

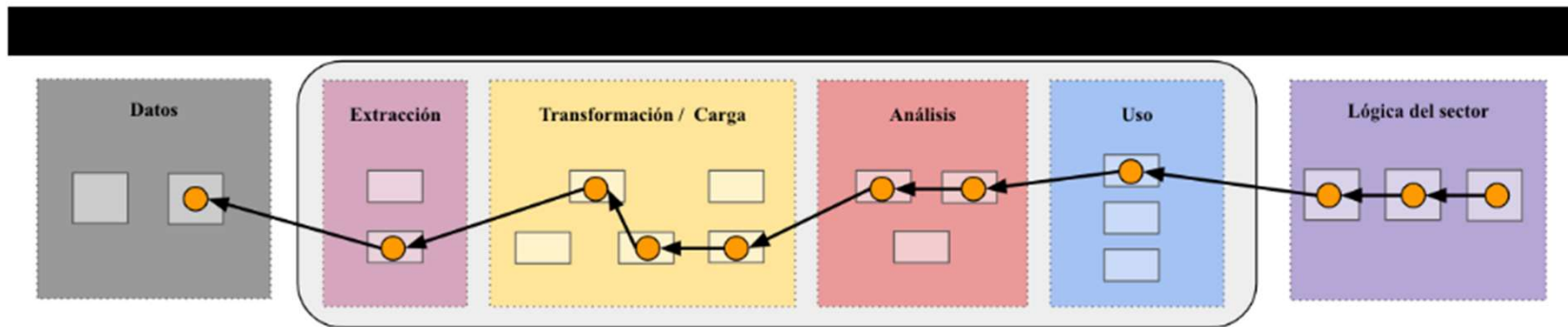


# *Sistema de Detección de Anomalías para Clientes Libres Electro Dunas*

**GRUPO 3**

18 May 2024

# Prototipo Fachada



# Logica del Sector

- ¿A qué audiencia está orientado el producto?
  - Está orientado a los usuarios del área de Operaciones para la detección y control de anomalías en la energía entregada activa y reactiva de los usuarios no regulados de ElectroDunas
- ¿Qué conceptos, procesos, actores, factores o métricas clave están involucrados en el análisis?
  - Stakeholders: Equipo de Operaciones y Clientes no regulados propios y terceros de ElectroDunas. Métrica: cantidad total de clientes según el intervalo de tiempo disponible
  - Consumo de energía activa y reactiva por parte de los clientes no regulados. Métrica: energía activa entregada – kWh, energía reactiva entregada - kVarh
  - Detección de anomalías presentes en el consumo de energía entregada activa y reactiva para cada cliente no regulado. Métrica: conteo de anomalías históricas, cantidad de anomalías pronosticadas, voltaje anormal (porcentaje), nivel de confianza para la detección de anomalías en la energía activa y reactiva.
- ¿Qué requerimientos se deben cumplir y como medir o verificar su cumplimiento?
  - Visualización de datos históricos de consumo de energía.
    - Requerimientos: Almacenamiento y recuperación de datos históricos, visualización gráfica intuitiva con opciones de filtrado.
    - Verificación: Pruebas funcionales y de usabilidad, pruebas de rendimiento para la recuperación de datos.
  - Detección de anomalías en el consumo histórico de energía
    - Requerimientos: Implementación y selección de modelos para identificar y clasificar anomalías históricas por cliente (DBSCAN, Isolation Forest, Método Descriptivo y KernelCPD)
    - Verificación: Se llevarán a cabo una serie de experimentos para cada enfoque con el propósito de determinar el método más adecuado para identificar las anomalías. Los siguientes aspectos serán considerados en la evaluación: porcentaje de anomalías identificadas, interpretabilidad, coherencia interna, robustez y tiempo de ejecución. Adicionalmente, se utilizará una métrica conocida como "precisión subjetiva", que oscila entre 0 y 1, para evaluar la eficacia en la detección de anomalías.
  - Proyección de consumos futuros.
    - Requerimientos: Implementación y selección de modelos de pronóstico diario de consumo (Modelos: LSTM, RNN, Prophet y SRV).
    - Verificación: se llevarán a cabo una serie de experimentos para cada enfoque y se utilizarán las métricas de MAE y el MSE para la selección de la mejor alternativa. El MAE nos proporcionará información sobre el rendimiento general del modelo, mientras que el MSE nos indicará la calidad del ajuste del modelo a los datos.
  - Detección de anomalías en proyecciones de consumo futuro.
    - Requerimientos: Implementación y selección de modelos para clasificar anomalías futuras (Logistic Regression, SVM, Random Forest y Gradient Boosting).
    - Verificación: Se llevarán a cabo una serie de experimentos para cada enfoque y se utilizará el área bajo la curva ROC (AUC) como métrica de evaluación para seleccionar el mejor modelo.
- ¿Qué externalidades o consideraciones (éticas, organizacionales) tener en cuenta?
  - La medición realizada por los contadores o la captura de información pueden presentar problemas de calidad, lo que conlleva a una detección imprecisa de las anomalías.
  - Es crucial garantizar la privacidad de los datos de los clientes. La construcción del tablero debe asegurar que la información del consumo de energía no se utilice de manera inapropiada o se divulgue sin consentimiento.
  - Es fundamental cumplir con todas las regulaciones y normativas pertinentes relacionadas con la privacidad de datos, la seguridad informática y la gestión de la energía.

