



Manual de Usuario:
**Analítica de Datos para la Detección de Anomalías en los Patrones de
Consumo de Clientes No Regulados en Electroductos**

Natalia Angarita
Johnny Castañeda
Carolina Osorio
Marcell Piraquive

Mayo 22 DE 2024

UNIVERSIDAD DE LOS ANDES
Versión Final

Tabla de contenido

1. Introducción..... 3

2. Instalación y Configuración 4

3. Casos de Uso..... 5

3.1. Cargar nuevos datos 5

3.2. Especificar una consulta..... 5

3.3. Dashboard..... 6

1. Introducción

El artefacto del Proyecto ElectroDunas, es una herramienta analítica diseñada (Dashboard) para permitir a los analistas de pérdidas monitorear el consumo de energía de los clientes no regulados, identificar posibles anomalías y tomar medidas para reducir las pérdidas de energía.

La herramienta permitirá la visualización de información de manera dinámica, con datos históricos segmentados según requerimiento del usuario, cálculos de KPI-Mediciones que facilitarán la toma de decisiones y, además, identificación de posibles anomalías mediante tablas estadísticas. También mostrará un diagrama donde podremos evidenciar la distribución de clústeres que el algoritmo de aprendizaje no supervisado asignó.

Objetivos del artefacto

- Analizar datos históricos de clientes mediante series temporales y permitir segmentación por sectores.
- Monitorear el consumo de energía de clientes no regulados.
- Identificar patrones de comportamiento y anomalías en el consumo.
- Proporcionar herramientas visuales para facilitar la toma de decisiones informadas.

Ventajas

- **Monitoreo eficiente:** Permite un seguimiento detallado del consumo de energía. Además, facilita la segmentación por clientes y por sectores.
- **Detección de anomalías:** Utiliza algoritmos avanzados de Machine Learning para identificar comportamientos anómalos.
- **Visualización interactiva:** Ofrece gráficos en un Dashboard intuitivo y personalizable.
- **Optimización de recursos:** Facilita la identificación de áreas para mejorar la eficiencia energética y reducir costos.

Limitaciones

- **Dependencia tecnológica:** Necesita de una infraestructura tecnológica adecuada, incluyendo acceso a herramientas como Power BI y Snowflake. Asimismo, el equipo de trabajo sugiere otras herramientas para una arquitectura más completa: Amazon Redshift para almacenamiento de datos, Apache Kafka para procesamiento en tiempo real y AWS SageMaker para despliegue de modelos.
- **Curva de aprendizaje:** Los usuarios deben estar familiarizados con conceptos de análisis de datos y herramientas de visualización como Power BI para utilizarlo.
- **Requerimientos de uso:** Para el caso de Power BI, es necesario comprar las licencias y para Snowflake adquirir una suscripción. De igual manera, es mandatorio tener licencias si ElectroDunas desea escalar la solución.

2. Instalación y Configuración

Requisitos previos

- **Hardware:** Computadora con capacidad suficiente para ejecutar Power BI y otras herramientas analíticas.
- **Software:**
 - Power BI Desktop.
 - Python (para procesamiento de datos).
 - Snowflake (para almacenamiento de datos).
 - Amazon Redshift (para almacenamiento de datos) – OPCIONAL, Solución a gran escala.
 - Apache Kafka (para procesamiento en tiempo real) – OPCIONAL, Solución a gran escala.
 - AWS SageMaker (para despliegue de modelos) – OPCIONAL, Solución a gran escala.
- **Conexión a Internet:** Necesaria para la descarga de datos y actualizaciones.

Descarga e instalación

1. **Descarga de Power BI Desktop:**
 - Visite el sitio oficial de Microsoft Power BI y descargue la versión más reciente de Power BI Desktop.
 - [Obtener Power BI Desktop - Power BI | Microsoft Learn](#)
2. **Instalación de Python:**
 - Descargue e instale la última versión de Python desde su sitio oficial. Asegúrese de incluir pip durante la instalación.
 - [Welcome to Python.org](#)

Actualización del sistema

- **Actualizaciones de Power BI:** Siga las notificaciones de Microsoft para mantener Power BI Desktop actualizado.
- **Actualizaciones de scripts Python:** Mantenga actualizados los scripts de procesamiento de datos a través de un sistema de control de versiones como Git.
- **Actualizaciones de Snowflake:** Asegúrese de que la configuración y los datos en Snowflake estén actualizados regularmente.

