**Logotipo

Descripción generada automáticamente**

**Título del trabajo**

Pronóstico de generación de energía reactiva en una planta de productos lácteos a partir del año 2020

Javier Andrés Causil Martínez

Yohiner Andrés Borja Góez

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
Nombres completos, Título académico más alto

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2024

Tabla de contenido

[1. Descripción del problema 5](#_Toc167743136)

[1.1. Problema de negocio 5](#_Toc167743137)

[1.2. Aproximación desde la analítica de datos 6](#_Toc167743138)

[1.3. Origen de los datos 6](#_Toc167743139)

[1.4. Métricas de desempeño 6](#_Toc167743140)

[2. Objetivos 7](#_Toc167743141)

[2.1. Objetivo general 7](#_Toc167743142)

[2.2. Objetivos específicos 7](#_Toc167743143)

[3. Datos 8](#_Toc167743144)

[3.1. Datos originales 8](#_Toc167743145)

[3.2. Datasets 9](#_Toc167743146)

[3.3. Analítica descriptiva 11](#_Toc167743147)

[Referencias 16](#_Toc167743148)

**Lista de tablas**

[Tabla 1. Variables disponibles en la tabla iot\_power\_meters 8](#_Toc167742374)

[Tabla 2. Tabla original de valores de consumo de energía 10](#_Toc167742375)

[Tabla 3. Tabla depurada de valores de consumo de energía 10](#_Toc167742376)

[Tabla 4. Identificación de valores atípicos utilizando el rango intercuartil 15](#_Toc167742377)

**Lista de figuras**

[Figura 1. Correlación gráfica de variables de consumo de energía 12](#_Toc167742531)

[Figura 2. Correlación de variables de consumo de energía utilizando coeficiente de Pearson 13](#_Toc167742532)

[Figura 3. Distribución de variables de consumo de energía 14](#_Toc167742533)

# Descripción del problema

Una empresa de productos lácteos requiere disminuir los costos asociados al consumo de energía en una planta de producción. Se aplicarán modelos para pronosticar la generación de energía reactiva que afecta el consumo pero no agrega valor a la empresa. Se cuenta con datos del historial de consumo de energía eléctrica desde el año 2020 y con datos capturados por sensores instalados en las máquinas de la planta desde el año 2023. Las métricas de desempeño son las asociadas a la precisión de los modelos y el valor de la facturación mensual.

## Problema de negocio

Una empresa de productos lácteos requiere disminuir los costos asociados al consumo de energía en una planta de producción.

Según (CELSIA, 2022), la energía reactiva es un tipo de energía eléctrica absorbida o inyectada a la red por algunos equipos que para su funcionamiento necesitan un campo magnético, tales como motores, transformadores, ascensores, sistemas de bombeo de agua, motores de aireación de piscinas, iluminación eficiente, entre otros. La unidad de medida de este tipo de energía es kVArh. En complemento, de acuerdo con (EPM, 2023), la energía reactiva se puede entender como una energía que ocupa espacio de las redes eléctricas, pero no es útil a la hora de hacer trabajo. Como esta energía reactiva satura las redes, es necesario para las empresas reducirla a su mínima expresión para evitar problemas en la calidad de la energía, sobrecargas e ineficiencias que redundaría en mayores costos para prestar el servicio.

De acuerdo con lo indicado por el personal técnico de la Planta, en los últimos años, el nivel de generación de energía reactiva en planta ha ido en aumento, y aunque la información de energía global se conoce, no es posible identificar la totalidad de las fuentes que la están generando, lo cual tiene un impacto significativo sobre los costos de producción. La energía reactiva puede ser generada por múltiples fuentes (motores, equipos eléctricos, instalaciones, etc.) y puede variar según el estado de la maquinaria, la tecnología de las máquinas, el mantenimiento y otras variables no identificadas, por lo cual no es posible discriminar el aporte de cada fuente al sobrecosto establecido por el prestador de servicios de energía, lo cual dificulta la toma de decisiones para optimizar su uso.