# Uso

¿Quién es el usuario y como esto afecta el diseño?

- Analistas y operadores que apoyan el área de Operaciones de ElectroDunas

¿Qué tipo de herramienta de uso supone (alerta, dashboard, reporte, appy)?

- Un tablero de control

¿Qué funcionalidades o características debe tener?

- Presenta información relevante de manera visual y fácil de entender, donde la interfaz sea amigable en tanto a uso y navegación, utilizando gráficos, tablas y otros elementos visuales que soporten las respuestas de cara a los usuarios.
- Permite a los usuarios personalizar el tablero según sus necesidades específicas, como seleccionar qué datos mostrar, cómo se presentan y qué métricas se destacan.
- Proporciona datos actualizados con una frecuencia específica, lo que permite a los usuarios tomar decisiones informadas basadas en la información más reciente.
- Permite a los usuarios interactuar con los datos, como filtrar información, cambiar el intervalo de tiempo o profundizar en detalles específicos, para obtener una comprensión más completa y detallada.

# Análisis

¿Qué pregunta(s) ¿Qué ) debe abordar el análisis (y cómo se relacionan con modelos analíticos, descriptivos, predictivos, o prescriptivos)?

1. ¿Cuáles son los patrones de consumo de energía eléctrica de los diferentes clientes no regulados de Electro Dunas?
2. ¿Qué métodos predictivos pueden utilizarse para anticipar el consumo futuro de energía eléctrica de los clientes?
3. ¿Cómo se pueden detectar y evaluar las anomalías en el consumo de energía eléctrica de los clientes?
4. ¿Identificar los criterios relevantes que permitan obtener un equilibrio entre la detección de anomalías extremas y sutiles manteniendo una solución eficiente en términos de precisión y tiempo de ejecución.

¿Qué técnicas, modelos o algoritmos puede involucrar (text-mining, económico, clasificación, pronóstico, análisis de redes, anomalías, simulación, optimización)

Para la pregunta (1) se implementará modelos de series temporales (STL y ETS ) para analizar aquellos momentos en donde cada cliente tiene anomalías en su consumo promedio efectuado.

Para la pregunta (2) se Implementará modelos predictivos que puedan predecir el consumo futuro de energía eléctrica, como las redes neuronales recurrentes (RNN) y los modelos de descomposición de series temporales (LSTM, RNN, Prophet y SRV), en donde se escogerá el que tenga mejor desempeño con métricas del tipo MSE, RMSE y R2 ajustado.

Para la pregunta (3) Vamos a utilizar modelos no supervisados diseñados para la detección de anomalías, como los autoencoders o métodos basados en densidad principalmente DBSCAN y en series de tiempo (Isolation Forest, Método Descriptivo y KernelCPD) . Con el fin de detectar los clientes que clasifiquen como outliers y poder encontrar esas anomalías en el consumo de energía.