Conocimientos y habilidades necesarias

- **Análisis de datos:** Familiaridad con conceptos de análisis descriptivo y técnicas de clustering.
- **Herramientas analíticas:** Experiencia previa con Power BI y Python.

3. Casos de Uso

3.1. Cargar nuevos datos

1. Preparación de archivos CSV:

- Asegúrese de que los archivos CSV estén correctamente formateados y contengan todos los campos necesarios.

Los campos esperados del CSV son:

Fecha
Active_energy
Reactive_energy
Voltaje_FA
Voltaje_FC
NumeroCliente
Mes
Día
Hora
Sector Económico

2. Cargar en Snowflake:

- Utilice los scripts Python proporcionados para cargar los datos en Snowflake. Ejecute el script de carga y verifique que los datos se han importado correctamente. Estos scripts serán los encargados de realizar el proceso de ETL al formatear y escalar la data en los inputs necesarios de los algoritmos.

3. Actualizar Power BI:

- En Power BI Desktop, actualice la conexión a la base de datos Snowflake para reflejar los nuevos datos.

3.2. Especificar una consulta

1. Abrir Power BI Desktop:

- Inicie Power BI Desktop y abra el proyecto correspondiente al artefacto (Archivo Dashboard.pbix).

2. Filtrar datos:

- Utilice los filtros de fecha, hora, cliente y sector económico para especificar la consulta deseada.

3. Ejecutar consulta:

- Aplique los filtros y ejecute la consulta para actualizar las visualizaciones con los datos filtrados.

3.3. Dashboard

A continuación, el informe detallado de los aspectos técnicos del Dashboard Ilustración 1:



Ilustración 1. Página 1 del Dashboard interactivo en PowerBI del proyecto ElectroDunas