La Comisión Reguladora de Energía y Gas CREG, mediante Resoluciones 15 de 2018 (CREG, 2018), 199 de 2019 (CREG, 2020) y 195 de 2020 (CREG, 2022), ha emitido los lineamientos con los cuales se determina la variable M que establece el nivel en el que se encuentra una empresa, de acuerdo con la cantidad energía reactiva generada en determinado periodo de tiempo. Esta variable es la que utilizan los prestadores de servicios de energía para establecer las penalizaciones aplicadas por sobrepasar los límites establecidos en la normativa. En el año 2024 la Empresa se encuentra en el nivel 6 de 12 niveles posibles y ha venido en aumento, donde el nivel 12 es el de mayor sanción. Lo anterior se ve reflejado en un aumento en los pagos realizados a EPM asociados a la penalización por exceso de generación de energía reactiva.

La situación se vuelve en un tema crítico para la gerencia de la Planta, al ver que los costos de producción aumentan por la generación de energía reactiva, y no se tiene establecido un plan para controlarla.

## Aproximación desde la analítica de datos

Se aplicarán modelos estadísticos y de machine learning para pronosticar la generación de energía reactiva que afecta el consumo energético, pero no agrega valor a la empresa. Se espera que, con los resultados del modelo, el equipo técnico pueda tomar decisiones para controlar la energía reactiva generada que lleve a la reducción del valor de la facturación mensual.

## Origen de los datos

Se cuenta con información de la energía activa y reactiva por hora para el periodo comprendido entre febrero de 2020 y mayo de 2024. Estos datos son proporcionados directamente por la empresa prestadora de servicios de energía eléctrica y son un instrumento para el control del pago de la facturación mensual.

Adicionalmente, se tienen mediciones de los sensores de los equipos, los cuales toman los datos por segundo de valores asociados al estado de los equipos para el seguimiento en la producción desde el año 2023, los cuales son almacenados en la nube a través del servicio de AWS.

## Métricas de desempeño

Se considerarán las siguientes métricas de desempeño del modelo:

* Error absoluto medio (MAE): Mide la diferencia promedio absoluta entre los valores pronosticados y los valores reales. Un MAE bajo indica mayor precisión.
* Error cuadrático medio (RMSE): Similar al MAE, pero considera los errores al cuadrado, penalizando más los errores grandes. Un RMSE bajo indica mayor precisión.
* Error porcentual absoluto medio (MAPE): Expresa el error en términos porcentuales, siendo útil para comparar series de tiempo con diferentes escalas. Un MAPE bajo indica mayor precisión.

Como métrica del negocio se considerará el valor de la facturación mensual.

# Objetivos

## Objetivo general

Desarrollar un modelo utilizando técnicas estadísticas y de aprendizaje automático para pronosticar la generación de energía reactiva en una planta de productos lácteos, con datos del consumo energético en el periodo 2020 a 2024, que permita disminuir los costos asociados al consumo de energía en la planta.

## Objetivos específicos

* Realizar los pasos contemplados en la metodología CRISP-DM para la creación de los modelos de analítica
* Evaluar la precisión de los modelos aplicados considerando errores inferiores al 30%
* Establecer, mediante el uso del modelo, los días del mes que hacen que la generación de energía reactiva sobrepase los 12 días permitidos.

# Datos

## Datos originales

Se dispone de dos fuentes de datos.

