

Contexto

Compañía de comestibles desea pronosticar las ventas (en unidades) de sus 2 productos estrella para el próximo mes. El objetivo del presente proyecto es encontrar el mejor modelo de serie de tiempo para realizar dicha labor.

1) Tipos de modelos a probar.

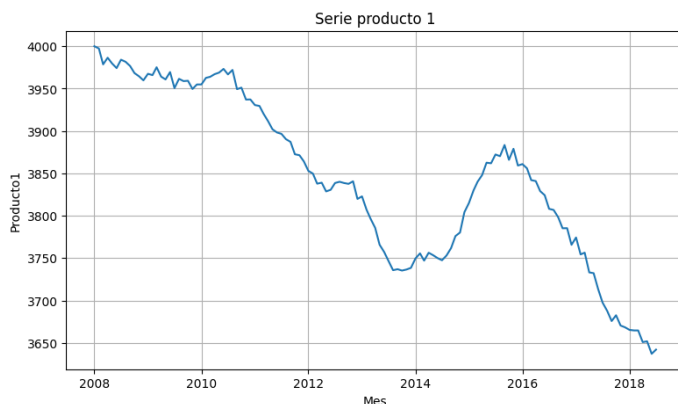
Para realizar el pronóstico sugerido, se probaron los siguientes tipos de modelos:

- Promedio Móvil (pm).
- Suavización exponencial (ETS).
- Regresión lineal (ols).
- Arima.

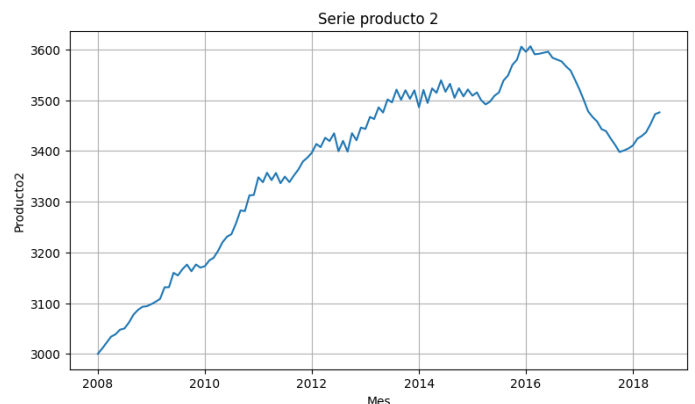
2) Protocolo de evaluación.

El protocolo elegido por temas computacionales y de velocidad del algoritmo, fue ventana móvil. Después de realizar una primera corrida de optimización bayesiana con todos los modelos y sus respectivos hiperparámetros, se definen los parámetros para protocolo de evaluación. Es importante aclarar que individualmente cada modelo presenta resultados diferentes respecto a la mejor configuración del protocolo de evaluación, sin embargo, se estandariza el protocolo de evaluación a las características Tamaño de muestra y Cantidad de meses para pruebas, del mejor modelo para hacer una comparación equitativa y sin asimetrías respecto al resto de modelos. También, hay que tener en cuenta que este protocolo de evaluación no aplica para el modelo de promedio móvil (pm).

En las gráficas siguientes se puede visualizar las series temporales, representando la evolución de las ventas en unidades de los productos a lo largo del tiempo:



- ✓ Desde 2008 hasta aproximadamente 2013, se observa una caída continua en el valor del producto.
- ✓ Entre 2013 y 2016, hay una recuperación del valor.
- ✓ Después de 2016, se observa otra caída pronunciada, hasta llegar a niveles bajos en 2018.



- ✓ Desde 2008 hasta 2016, se observa una tendencia ascendente consistente.
- ✓ Después de 2016, hay una caída considerable, seguida por una pequeña recuperación alrededor de 2018.