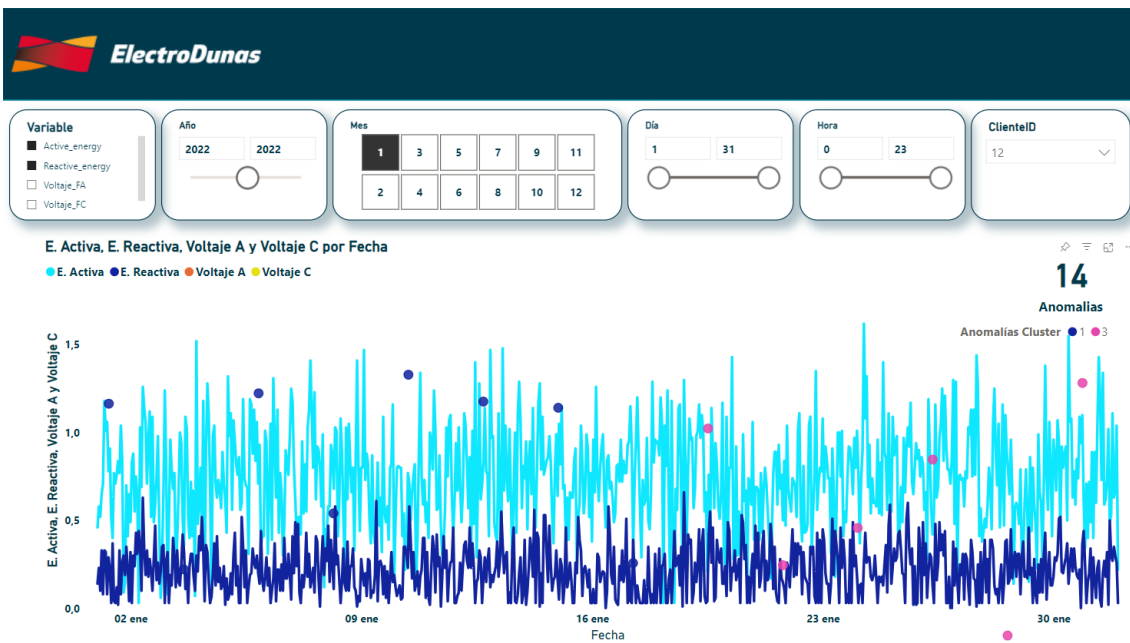


Ilustración 2. Página 2 del Dashboard interactivo en PowerBI del proyecto ElectroDunas

Página 1: Visualización de Energía Reactiva

1. Gráfico de Líneas: Suma de Energía Reactiva por Hora y Mes

- **Descripción:** Este gráfico de líneas muestra cómo varía el promedio de energía reactiva a lo largo del día y los meses.
- **Datos Utilizados:**
 - **Eje Fecha:** Fecha de la observación.
 - **Eje Y:** Suma de Energía Reactiva.
 - **Segmentación:** Meses (1 a 12), Clientes y sector económico.
- **Filtros Aplicados:**
 - **Fecha:** Rango desde 1/1/2021 hasta 4/1/2023.
 - **Cliente:** Todos los clientes o uno específico.
 - **Sector Económico:** Ejemplo, "Cultivo de Hortalizas".
 - **Cluster:** Cluster asignado por el algoritmo de aprendizaje no supervisado.
- **Propósito:** Ayudar a identificar patrones temporales en el consumo de energía reactiva, facilitando la detección de anomalías en diferentes horas del día y meses del año.

2. Indicadores Clave de Rendimiento (KPIs)

- **Penalización:**
 - **Descripción:** Indica el valor de las penalizaciones aplicadas debido a un consumo no adecuado.
 - **Datos Utilizados:** ERi (Energía reactiva inductiva), EAFP (energía activa fuera de punta), EAHP (energía activa presente hora de punta).
 - **Consideraciones:**
 - Aplica sólo cuando $ER \Rightarrow 0$.
 - Para el cálculo, la medición debe estar en unidad de energías: KWH, KVARH. La energía activa y reactiva en hora de punta y fuera de punta se calcula como la suma durante el día:
 - $\sum EA$ en Hora Punta 18:00 hasta 23:00hrs,
 - $\sum EA$ en Fuera de Punta: 00:00 a 17.59hrs y 23:01 hasta 24:00hrs
- **Correlación:**
 - **Descripción:** Mide la correlación entre el consumo de energía reactiva y otros factores.
 - **Datos Utilizados:** Cálculo de correlación entre energía activa y reactiva.
- **Factor de Potencia:**
 - **Descripción:** Indica el factor de potencia, que es una medida de eficiencia energética.
 - **Datos Utilizados:** Cociente entre energía reactiva y energía activa

3. Tabla descriptiva

- **Desviación Estándar (kWh):**
 - **Descripción:** Variabilidad del consumo de energía reactiva.
 - **Datos Utilizados:** Desviación estándar del consumo.

- **Consumo Máximo (kWh):**
 - **Descripción:** Valor máximo de consumo de energía reactiva registrado en el período seleccionado.
 - **Datos Utilizados:** Consumo de energía reactiva (máximo).
- **Consumo Mínimo (kWh):**
 - **Descripción:** Valor mínimo de consumo de energía reactiva registrado en el período seleccionado.
 - **Datos Utilizados:** Consumo de energía reactiva (mínimo).
- **Consumo Promedio (kWh):**
 - **Descripción:** Valor promedio de consumo de energía reactiva registrado en el período seleccionado.
 - **Datos Utilizados:** Consumo de energía reactiva (promedio).

