas de gestión de mantenimientos o archivos internos de la misma compañía. **Datos financieros del cliente:** Integrar información económica para correlacionar patrones de consumo con la actividad económica de los clientes industriales. Obtenidas directamente con el cliente, si es posible.

1. **Instrumentos para la Recolección de Datos:** Se deberán desarrollar instrumentos de recolección para obtener datos que no están digitalizados. Esto incluye:
   1. **Encuestas de mantenimiento:** Para recopilar información cualitativa sobre el estado de los medidores y tuberías, así como eventos que no están registrados digitalmente.
   2. **Monitoreo manual de condiciones ambientales:** Si no se cuentan con sensores automáticos para recopilar datos meteorológicos, se requerirán registros manuales, lo que podría extender el tiempo necesario para su obtención.
2. **Tipos de Datos:** Las fuentes de adquisición implicarán los siguientes tipos de datos:
   1. **Datos Estructurados:** Registros de presión, volumen y temperatura, así como datos financieros, que se pueden almacenar en bases de datos.
   2. **Datos No Estructurados:** Informes de mantenimiento o encuestas que pueden contener información valiosa pero que no están organizados en un formato fácilmente analizable.
   3. **Datos de Series de Tiempo:** Datos operativos que cambian a lo largo del tiempo, necesarios para el análisis de patrones y tendencias.
3. **Transformaciones Necesarias.** Es probable que se necesiten transformaciones para generar datos útiles para el análisis. Estas pueden incluir:
   1. Normalización de Datos: Ajustar datos de diferentes fuentes y formatos para facilitar su comparación.
   2. Creación de Variables Derivadas: Generar indicadores a partir de los datos existentes, como promedios móviles o detección de outliers.
   3. Procesamiento de Datos No Estructurados: Aplicar técnicas de procesamiento de lenguaje natural (NLP) para extraer información relevante de informes y encuestas.
4. **Departamentos involucrados:** Las siguientes áreas de la organización están involucradas con las fuentes de datos:
   1. Operaciones: Encargada de la recolección de datos operativos y registros de mantenimiento
   2. Tecnología de la Información (TI): Responsable de la gestión, calidad y seguridad de los datos.
   3. Mantenimiento: Proporciona información sobre intervenciones en medidores y tuberías.
   4. Finanzas: Colabora con datos económicos y patrones de consumo de clientes industriales.

**Fase B) Definir los Objetivos del Análisis de Datos**

**Objetivos del Análisis de Datos.**

1. Identificar patrones y posibles “Anomalías”: Detectar patrones de consumo normal y anomalías en variables como presión, volumen y temperatura en los clientes industriales de Contugas.

2. Monitorear consumo de gas: Implementar un sistema de monitoreo de consumo de gas de los clientes industriales, datos históricos e integrando el modelo desarrollado.

3. Generar tablero de control con alertas de anomalías: El tablero de control apoya a la toma de decisiones y acciones respecto a los patrones de consumo de gas de los clientes industriales.

**Resultados Esperados**

1. Monitoreo de consumo de gas de clientes industriales, para apoyar la toma de decisiones y acciones más eficientes.

2. Dashboard de Control sobre el consumo de gas de clientes industriales, facilitando la planificación y la asignación de recursos.

**Acciones para Abordar el Problema de Negocio**

1. Recolección y Preprocesamiento de Datos: Obtener datos históricos y externos relevantes para enriquecer el análisis.
2. Análisis Exploratorio de Datos: Evaluar patrones de consumo para establecer comportamientos normales.
3. Entrenamiento y Validación de Modelos: Desarrollar y validar modelos de detección de anomalías utilizando datos históricos.
4. Implementación de Sistemas de Alerta: Desplegar un sistema para notificar al equipo de operaciones y a los clientes sobre anomalías en tiempo real.
5. Capacitación del Personal: Formar al equipo operativo en el uso de herramientas de análisis y monitoreo.
6. Monitoreo y Ajuste Continuo de Modelos: Realizar ajustes regulares a los modelos para mantener su efectividad ante cambios operativos.