Durante la optimización bayesiana, se probaron diferentes tamaños de muestra, incluyendo intervalos de 48, 72, 96 y 127 meses, y se evaluaron periodos de prueba de 6, 12, 18 y 24 meses. Al finalizar el proceso, se determinó que los mejores hiperparámetros para cada producto, de acuerdo con el protocolo de evaluación, son los siguientes:

Producto	Tamaño muestra analizada	Tamaño de la ventana móvil	Cantidad de meses (test_period)
Producto 1	96	90	6
Producto 2	48	42	6

3) Esquema de optimización: parámetro de escogencia del mejor modelo e hiperparámetros

Para realizar este paso, se usó una segunda corrida de optimización bayesiana (luego de haber encontrado y estandarizado los parámetros tamaños de muestra y ventana móvil para el protocolo de evaluación), buscando minimizar el **parámetro RMSE**, es decir que se encontró el modelo con sus hiperparámetros respectivos, con menor error cuadrático medio, bajo un protocolo de evaluación estandarizado descrito en el punto anterior.

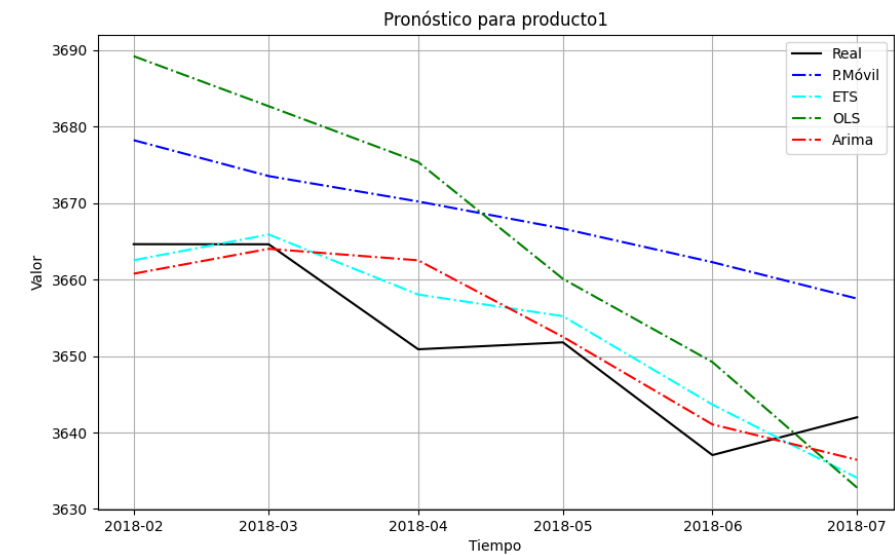
Los valores listados en el presente documento se presentan a 4 decimales, en el jupyter notebook se pueden visualizar con todos los decimales.

4) Comparativo RMSE

Los resultados y mejores parámetros obtenidos bajo el protocolo de evaluación y esquema de optimización mencionados son los siguientes:

Producto1

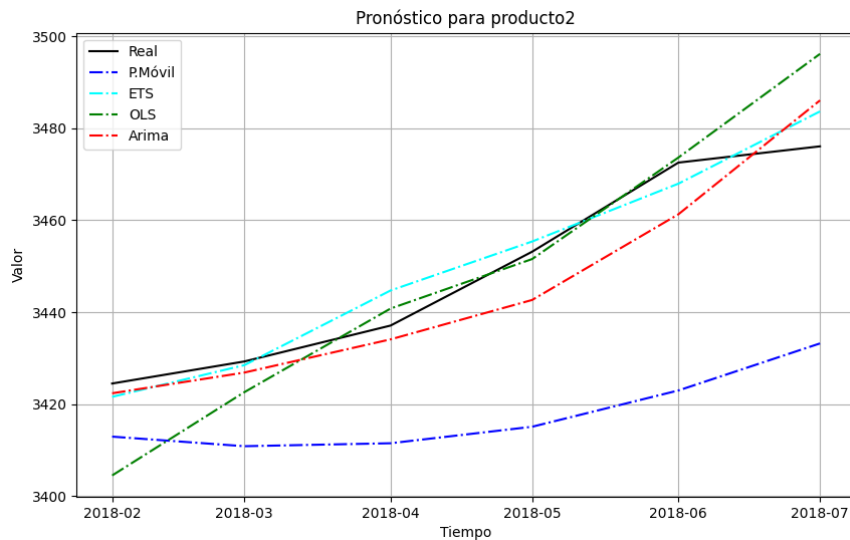
Ranking	Tipo modelo	Tamaño muestra	Test period	RMSE	Hiperparametros
1	ETS	96	6	5.4044	Error: Add Trend: None Seasonal: Add Alpha: 0.7741 Gamma: 0.3295
2	ARIMA	96	6	5.7481	p: 3 d: 1 q: 1
3	PM		6	17.0111	Ventana: 7 Períodos Mínimos: 2
4	OLS	96	6	17.4691	Estacionalidad: Si Grado Polinómico: 7



