**Título del Proyecto**: Pronóstico de generación de energía reactiva en una planta de productos lácteos.

Seminario de Analítica y Ciencia de Datos

Resumen Descriptivo del Proyecto

[Elabore un párrafo corto describiendo de manera más madura las características del

mismo. Ésta ya es una referencia, debido a que en el momento evaluativo anterior ud ya se

esforzó por “vender la idea” de manera clara]

Basado en el Data Product Canvas

**Objetivo:** Desarrollar un modelo para pronosticar con precisión la generación de energía reactiva en una planta de productos lácteos.

**Motivación:** La generación de energía reactiva en plantas industriales como las de productos lácteos puede generar costos adicionales y afectar la eficiencia energética. Un modelo de pronóstico preciso puede ayudar a la planta a optimizar su consumo de energía, reducir costos y mejorar su sostenibilidad.

**Beneficios potenciales:**

* Reducción de costos de energía reactiva.
* Optimización del consumo de energía.
* Mejora de la eficiencia energética.
* Mayor sostenibilidad de la planta de productos lácteos.

**Recursos necesarios:**

* Datos históricos de consumo de energía.
* Información sobre la producción y variables ambientales.
* Conocimiento en aprendizaje automático y pronóstico de series de tiempo.
* Herramientas computacionales (Python).

**Resultado esperado:**

Modelo preciso haciendo uso de herramientas de analítica que pueda pronosticar la generación de energía reactiva en la planta de productos lácteos de Auralac.

1

resuelve:

**1. Problema.**

a. ¿Cuál es mi problema alrededor de mi proyecto?

Según (CELSIA), 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), 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 redundarían en mayores costos para prestar el servicio.

De acuerdo con lo indicado por el personal técnico de Auralac, 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.

b. ¿Por qué se manifiesta el problema?

La Comisión Reguladora de Energía y Gas CREG, mediante Resoluciones 15 de 2018 (CREG), 199 de 2019 (CREG) y 195 de 2020 (CREG), 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 Auralac 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.

c. ¿De quién es el problema?

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

**2. Datos.**

**a. Describa la fuente de datos (Sea más específico que en documento pasado)**

De las reuniones que se han realizado el encargado de la infraestructura nos ha aclarado que la empresa cuenta con una infraestructura On-premises y se encuentran migrando a la nube de AWS. Actualmente el origen de los datos proviene de sensores con una actualización segundo a segundo, esta información es alojada en DynamoDB.  
  
El equipo desarrollador de Auralac, nos dio acceso a la nube mediante un usuario de AWS con ciertos privilegios, que nos permiten interactuar con la base de datos no relacional DynamoDB, se está tramitando un diccionario de datos en el cual se explique cada contexto de las variables, por el momento se tiene una relación de las tablas sobre las cuales se tiene acceso para el proyecto, en las cuales se identifica el tipo de colecciones y su relación con la planta de producción.

* **auralac\_iot\_calderas:** Esta colección se encarga de almacenar la información que los sensores captan de la pasteurización a través de vapor.

| **Columna** | **Tipo** |
| --- | --- |
| equipo | str |
| ts | datetime |
| combustion\_air\_temp | float |
| efficiency | float |
| firing\_rate | float |
| oxigeno\_actual | float |
| setpoint | int |
| stack\_temp | float |
| steam\_pres | float |
| steam\_total | int |
| vfd\_frequency | float |
| water\_outdoor\_ret\_temp | float |

* **auralac\_iot\_esterilizadores:** Está colección se encarga de almacenar los datos de la pasteurización de la leche.

| **Nombre de la columna** | **Tipo** |
| --- | --- |
| equipo | str |
| ts | datetime |
| h00c01bt335 | float |
| h00e21bf100 | float |
| h00e51bt125 | float |
| h00h51bt080 | float |
| h00h51bt085 | float |
| h00h61bp711 | float |
| h00w01bf050 | float |
| seq\_step\_no | int |