**Fase c) Determinar criterios de éxito en el análisis**

**Métricas y Benchmarking para la evaluación de Modelos.**

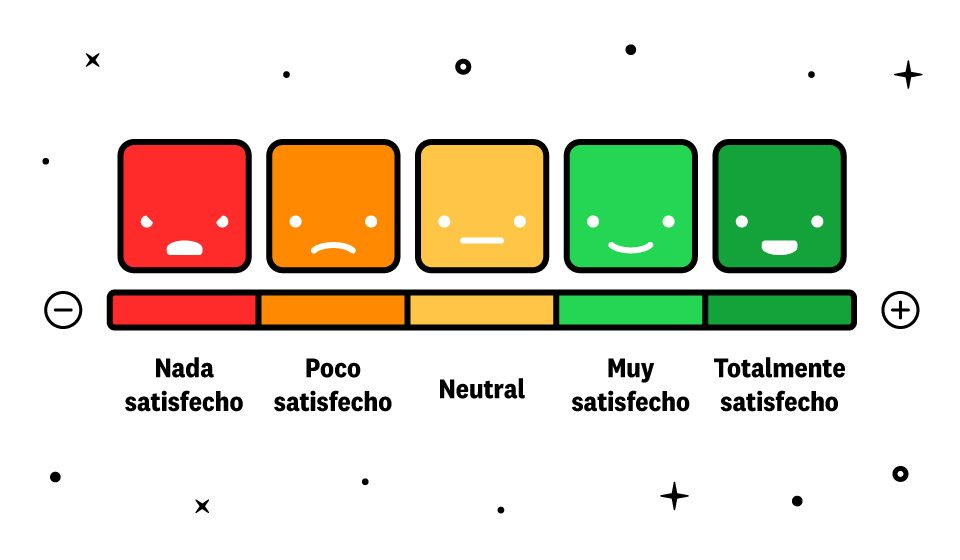
|  |  |  |
| --- | --- | --- |
| **Métrica** | **Descripción** | **Objetivo** |
| Precisión (Accuracy) | Proporción de predicciones correctas sobre el total de predicciones. | Superior al 85% para identificar patrones de consumo normal y anomalías. |
| Tasa de Falsos Positivos | Porcentaje de detecciones erróneas de anomalías; importante para minimizar alertas innecesarias. | Debe estar por debajo del 5% para evitar una sobrecarga de alertas. |
| Tasa de Falsos Negativos | Porcentaje de anomalías no detectadas; debe ser bajo para evitar riesgos operativos. | Inferior al 10%, dado el impacto crítico de las detecciones. |
| AUC-ROC | Área bajo la curva ROC; evalúa la capacidad del modelo para distinguir entre clases. | Superior a 0.90, indicando un modelo confiable. |
| Tiempo de Respuesta | Tiempo requerido para detectar y alertar sobre anomalías; crítico en sistemas de detección en tiempo real. | Latencia inferior a 5 segundos para notificaciones en tiempo real. |
| Recall (Sensibilidad) | Capacidad del modelo para identificar anomalías verdaderas; un alto recall es esencial. | Superior al 90% para asegurar la correcta identificación de anomalías. |
| F1 Score | Métrica que combina precisión y recall; útil para balancear falsos positivos y negativos. | Superior a 0.85 para lograr un equilibrio efectivo entre precisión y sensibilidad. |

**Métricas Subjetivas de Éxito**

* Satisfacción del Equipo Operativo: La facilidad para interpretar y actuar sobre las alertas generadas es crucial. Un modelo complejo o confuso podría no ser considerado exitoso, a pesar de su precisión.

Método de Medición:

Encuestas o cuestionarios al equipo operativo para medir la facilidad de uso y comprensión del modelo de la herramienta. Y por ejemplo al equipo operativo de Contugas se le puede medir la satisfacción con la herramienta por medio de las escalas Likert 1 para medir qué tan intuitivo encuentran las alertas generadas de anomalías.



Fuente: Escala de Likert. [Zunzunegui](https://acumbamail.com/blog/author/alicia-zunzunegui/) , A. (2023).

Indicadores:

Porcentaje de usuarios que califican de forma positiva la herramienta, entonces se mide los porcentajes de comentarios positivos versus los negativos en los instrumentos de recolección. Esto en términos de frecuencia de medición: El tiempo en que se puede implementar durante la ejecución y después la aplicación en Contugas.

* Adopción de Decisiones: La efectividad con la que se utilizan los resultados analíticos para la toma de decisiones operativas es fundamental.

Método de Medición:

Análisis de registros operativos para validar y revisar si las decisiones tomadas están respaldadas por los resultados de la herramienta. Además, es importante realizar entrevistas con los responsables de tomar decisiones sobre la utilidad o uso de la herramienta.

Indicadores:

Porcentaje de decisiones operativas a raiz de la herramienta, tiempo entre las alertas generadas por la herramienta y la toma de desición y la frecuencia de medición de estas decisiones puede ser mensuales.

* Capacidad de Adaptación: La flexibilidad del modelo para ajustarse a cambios operativos sin requerir constantes reorientaciones es un factor importante.

Método de Medición:

Llevar registro de pruebas y tiempos requeridos al aplicar o introducir cambios en la herramienta por parte del equipo desarrollador, así se mide la flexibilidad del modelo y de la herramienta

Indicadores:

Tiempo promedio de ajuste de la herramienta y modelo con los cambios o parámetros solicitados. Y este indicador se puede mediar antes y después de que se implementan estos cambios.