Fuente de datos 1: se cuenta con la información para 1404 días del consumo energía activa y reactiva para el periodo comprendido entre febrero de 2020 y mayo de 2024. Estos datos son proporcionados directamente por la empresa prestadora de servicios de energía eléctrica y son un instrumento para el control del pago de la facturación mensual. Se cuenta con un archivo en excel por año. Cada archivo contiene cinco variables de las cuales se tiene información para cada hora, así:

* + Energía Activa Consumo (kWh)
  + Energía Activa Generación (kWh)
  + Energía Reactiva Inductiva (kVarh)
  + Energía Reactiva Capacitiva (kVarh)
  + Fecha

Fuente de datos 2: Mediciones de los sensores de equipos de la empresa productora de lácteos. Los datos de los sensores están alojados en DynamoDB con una actualización segundo a segundo. Los sensores toman los datos de ciertas métricas asociadas al estado de los equipos para el seguimiento en la producción, los cuales son almacenados en la nube a través del servicio de AWS. Actualmente el acceso a las bases de datos se encuentra limitado por parte de la oficina de TI. Una vez se restablezca el servicio, se procederá a realizar el análisis correspondiente de los datos disponibles. Se está a la espera de la entrega del diccionario de datos generados por los sensores por parte del equipo de TI. Las variables que contempla esta fuente de datos se relacionan en la tabla 1 y su análisis se realizará para una entrega posterior.

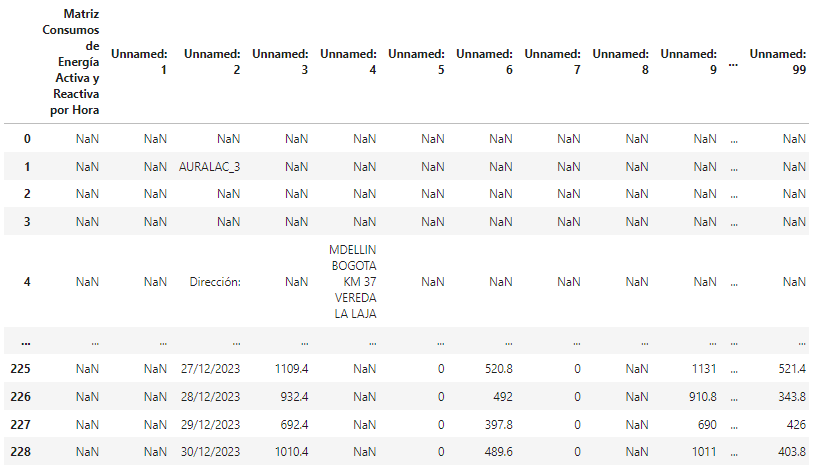
Tabla 1. Variables disponibles en la tabla iot\_power\_meters



## Datasets

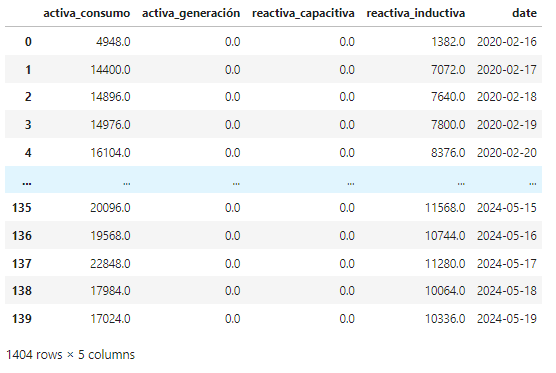
El análisis se realiza para el dataset de la fuente de datos 1. Debido a que los archivos de cada año se encuentran en Excel, es necesario extraer la información de cada uno de ellos para consolidar una tabla global que consolide la información de todos los años. Se hace necesario identificar las celdas que contienen información y eliminar columnas con valores nulos. Con la información de las tablas se hace necesario consolidar para cada variable la información disponible en horas y sumar los valores correspondientes para tener la información completa por día. La estructura de los datos originales se muestra en la tabla 2.

Tabla 2. Tabla original de valores de consumo de energía



Una vez realizada la limpieza de datos se obtiene un dataset consolidado por días para los valores de la energía. El resultado se observa en la tabla 3.

Tabla 3. Tabla depurada de valores de consumo de energía

****

## Analítica descriptiva

Como se observa en la figura 1, se identifica una posible relación directa entre las variables activa\_consumo y reactiva\_inductiva. Lo anterior se valida con el análisis de correlación de la figura 2.