Para la pregunta (4) nos apoyaremos con las diferentes métricas comparando diferentes modelos ya mencionados para seleccionar el del mejor desempeño o posiblemente ensamblar modelos si su resultado resulta ser más ajustado a la realidad.

# Transformación / Carga

¿Qué características deben tener los datos que requieren los análisis considerados(formato, tipo de dato, balanceo, errores) para producir conclusiones válidas?

La información de energía activa, energía reactiva, voltaje en la fase A y voltaje en la fase B proporcionada por el medidor de cada cliente se compila en archivos con formato CSV, un archivo por cliente, con un mapeo de campos estandarizado en los archivos.

Columna	Descripción	Tipo
Fecha	Fecha y hora del consumo	Datetime64
Active_energy	Cantidad de energía eléctrica entregada que realiza trabajo efectivo medida en kilovatios-hora	float64
Reactive_energy	Cantidad de energía intercambiada entre la fuente de energía y una carga sin realizar trabajo útil medida en kilovatios-ampere-reactivos-hora (kVarh);	float64
Voltaje_FA	Voltaje en la fase A del sistema trifásico medido en voltios	float64
Voltaje_FC	Voltaje en la fase C del sistema trifásico, medido en voltios	float64

Se observó que 10 clientes tienen datos completos para el periodo entre el 01-Ene-2021 y el 01-Apr-2023, mientras que 5 clientes tienen datos con un nivel de completitud del 99%. Se planea imputar valores faltantes en los clientes con 99% de completitud usando diferentes enfoques como la media móvil y la regresión temporal. Después de seleccionar la mejor técnica, se trabajará con una muestra de 15 clientes. Además, se destaca que otros clientes tienen niveles de completitud menores y se sugiere confirmar las razones detrás de esto antes de imputar datos, para garantizar la integridad y representatividad de los mismos.

En el contexto de un sistema de detección de anomalías, resulta crucial conservar los valores extremos, ya que estos sirven como indicadores, como se ilustra en la figura abajo, de posibles anomalías en el comportamiento del consumo de energía por parte de cada cliente. La identificación de valores atípicos actuales ofrece la oportunidad de descubrir patrones que proporcionan información para abordar y mitigar estas anomalías.

## ¿ Qué transformaciones son necesarias para llevar los datos a la forma requerida?

Se transformarán el campo fecha para poder acceder por esta información por jornada, día, mes y año.

# Extracción

¿Qué fuentes de datos es necesario incluir y cómo se accede a ellas (API, streaming, requiere captcha solicitud)?  
Se encuentran en formato csv y son proporcionados por el área encargada en ElectroDunas.

Se manejará un repositorio de información sharepoint, en donde se cargarán los archivos .csv suministrados por la compañía que contienen la información de los clientes, y posteriormente una vez realizada la limpieza a través de códigos de Python con pandas y numpy quedarán disponibles como entrada para el tablero.

¿Cómo especificar la consulta de los datos (filtros por fecha, lugar, modalidad, bien servicio...)?  
De acuerdo con la periodicidad de actualización o cargue de nuevos registros en los archivos csv, se realizará una validación en cuanto a la última fecha de cargue de información en el tablero y se tomarán aquellos archivos que tengan una fecha posterior a dicha fecha para realizar la actualización con los nuevos datos.

