**PARCIAL PRÁCTICO DE SERIES DE TIEMPO**

**Contexto Inicial**

En este documento se realiza un ejercicio de aplicación de diferentes metodologías de series de tiempo con el objetivo de predecir el comportamiento de las unidades vendidas de dos de los productos más importantes de una empresa de comestibles. Para este caso, se emplearon cuatro técnicas diferentes que abarcan desde enfoques simples hasta métodos que requieren un análisis más riguroso y detallado, en el tratamiento de hiperparámetros y la evaluación de supuestos. Entre las mencionadas se encuentran: **Medias Móviles**, **Suavización Exponencial de Holt-Winters**, **Modelos de Regresión** y **Modelos ARIMA**. Este análisis se trabajará a partir de un conjunto de información que contiene 127 observaciones que datan del mes de enero de 2008.

**Metodología**

En esta sección se presentan los principales pasos realizados para desplegar la aplicación y ejecución de los modelos propuestos. Se abordarán las estrategias de manipulación o preparación de los datos, las técnicas aplicadas para el tratamiento de los conjuntos de entrenamiento y validación y el despliegue de los modelos.

1. **Preparación de los Datos**

Se llevó a cabo el establecimiento de la columna ‘Date’ como índice principal de la serie, de tal forma que se asegure un manejo adecuado de las fechas que marcan la evolución de ambos productos. Esto asegura el correcto funcionamiento de los análisis posteriores y las visualizaciones.

Adicionalmente, dado el contexto de los datos, se formateo el 01 de enero de 2008 como punto de partida y se configuraron los valores como unidades diarias desde la fecha mencionada.

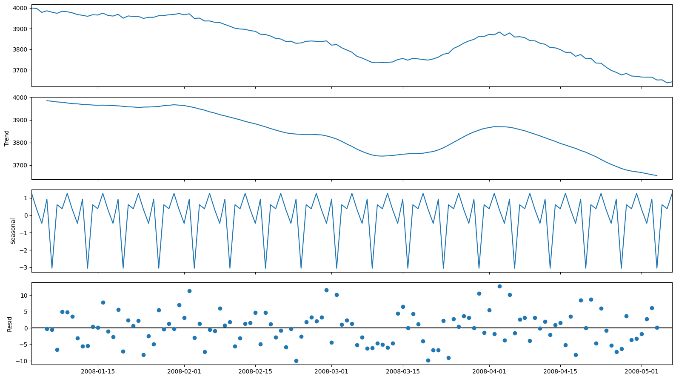
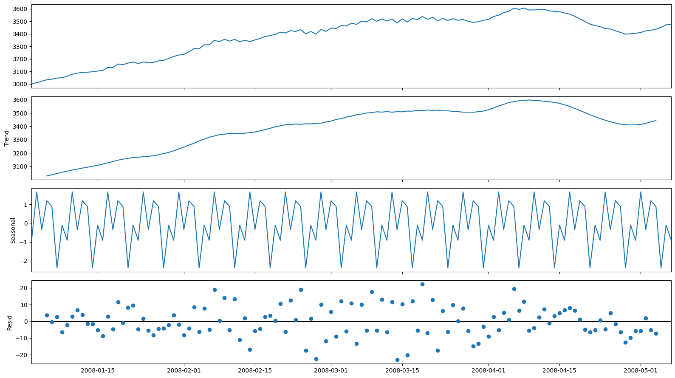
1. **Análisis Exploratorio de Datos**

En términos de la descomposición de las series, se encontró para el *producto 1* la evidencia de una tendencia en mayor medida decreciente, aunque se observa una ligera estabilización seguida por un pequeño incremento antes de una nueva caída. De otro lado, se identificó la presencia de estacionalidad, pues este componente exhibe un comportamiento regular con picos y caídas constantes.

Con respecto al *producto 2*, se identificó un comportamiento creciente de las ventas a lo largo del tiempo, sin embargo, se observan algunos altibajos en los valores observados. Adicionalmente, el componente estacional tiene una fuerte presencia en la serie, dado que se observa un comportamiento cíclico a lo largo del tiempo que se repiten de manera consistente en intervalos muy parecidos.

