Mayo 2024

**\*\*Objetivos\*\***

3.1 Establecer las medidas de calidad del modelo de aprendizaje automático.

3.2 Proporcionar un marco de referencia para evaluar y mejorar modelos más avanzado

.

**\*\*\*Desarrollo\*\***\*

Para nuestro modelo baseline vamos a elegir un modelo de Random Forest para la demanda (volumen\_despachado) y un modelo genético para la determinación del precio. Elegimos este tipo de algortimo porque nos facilita capturar las relaciones de mayor complejidad entre variable dependiente y variables independientes. De igual manera, el Random Forest ayuda a la reducción de *\*overfitting\**. También porque nos permite identificar las variables importantes en nuestra predicción, de tal manera que podamos  eliminar aquellas que no afectan significativamente a la variable objetivo y puedan causar ruido en la predicción si son incluidas. De igual manera, utilizamos un modelo genético debido a su capacidad de capturar relaciones complejas entre las variables, la capacidad de incorporar restricciones a las diferentes variables del negocio, la penalización a la variable objetivo, que en este caso evita que el precio óptimo se encuentre en un rango fuera de lo normal (es decir, evita que el precio óptimo sea absurdamente alto porque a mayor precio mayor rentabilidad, y lo ubica dentro de un rango razonable dentro de las datos observados).

Para nuestro modelo de optimización de precios de gasolina, reiteramos que vamos a trabajar sobre 3 tipos de combustible por separado (MAGNA, PREMIUM, DIESEL), de tal manera que obtendremos 3 modelos diferentes de optimización de precios.

Para los features de nuestro modelo estamos incluyendo precio\_neto, costo\_neto, precio\_brent (rezagado), y variables temporales de hora, día y mes. Nuestra variable objetivo es el volumen despachado o como lo definimos la demanda de la gasolinera.

Antes de desarrollar el modelo realizamos una codificación de las variables temporales para capturar el ciclo temporal en los datos. Esto se realiza con el seno y coseno de las variables temporales de hora, día y mes.

Otra variable que estamos agregando para el modelo genético es la elasticidad de la demanda. Esta variable también es relevante porque nos ayudará a determinar como cambia la demanda cuando el precio se modifica. La elasticidad es importante porque nos da información sobre el producto, por ejemplo, si una pequeña variación del precio genera un cambio significativo en la demanda entonces el producto se considera elástico); o pese a un cambio importante en el precio, la cantidad demandada se mantiene igual, el producto se considera inelástico. Para obtener la elasticidad de la demanda estamos realizando una regresión linear con los logaritmos de nuestros datos de precio\_neto y volumen\_despachado.

Una vez que obtenemos el RandomForest del volumen despachado, procedemos a realizar un algoritmo genético para la optimización de precio. Dentro del modelo genético estamos generando individuos, población, cruce, mutación y selección. Una vez que definimos estos parametros procedemos a calcular el margun bruto utilizando nuestra base de datos de entrenamiento, nuestro modelo RandomForest, la elasticidad del precio (que calculamos con regresión lineal) y el precio actual. Por último, se incluye una penalización calculada en función a la diferencia del precio\_pred con el precio actual.

*\*¿El modelo está sub/sobreajustando los datos de entrenamiento?\**

*\*¿Cuál es la métrica adecuada para este problema de negocio? \**

*\*¿Cuál debería ser el desempeño mínimo obtener?\**