Al realizar el análisis de correlación por medio del coeficiente de Pearson en la figura 2, se valida lo identificado gráficamente pues se obtuvo, para las variables activa\_consumo y reactiva\_inductiva, un coeficiente de correlación de 0.81, el cual se considera significativamente alto. Esto nos lleva a analizar la posibilidad de eliminar alguna de las dos variables. Sin embargo se hace necesario contrastarlo con el personal técnico de la empresa para conocer detalles de los expertos antes de tomar la decisión.

Figura 1. Correlación gráfica de variables de consumo de energía

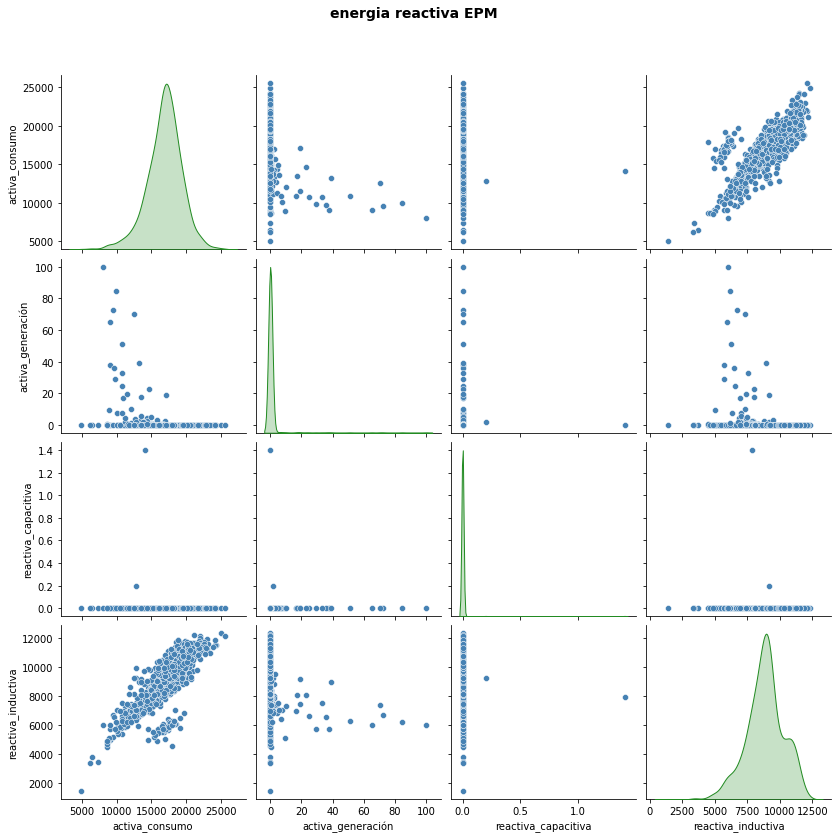
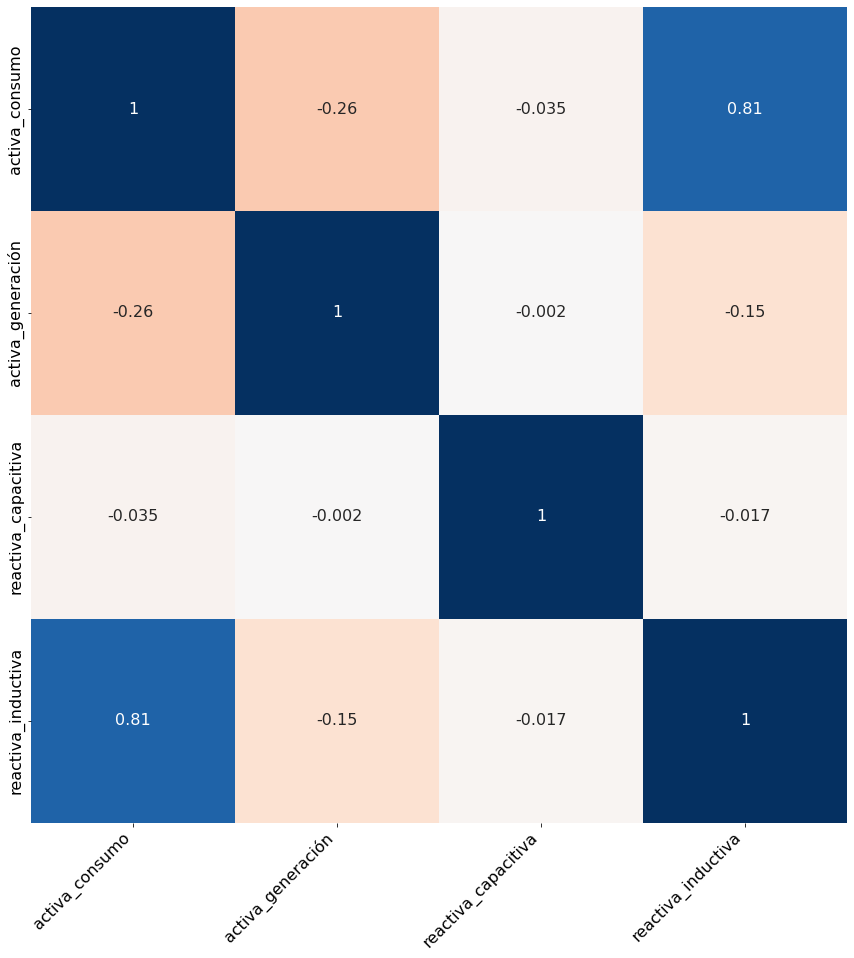