Gráfico 2. Descomposición de la serie de  
 ventas (unidades) del producto 2

Gráfico 1. Descomposición de la serie de   
ventas (unidades) del producto 1



1. **Protocolo de Evaluación**

Buscando que las aproximaciones que se realicen a partir de los modelos sean consistentes en la media que ofrezcan resultados precisos y útiles para el pronóstico del número de unidades vendidas para cada uno de los productos, se optó por escoger un enfoque de protocolo de evaluación basado en ventanas móviles. En este ejercicio se construyó una función denominada *rolling\_window\_evaluation*, en la que se buscó la mayor eficiencia posible a la hora de implementar la técnica en los diferentes modelos propuestos.

Esta función recibe siete hiperparametros principales que son modificables según las necesidades del investigador: model\_class, target\_column, initial\_train\_size, num\_iterations y forecast\_horizon. Adicionalmente, según la característica del modelo a evaluar permite unos argumentos de inicialización y ajuste (init\_args y fit\_args).

Para ilustrar cómo trabaja esta función, se tomará el modelo más básico -*medias móviles (MMM)-*. En términos generales, la función evalúa el modelo en múltiples ventanas de datos, en las cuales calcula un error para cada ventana. Posteriormente, el cálculo del RMSE se lleva a cabo tomando la raíz cuadrada del promedio de los errores cuadráticos medios almacenados en cada iteración. En la construcción del código también se generó una función denominada *plot\_forecast* para graficar los valores reales versus los predichos, la cual se utilizó en la implementación de cada uno de los modelos propuestos.

1. **Comparación de los Modelos**

Para el *producto 1*, el modelo ARIMA (2,2,1) fue el que evidenció el menor RMSE en términos comparativos, lo cual sugiere que pareciera es este modelo el que tendrá la mejor representación de las fluctuaciones en los datos. Por su parte, al evaluar el *producto 2*, el modelo que muestra el mejor desempeño es el ARIMA (1,1,2) con un RMSE de 7,92. Esta comparación, evidencia la necesidad de ajustar y seleccionar los modelos en función de las características específicas de cada una de las series propuestas.

Cabe resaltar que ambos casos, el modelo con el segundo RMSE más bajo corresponde a Holt-Winters. Para esta aproximación, se llevó a cabo un proceso de *optimización bayesiana*, que permitió encontrar los mejores hiperparámetros a partir de un enfoque probabilístico. Dado que los modelos ARIMA requieren la evaluación de supuestos, los modelos de Winters actúan como la mejor opción en caso de que se presenten incumplimientos.

Tabla 1. Comparación de modelos ajustados según RMSE para cada producto

|  |  |  |  |
| --- | --- | --- | --- |
| **Producto 1** | | **Producto 2** | |
| **Promedio Móvil** | 9,68 | **Promedio Móvil** | 19,67 |
| **Holt - Winters** | 8,76 | **Holt - Winters** | 12,11 |
| **Polinomio Grado 11 y Estacionalidad** | 12,58 | **Polinomio Grado 10 y Estacionalidad** | 36,89 |
| **ARIMA (2,1,1)** | 5,39 | **ARIMA (1,1,2)** | 7,92 |