# Prototipo del Dashboard





Inicio

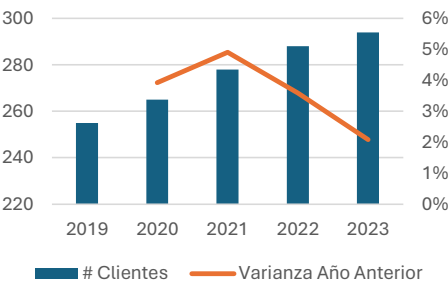
Ficha técnica

Resumen Descriptivo

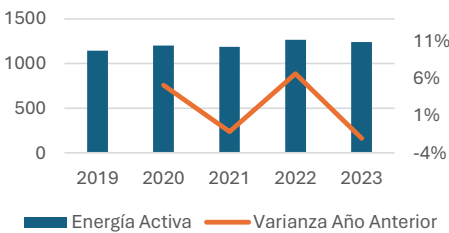
Detección de Anomalías

Cientes Libres de Electrodunas

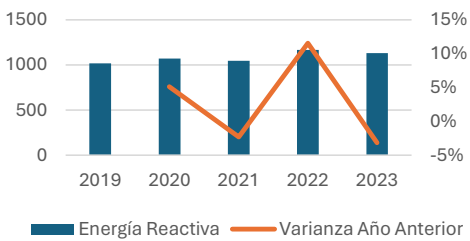
Cientes



Energía Activa



Energía Reactiva



Fecha inicial

Fecha final

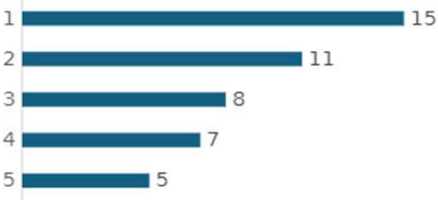
Cientes por Sector

Captación, tratamiento y distribución de agua	Cultivo de Hortalizas	Venta al por mayor de metales y minerales metalíferos
Cultivo de Árboles Frutales y Nueces	Elaboración de cacao y chocolate y de productos de confitería	Cultivo de otros frutos y nueces de árboles y arbustos