Figura 2. Correlación de variables de consumo de energía utilizando coeficiente de Pearson



De acuerdo con la información disponible en la figura 3, se observa para cada una de las variables, lo siguiente:

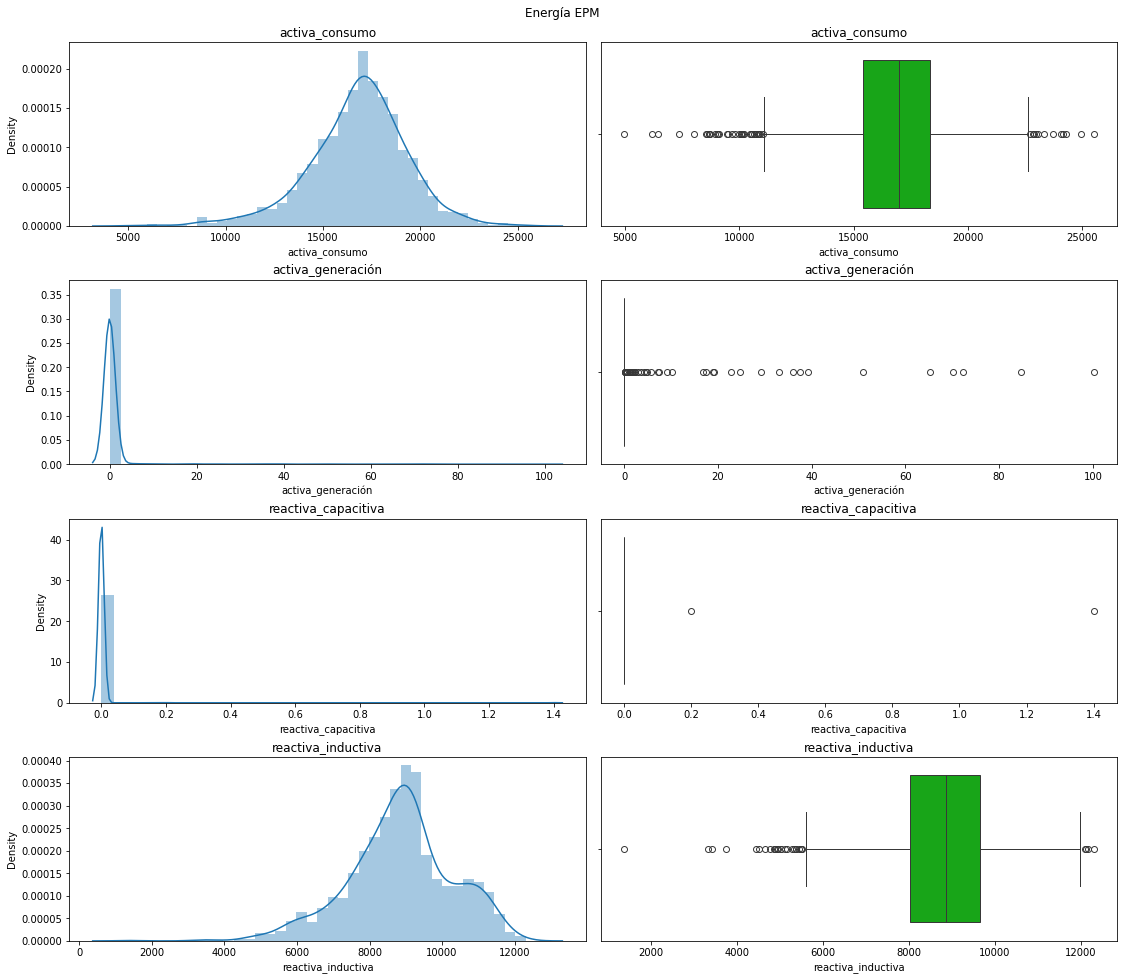
activa\_consumo: se observa una posible distribución normal de los datos con media en 17.500. Se identifica la existencia de algunos datos por fuera de los percentiles 25 y 75.