**Fase d) Tipos de problemas de Analytics:**

* **Enfoques y técnicas candidatos**: Para abordar el problema de detección de anomalías en el consumo de gas, se proponen técnicas como análisis de series temporales para entender patrones históricos y proyectar comportamientos futuros. Se sugiere el uso de modelos no supervisados, como clustering (k-means, DBSCAN) y autoencoders, que son efectivos para detectar anomalías sin datos etiquetados. También se consideran modelos supervisados, como árboles de decisión o bosques aleatorios, para clasificar consumos y modelos de predicción para anticipar comportamientos futuros.
* **Respuesta a los objetivos de análisis**: Las técnicas de clustering permiten segmentar clientes según su comportamiento de consumo, facilitando la identificación de anomalías. Los autoencoders ayudan a detectar desviaciones en tiempo real al aprender patrones normales. El análisis de series temporales proporciona un marco para entender el comportamiento pasado y prever anomalías, mientras que los modelos supervisados ofrecen un enfoque robusto para clasificar consumos en normales o anómalos basados en datos históricos.
* **Alineación con el problema de negocio**: Los resultados de estas técnicas están alineados con los objetivos de Contugas, mejorando la eficiencia operativa al detectar anomalías en tiempo real, incrementando la satisfacción del cliente al prevenir interrupciones y optimizando recursos para reducir costos.
* **Acciones para implementar técnicas**: Para aplicar estas técnicas, se recomienda implementar un sistema de monitoreo en tiempo real que agrupe los consumos y alerte sobre anomalías. Los patrones aprendidos por autoencoders deben actualizarse periódicamente para detectar desviaciones significativas. Además, se pueden hacer proyecciones de consumo con modelos de series temporales y aplicar algoritmos de clasificación en modelos supervisados para asignar probabilidades de anomalía.

**Fase e) Planteamiento del Problema de Analytics:**

El análisis de datos busca desarrollar un sistema de detección de anomalías en el consumo de gas para Contugas, utilizando modelos de inteligencia artificial y técnicas avanzadas. Este sistema se enfocará en identificar patrones atípicos en tiempo real, facilitando la anticipación de problemas operativos y optimizando recursos. Para ello, se requerirán datos precisos de consumo de gas, presión y temperatura, que deben ser actualizados y etiquetados cuando sea posible. Las métricas de éxito incluirán una tasa de detección de anomalías superior al 85% y una mejora en la eficiencia operativa, lo que permitirá una mejor planificación y reducción de costos.

Los resultados del análisis ofrecerán a Contugas herramientas robustas para gestionar el consumo de gas, permitiendo identificar anomalías de manera proactiva y optimizar la asignación de recursos. Se estima un tiempo de implementación de seis meses, tras los cuales se espera una reducción de costos operativos y una mejora en la satisfacción del cliente. Este enfoque alineará el análisis de datos con los requerimientos del negocio, contribuyendo a la sostenibilidad y competitividad de la empresa en el mercado.

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

1. **Artificial Intelligence-Based Anomaly Detection of Energy Consumption in Buildings: A Review, Current Trends, and New Perspectives. Referencia: Y. Y. Li et al. (2021). Energy and Buildings**

Este estudio revisa el uso de IA para detectar anomalías en el consumo energético de edificios, evaluando técnicas como redes neuronales profundas y SVM. Los autores concluyen que, aunque efectivas, estas técnicas requieren grandes volúmenes de datos y enfrentan desafíos en la heterogeneidad de sistemas energéticos. Este enfoque es relevante para Contugas en la detección de consumos anómalos.

1. **Optimal Control of Cooling Management System for Energy Conservation in Smart Home with ANNs-PSO Data Analytics Microservice Platform. Referencia: S. Sirisumrannukul et al. (2024).** [**Heliyon**](https://www.sciencedirect.com/science/article/pii/S2405844024029682)

Este artículo investiga la optimización del consumo energético en sistemas de refrigeración de hogares inteligentes utilizando ANNs y el algoritmo PSO. Se logró una mejora significativa en la eficiencia energética, lo que sugiere que el uso de ANNs en Contugas podría optimizar el consumo de gas y detectar anomalías.

1. **Enhancing Security and Energy Efficiency of Cyber-Physical Systems using Deep Reinforcement Learning.Referencia: J. Zhang et al. (2024).** [**Procedia Computer Science**](https://www.sciencedirect.com/science/article/pii/S1877050924013735)

Este estudio utiliza DRL para mejorar la seguridad y eficiencia energética en sistemas ciberfísicos. Los resultados muestran una reducción en el uso de energía y una operación más segura, lo cual es pertinente para Contugas al optimizar decisiones en tiempo real para el manejo de anomalías en el consumo de gas.

1. **Detección de anomalías en datos industriales usando deep learning. Referencia: Y. Zhao et al. (2022).** [**IEEE Transactions on Industrial Informatics**](https://ieeexplore.ieee.org/document/10643234)

Este trabajo explora el uso de redes neuronales profundas para detectar anomalías en sistemas IoT industriales, logrando identificar el 90% de las anomalías en tiempo real. Este enfoque puede ser altamente relevante para Contugas, mejorando la gestión de datos en tiempo real de sensores de gas.

1. **Modelos de predicción de consumo de energía para la detección de anomalías. Referencia: X. Li et al. (2020).** [**Applied Energy**](https://www.researchgate.net/publication/349193113_Artificial_intelligence_based_anomaly_detection_of_energy_consumption_in_buildings_A_review_current_trends_and_new_perspectives)

Este estudio utiliza un enfoque de aprendizaje en conjunto combinando modelos como ARIMA y LSTM para detectar patrones anómalos en datos de consumo energético, logrando una tasa de detección del 85%. Este enfoque híbrido es relevante para Contugas, permitiendo una detección anticipada de anomalías en el consumo de gas.