* **auralac\_iot\_pasteurizadores:** La colección aloja la cantidad de producto que se ingresa en el equipo

| **Nombre de la columna** | **Tipo** |
| --- | --- |
| equipo | str |
| ts | datetime |
| cream\_step\_no | int |
| entregado\_cuajada\_total | float |
| entregado\_leche\_total | float |
| entregado\_quesito\_total | float |
| entregado\_termizada\_total | float |
| entregado\_yogurt\_total | float |
| lp3fit01 | int |
| lp3fit041 | int |
| lp3le065 | float |
| lp3te040 | float |
| lp3te043 | float |
| lp3te044 | float |
| lp3te045 | float |
| lp3te052 | float |
| lp3te064 | float |
| lp3te071 | float |
| main\_step\_no | int |
| plan\_cuajada | int |
| plan\_leche | int |
| plan\_quesito | int |
| plan\_termizada | int |
| plan\_yogurt | int |
| recibido\_fit01\_cuajada | int |
| recibido\_fit01\_leche | int |
| recibido\_fit01\_quesito | int |
| recibido\_fit01\_termizada | int |
| recibido\_fit01\_yogurt | int |

* **auralac\_iot\_power\_meters:** Almacena la información de la energía de los equipos utilizados para abastecer la planta, en está colección se encuentra el histórico de la energía reactiva.

| **Nombre de la columna** | **Tipo** |
| --- | --- |
| equipo | str |
| ts | datetime |
| corriente\_a | float |
| corriente\_b | float |
| corriente\_c | float |
| corriente\_media | float |
| distorsion\_media\_total | float |
| energia\_activa | float |
| energia\_aparente | float |
| energia\_reactiva | float |
| factor\_potencia | float |
| potencia\_activa | float |
| potencia\_aparente | float |
| potencia\_reactiva | float |
| tension\_media | float |
| voltaje\_ab | float |
| voltaje\_bc | float |
| voltaje\_ca | float |

* **auralac\_iot\_silos:** La colección almacena las variables que alojan la cantidad de leche por silo y la temperatura a la que se encuentra.

| **Nombre de la columna** | **Tipo** |
| --- | --- |
| equipo | str |
| ts | datetime |
| as1lit01 | float |
| as1tit01 | float |

* **auralac\_iot\_yogurteras:** Contiene las variables que describen el peso que tiene el equipo y su temperatura.

| **Nombre de la columna** | **Tipo** |
| --- | --- |
| equipo | str |
| ts | datetime |
| ly11tit01 | float |
| ly11wt01 | float |

b. ¿Cuál es la calidad que posee dicha fuente?

Con ayuda de chatGPT podemos aproximarnos al concepto de calidad de datos si tenemos las siguientes características:

* **Exactitud:** Los datos que se han explorado son correctos y concuerdan con el tipo de dato correspondiente y tienen una gran veracidad ya que son directamente proporcionados por sensores que a su vez tienen una monitorización para su buen funcionamiento.
* **Integridad:** En la exploración que hemos realizado las variables necesarias están presentes.
* **Consistencia:** Los nombres de las variables muestran una estructura relacionada a su entorno, por ejemplo tit significa temperatura, wt peso, fit flujo no acumulativo, así sucesivamente de las reuniones que hemos realizado hemos visto que estás describen la zona, el equipo y el tipo, entre otras características. La documentación de las variables está en proceso de creación y una vez lo tengan terminado nos lo pasan para tener un mejor contexto de las variables en cada colección.
* **Actualidad:** La actualización de los datos, de las colecciones que nos han compartido cuentan con actualización segundo a segundo. Ellos no has explicado qué es importante ya que si la leche cambia de temperatura en una etapa del procedimiento el producto ya no sirve, por lo cual es crucial la actualización constante de los datos.
* **Disponibilidad:** La fuente de datos se encuentra totalmente disponible para su uso, manteniendo la debida precaución.

c. ¿Es accesible y disponible?  
Dado que la empresa se encuentra en una migración hacía la nube y nos han otorgado un usuario a la plataforma AWS, se evidencia accesibilidad y disponibilidad de la data compartida.