activa\_generación: se observa una clara concentración de los datos alrededor de cero con presencia de posibles datos atípicos muy alejados de este valor.

reactiva\_capacitiva: se observa una clara concentración de los datos alrededor de cero. Se identifican 2 posibles valores atípicos en valores de 1.2 y 1.4.

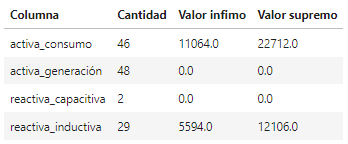
reactiva\_inductiva: se observa que los datos posiblemente no provengan de una distribución normal, por la presencia de una cola a la izquierda. Se identifican posibles datos atípicos alejados del percentil 25.

Figura 3. Distribución de variables de consumo de energía



Para la identificación de datos atípicos se tomaron aquellos valores que se encuentran un 50% por encima y por debajo del rango intercuartil de cada una de las variables. El resumen se observa en la Tabla 4. La decisión de eliminarlos está condicionada a la validación por parte de los expertos del negocio.

Tabla 4. Identificación de valores atípicos utilizando el rango intercuartil



# Referencias

CELSIA. (2022, febrero 20). *¿Qué es la energía reactiva?* Celsia. Retrieved abril 27, 2024, from https://www.celsia.com/en/blog-celsia/conoce-todo-lo-que-debes-saber-sobre-energia-reactiva-por-que-se-refleja-en-tu-factura-la-energia-reactiva/

CREG. (2018, febrero 03). *Alejandría - Resolución 15 de 2018 CREG*. Gestor normativo CREG. Retrieved abril 19, 2024, from https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_creg\_0015\_2018.htm

CREG. (2020, enero 20). *Alejandría - Resolución 199 de 2019 CREG*. Gestor normativo CREG. Retrieved abril 19, 2024, from https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_creg\_0199\_2019.htm

CREG. (2022, octubre 22). *Alejandría - Resolución 195 de 2020 CREG*. Gestor normativo CREG. Retrieved abril 19, 2024, from https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_creg\_0195\_2020.htm

EPM. (2023). *Energía reactiva*. EPM. Retrieved abril 19, 2024, from https://www.epm.com.co/clientesyusuarios/energia/tarifas-energia/energia-reactiva.html