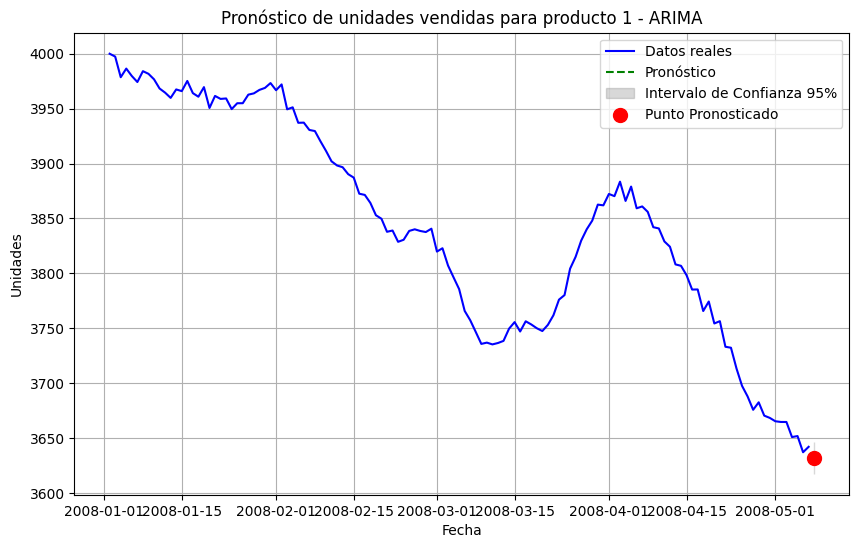
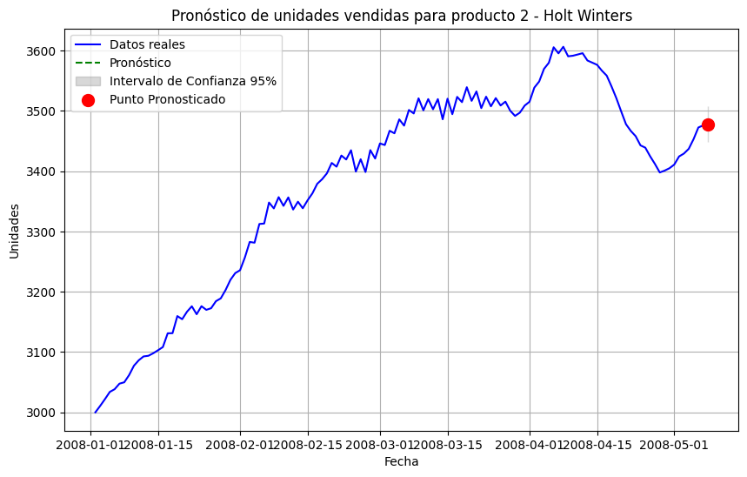
1. **Predicciones y Evaluación de Supuestos**

Haciendo uso de todo el conjunto de datos, para el *producto 1* se hizo uso de un modelo ARIMA (2,1,1) para pronosticar el número de unidades vendidas. Esta decisión se toma, por un lado, teniendo en cuenta el RMSE evidenciado en la sección anterior y por otro, gracias a la evaluación y el buen comportamiento de los supuestos del modelo. En este caso, para los residuales del modelo se identificó un comportamiento normal, la no presencia de autocorrelación y la ausencia de heterocedasticidad. A través de este modelo, se pronosticaron un total de 3.631 unidades vendidas. En el caso del *producto 2*, no se pudo utilizar el modelo ARIMA (1,1,2), a pesar de que mostraba el menor RMSE, ya que incumplió el supuesto de no autocorrelación en los residuos. Por lo tanto, se optó por emplear el modelo Holt-Winters, que pronostica la venta de 3.477 unidades para este producto.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Shapiro** | **lLjung-Box** | **ARCH** |
| **Producto 1** | 0.298 | 0.432 | 0.156 |
| **Producto 2** | 0.187 | 0.00004 | 0.547 |

Gráfico 4. Pronóstico de unidades vendidas para el producto 1 Modelo Holt Winters

Gráfico 3. Pronóstico de unidades vendidas para el producto 1 Modelo ARIMA (2,1,1)



1. **Conclusión**

El análisis de las técnicas de modelado de series de tiempo ha revelado que el modelo ARIMA (2,2,1) es el más preciso para el producto 1, mientras que el modelo Holt-Winters es más adecuado para el producto 2, debido a problemas con los supuestos en el modelo ARIMA (1,1,2). La utilización de ventanas móviles y la optimización bayesiana han sido esenciales para afinar los modelos y mejorar la precisión de las predicciones. Finalmente, elegir y ajustar los modelos de acuerdo con las características de los datos y la validez de los supuestos es crucial para obtener pronósticos precisos y valiosos para la planificación de la empresa en cuestión.