**3. Hipótesis.**

a. ¿Qué es lo que pretende probar?

* Se pretende probar que la implementación de un modelo de pronóstico de energía reactiva puede contribuir al logro de ahorros significativos en costos energéticos para la planta de lácteos de Auralac.

b. ¿Cuáles pueden ser las respuestas esperadas a lo que pretende probar?

* Planificación: el pronóstico facilita la cuantificación de los costos de producción asociados al consumo de energía reactiva.
* Reducción de costos: el pronóstico contribuye a la toma de decisiones basadas en datos para identificar la necesidad de adquirir equipos para reducir la energía reactiva que se devuelve a la red, considerando el valor de los mismos y la tasa de retorno.
* Rentabilidad y competitividad: Se evaluarán los beneficios económicos y la mejora en la competitividad de la planta como resultado de la implementación del modelo.

c. ¿Qué acciones debo tomar para cada respuesta anterior?

* Crear modelos de pronósticos confiables y validados con las herramientas analíticas disponibles.
* Con base en los pronósticos logrados, cuantificar el valor de la generación de energía reactiva por día con el fin de generar una proyección de los costos energéticos mensuales.
* Crear herramientas de visualización para mostrar los resultados y facilitar la toma de decisiones a la gerencia.

**4. Solución.**

a. Tipo. (¿Qué tipo de algoritmo de ML es? ¿Qué técnica es?

La solución que se pretende aplicar es un modelo tipo SARIMA y una red neuronal. SARIMA es una técnica estadística para analizar y pronosticar series temporales que presentan patrones estacionales y no estacionarios.

b. ¿Cómo debe ser implementado?

Segun chatGPT debemos implementar las siguientes etapas para implementar una red neuronal

* Preparación de los datos
* Diseño de la arquitectura de la red neuronal.
* Inicialización de parámetros.
* Definición de la función pérdida y el optimizado.
* Entrenamiento del modelo.
* Validación y ajuste de hiperparámetros.

Según chatGPT para implementar un modelo SARIMA se debe

* Preparación de los datos.
* División de datos.
* Identificación de parámetros.
* Ajuste del modelo.
* Validación del modelo.
* Refinamiento del modelo.
* Predicción futura.
* Análisis de residuos.

c. ¿Qué resultados se esperan?

Se espera poder realizar predicciones de la energía reactiva que al equipo técnico le permita tomar decisiones para controlar la energía reactiva y poder crear planes para controlarla.

**5. KPI’s**

a. ¿Cómo espera evaluar el modelo?

Una vez aplicados los modelos, se seguirán los siguientes pasos para su evaluación:

* **Evaluar el rendimiento:** Calcular las métricas aplicables de acuerdo con la técnica utilizada.
* **Visualizar los resultados:** Graficar los pronósticos junto con los valores reales para observar su desempeño visualmente.
* **Pruebas de diagnóstico:** Realizar pruebas estadísticas como la prueba de Dickey-Fuller o ADF para evaluar la estacionariedad de los residuos del modelo.
* **Selección del mejor modelo:** Comparar el rendimiento de los diferentes modelos utilizando las métricas aplicadas, seleccionando el que mejor se ajuste a los datos y genere pronósticos precisos.
* **Refinamiento del modelo:** Ajustar los hiperparámetros del modelo seleccionado para optimizar su rendimiento.

b. ¿Qué métricas debe usar?

De acuerdo con (Mauricio), algunas de las métricas utilizadas para evaluar el rendimiento de una serie de tiempo serían las siguientes, serían los siguientes:

* **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.