El mejor modelo para pronosticar el producto1 es ETS, con un RMSE de 5.4044. Los hiperparámetros utilizados para lograr este nivel de error son: Error: "Add", Trend: "None", Seasonal: "Add", Alpha: 0.7741 y Gamma: 0.3295.

Producto2

Ranking	Tipo modelo	Tamaño muestra	Test period	RMSE	Hiperparametros
1	ETS	48	6	5.0117	Error: Add Trend: Mul Seasonal: Mul Alpha: 0.5746 Beta: 0.6970 Gamma: 0.0826
2	ARIMA	48	6	7.6979	p: 1 d: 0 q: 3
3	OLS	48	6	12.0082	Estacionalidad: No Grado Polinómico: 6
4	PM		6	33.8228	Ventana: 7 Períodos Mínimos: 2



El mejor modelo para pronosticar el producto2 es ETS, con un RMSE de 5.0117. Los hiperparámetros utilizados para lograr este nivel de error son: Error: "Add", Trend: "Mul", Seasonal: "Mul", Alpha: 0.5746, Beta: 0.6970 y Gamma: 0.0826.

5) Estimaciones

Usando el mejor modelo ETS, el pronóstico de ventas en unidades con un nivel de confianza del 95% para los productos 1 y 2, en el mes de agosto de 2018 el siguiente:

Producto	Pronóstico	lower_95	upper_95	Comentarios
Producto1	3643.1146	3460.9589	3825.2703	<p>El pronóstico para el Producto 1 es de 3643.11, que parece estar alineado con la continuación de esta tendencia bajista observada hacia el final de la serie.</p> <p>La caída constante en la gráfica sugiere que este producto podría continuar con una tendencia bajista a corto plazo, aunque el intervalo de confianza refleja la posibilidad de una fluctuación en torno a esa caída.</p>
Producto2	3491.8669	3317.2736	3666.4603	<p>El pronóstico para el Producto 2 es de 3491.87, lo cual sugiere que el valor futuro del producto se estabilizará alrededor de ese nivel. El intervalo de confianza al 95% es más ajustado que el del Producto 1, lo que implica una menor incertidumbre en la predicción.</p> <p>Lo anterior sugiere un comportamiento más estable, con una tendencia general positiva a pesar de la caída reciente. El pronóstico también refleja esta estabilidad, y el intervalo de confianza ajustado implica menos incertidumbre.</p>

6) Conclusiones

Tras realizar múltiples iteraciones con diferentes modelos y sus respectivos hiperparámetros, se puede concluir lo siguiente:

- El modelo ETS se destacó por ofrecer el mejor rendimiento al trabajar con los datos proporcionados en el archivo "examen.csv".
- A lo largo del análisis de las series de tiempo, el modelo ETS superó en desempeño a los demás modelos evaluados.
- Una ventaja clave del ETS es su versatilidad para hacer pronósticos, ya que no requiere la validación de supuestos estadísticos complejos, facilitando su uso en una amplia variedad de casos.
- En el análisis de series de tiempo, disponer de un mayor número de divisiones o "splits" de los datos no siempre resulta en una mejora en el error cuadrático medio (RMSE). Es importante no incorporar información innecesaria que no contribuya al modelo.
- La optimización bayesiana ha demostrado ser una herramienta eficiente para la búsqueda de hiperparámetros, superando en efectividad a la búsqueda exhaustiva o por fuerza bruta, ya que acelera significativamente el proceso de convergencia hacia los resultados óptimos.

El notebook se encuentra en el siguiente repositorio: https://github.com/ajapontes/FDA_2_Parcial_1