4. Gráfico de Dispersión: Clústeres asignados

- **Descripción:** Este gráfico muestra los clústeres que asignó el algoritmo de aprendizaje no supervisado.
- **Datos Utilizados:**
 - **Eje X:** Energía activa.
 - **Eje Y:** Energía reactiva.
- **Filtros Aplicados:**
 - **Fecha:** Rango desde 1/1/2021 hasta 4/1/2023.
 - **Cliente:** Todos los clientes o uno específico.
 - **Sector Económico:** Ejemplo, "Cultivo de Hortalizas".
- **Propósito:** Identificar posibles comportamientos anómalos mediante un agrupamiento por similitud con un algoritmo de aprendizaje no supervisado.

Página 2: Detección de anomalías por Isolation Forest

1. Filtro de Fecha

- **Descripción:** Permite seleccionar un rango de fechas específico para analizar los datos.
- **Datos Utilizados:** Fechas de los registros de consumo.
- **Propósito:** Filtrar los datos del dashboard según el período de interés, facilitando el análisis temporal.

2. Filtro de Hora

- **Descripción:** Permite seleccionar una hora específica del día para analizar los datos.
- **Datos Utilizados:** Hora de los registros de consumo.
- **Propósito:** Filtrar los datos del dashboard según la hora del día, permitiendo un análisis detallado de las variaciones horarias.

3. Filtro de Cliente

- **Descripción:** Permite seleccionar un cliente específico para analizar los datos.
- **Datos Utilizados:** ID de Cliente.
- **Propósito:** Filtrar los datos del dashboard según el cliente, facilitando un análisis específico por cliente.

4. Filtro de Sector Económico

- **Descripción:** Permite seleccionar un sector económico específico para analizar los datos.
- **Datos Utilizados:** Sector Económico.
- **Propósito:** Filtrar los datos del dashboard según el sector económico, permitiendo un análisis sectorial.

5. Gráfico de líneas

- **Descripción:** Este gráfico de líneas muestra cómo varía la energía reactiva a lo largo del tiempo y señala con puntos las anomalías detectadas por el algoritmo de Isolation Forest, cruzando con el clúster asignado por el algoritmo de aprendizaje no supervisado.
- **Datos Utilizados:**
 - **Eje Fecha:** Fecha de la observación.
 - **Eje Y:** Energía Activa, Energía Reactiva, Voltaje FC y Voltaje FA.
 - **Segmentación:** Meses (1 a 12), Clientes y sector económico.
- **Filtros Aplicados:**
 - **Fecha:** Rango desde 1/1/2021 hasta 4/1/2023.
 - **Cliente:** Todos los clientes o uno específico.
 - **Sector Económico:** Ejemplo, "Cultivo de Hortalizas".

Análisis Técnico

1. Modelo de Datos:

- **Tablas Utilizadas:** Tablas que contienen registros horarios y mensuales de consumo de energía reactiva, junto con información de clientes y sectores económicos.
- **Relaciones:** Las tablas están relacionadas mediante claves como ClienteID y Fecha, permitiendo un análisis integrado de los datos.

2. Medidas Calculadas:

- **Promedio de Energía Reactiva:** Función DAX **AVERAGE** para calcular el promedio de energía reactiva.
- **Consumo Máximo y Mínimo:** Funciones DAX **MAX** y **MIN** para calcular los valores máximos y mínimos de consumo.
- **Penalización:** Calculada mediante la fórmula:
$$ER(\text{penalización}) = ER_i - 30\%(EAFP + EAHP).$$

Donde:

- ERi=Energía reactiva inductiva.
- EAFP=energía activa fuera de punta.
- EAHP= energía activa presente hora de punta.

- **Correlación:** Calculada utilizando la función DAX **CORREL**.
- **Factor de Potencia:** Calculada realizando el cociente entre energía reactiva y activa.
- **Desviación Estándar:** Calculada utilizando la función DAX **STDEV**.

3. Visualizaciones:

- **Gráficos de Líneas:** Utilizados para mostrar tendencias a lo largo del tiempo.
- **Indicadores KPIs:** Utilizados para mostrar métricas clave de rendimiento.
- **Gráficos de Dispersión:** Utilizados para visualizar el agrupamiento realizado por el algoritmo.