Adicionalmente, se podrían considerar:

* **Coeficiente de correlación de Pearson:** Mide la correlación lineal entre los valores pronosticados y los valores reales. Un coeficiente cercano a 1 indica una fuerte correlación positiva, mientras que un coeficiente cercano a -1 indica una fuerte correlación negativa.
* **Índice de determinación R^2:** Indica la proporción de la variabilidad de los valores reales que se explica por los valores pronosticados. Un R^2 alto indica un mejor ajuste del modelo.

**6. Actores.**

a. ¿Quién es su cliente?

AURALAC S.A. Es una empresa que hace parte del grupo QBCo dedicada a la producción de lácteos (Leche, bebidas lácteas y yogurt) y de derivados lácteos (crema de leche, cuajada, queso crema, quesito, queso, kumis), con su planta principal ubicada en el municipio de Rionegro.

b. ¿Quiénes son los interesados (Stakeholders)?  
Hasta ahora en las reuniones con el equipo desarrollador hemos podido identificar tres principales interesados

* El gerente.
* El líder del equipo de desarrollo.
* El líder del equipo de IOT.

c. ¿Quién usará la solución?

La solución la usará el equipo de soporte técnico IOT y el equipo de desarrollo.

d. ¿A quién impactará?

La solución impactará primeramente a la empresa EPM, ya que se espera que al tener una predicción en la variable de la energía reactiva se pueda tomar acción para disminuir la generación energía reactiva, lo cual le beneficia a las redes eléctricas.

1

https://medium.com/@leandroscarvalho/data-product-canvas-a-practical-framework-for-building-high-

performance-data-products-7a1717f79f0

Trabajos citados

CELSIA. “¿Qué es la energía reactiva?” *Celsia*, 02 febrero 2022, 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/. Accessed 20 April 2024.

CREG. “Alejandría - Resolución 15 de 2018 CREG.” *Gestor normativo CREG*, 03 febrero 2018, https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_creg\_0015\_2018.htm. Accessed 19 April 2024.

CREG. “Alejandría - Resolución 195 de 2020 CREG.” *Gestor normativo CREG*, 22 octubre 2022, https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_creg\_0195\_2020.htm. Accessed 19 April 2024.

CREG. “Alejandría - Resolución 199 de 2019 CREG.” *Gestor normativo CREG*, 02 enero 2020, https://gestornormativo.creg.gov.co/gestor/entorno/docs/resolucion\_creg\_0199\_2019.htm. Accessed 19 April 2024.

EPM. “Energía reactiva.” *Clientes y usuarios*, 2023, https://www.epm.com.co/clientesyusuarios/energia/tarifas-energia/energia-reactiva.html. Accessed 19 April 2024.

Mauricio, José Alberto. *Introducción al análisis de series temporales*. Madrid, Universidad Complutense de Madrid, 2007, https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IAST-Libro.pdf. Accessed 20 04 2024.

OpenAI. (2024). ChatGPT (Version 2) [Modelo de lenguaje]. <https://openai.com/research/gpt-3-5>

**Rúbricas**

R01: Completitud y consistencia: el estudiante realiza una entrega del momento evaluativo

con todos los elementos desarrollados, manteniendo una relación coherente entre los

elementos solicitados. El estudiante usa una consistencia argumentativa que incluye el uso

de referencias cuando hay mérito de autoría entre otros.

R02: Pertinencia: el proyecto del estudiante está dentro del ámbito de la analítica y la

ciencia de datos y es evidente el desarrollo de cada uno de los elementos del Data Product

Canvas

R03: Solución: el estudiante es capaz de responder a las preguntas del Data Product

Canvas con la consistencia necesaria para relacionarla con el proyecto que desarrolla. Hay

una relación evidente entre ellos. Las herramientas de MAchine Learning son una solución

viable dentro de la descripción de la necesidad y objetivos del proyecto.

R04: Datos: los datos sugeridos son pertinentes al proyecto y al alcance delimitado en el

momento evaluativo anterior.