Anomalías por Sector



Top 5 clientes - cantidad anomalías detectadas



Fecha actualización: xx-xx-xxxx

Inicio

Ficha técnica

Detección de Anomalías

## Perfil del Cliente

Id cliente

Fecha inicial

Fecha final

Sector Económico

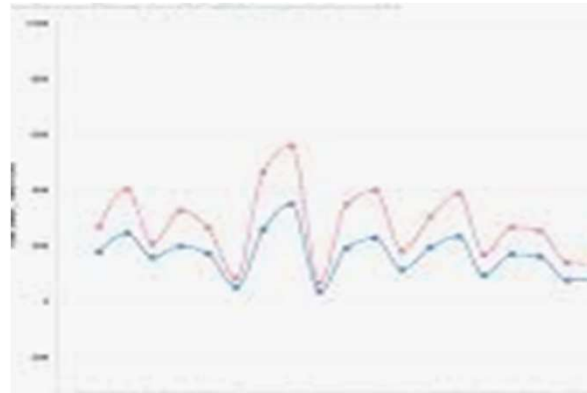
Consumo de  
energía activa

Consumo de  
energía reactiva

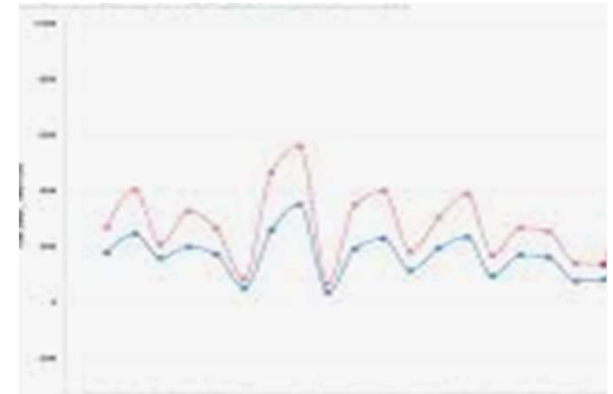
Información  
Disponible

Inicio: dd-mm-yyyy  
Fin: dd-mm-yyyy

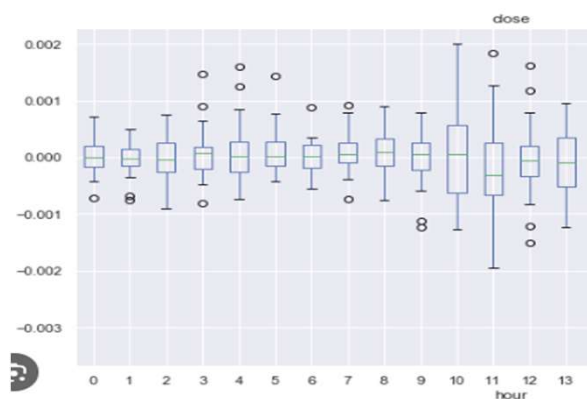
Energía Activa



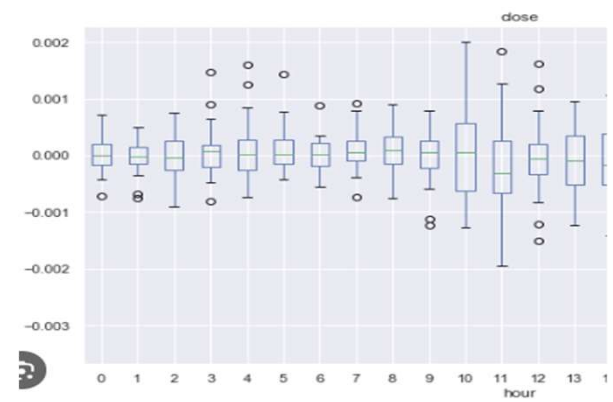
Energía Reactiva



Box Plot mensual energía activa



Box Plot mensual energía reactiva



Fecha actualización: xx-xx-xxxx

Inicio

Ficha técnica

Resumen Descriptivo

Detección de Anomalías

Anomalías por Cliente

Id cliente

Fecha inicial

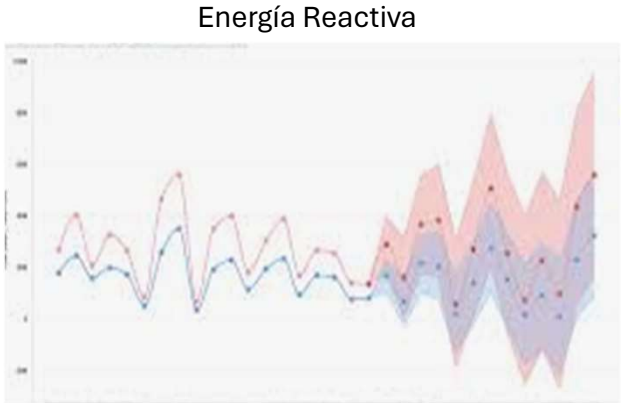
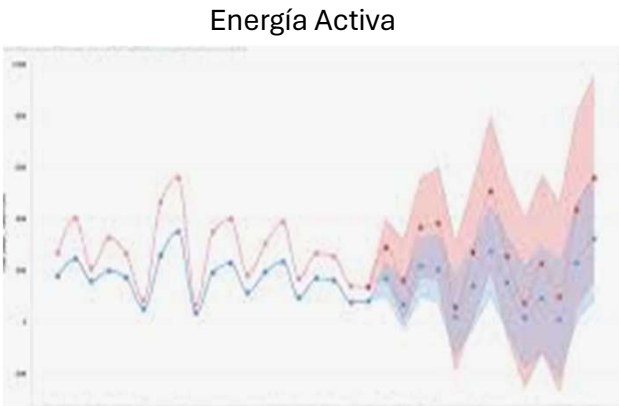
Fecha final

Nivel de confianza inicial

Nivel de confianza final

# Periodos Pronostico  
(1 – 30 días)

Sector Económico	
Consumo de energía activa	
Consumo de energía reactiva	
Posición en el ranking de anomalías	
Cantidad de anomalías detectadas actuales	
Cantidad de anomalías detectadas proyectadas	



Anomalías Históricas

Fecha y hora: 2021-04-05 08:45:00  
Energía Activa - Cliente 16.  
15% menos comparado con la línea base

Fecha y hora: 2024-04-09 10:15:00  
Energía Reactiva - Cliente 16  
25% menos comparado con la línea base

Anomalías Proyectadas

Fecha y hora: 2024-04-05 08:45:00  
Energía Activa - Cliente 16.  
15% menos comparado con la línea base

Fecha y hora: 2024-04-09 10:15:00  
Energía Reactiva - Cliente 16  
25% menos comparado con la línea base

Inicio

Ficha técnica

Resumen Descriptivo

Detección de Anomalías

Ficha Técnica

Información de construcción y que contiene el tablero.