

Predicciones de Ventas con Series Temporales

by

Jared Onan Vilorio Luque

A thesis submitted in conformity with the requirements
for the MSc in Economics, Finance and Computer Science

University of Huelva & International University of Andalusia

uhu.es

un
i **Universidad**
Internacional
de **Andalusia**
A

September 2023

Predicciones de Ventas con Series Temporales

Jared Onan Vilorio Luque

Máster en Economía, Finanzas y Computación

Dr. Nicola Rubino

Universidad de Huelva y Universidad Internacional de Andalucía

2023

Abstract

Being able to predict the future is something that humanity has been working on for decades. For this reason, I have focused my project on developing a sales forecasting system for the company Drogueria Proconsumo. This company is located in Tegucigalpa, Honduras, and is dedicated to the distribution, marketing, and promotion of leading brands of pharmaceutical and popular consumer products in the Honduran territory.

For this purpose, real sales data from the company were used, ranging from January 1, 2017, to July 31, 2023, and the following algorithms were applied: ARIMA, SARIMA, Holt Winters, Prophet and LSTM. The model with the minimum prediction error was Holt Winters, with very good MAPE and MASE values, indicating that the model performs well.

JEL classification: C32, C13, C45, M20, M30.

Keywords: ARIMA, SARIMA, Holt Winters, Prophet, LSTM.

Resumen

Poder predecir el futuro es algo que el ser humano ha venido trabajando durante décadas. Por este motivo, he enfocado mi proyecto en desarrollar un sistema de pronóstico de ventas para la empresa Droguería Proconsumo. Esta compañía se ubica en Tegucigalpa, Honduras y se dedica a la distribución, comercialización y mercadeo de marcas líderes de productos farmacéuticos y de consumo popular en el territorio hondureño.

Para este propósito, se utilizaron las ventas reales de la empresa desde el primero del enero del 2017 hasta el 31 de julio del 2023 y se aplicaron los siguientes algoritmos: ARIMA, SARIMA, Holt Winters, Prophet y LSTM; donde el modelo con el mínimo error de predicción fue el Holt Winters además con MAPE y MASE muy buenos que indican que el modelo es bueno.

Clasificación JEL: C32, C13, C45, M20, M30.

Palabras clave: ARIMA, SARIMA, Holt Winters, Prophet, LSTM.

Agradecimientos

A Dios y Jesus por mi vida y por permitirme concluir una etapa más en vida.

A mi familia por todo el apoyo incondicional, por creer en mí, por siempre motivarme a salir a delante y ser una mejor persona.

A Andrea Paredes por siempre ser una fuente de energía y motivación.

A mi tutor Nicola Rubino por todo el apoyo y horas extra de tutorías, por no limitarme y todas sus exigencias.

Al cuerpo de profesores del Master en Economía Finanzas y Computación.

A la Universidad de Huelva por creer en mi concediéndome una beca de estudios.

Tabla de Contenido

Lista de Tablas	vii
Lista de Figuras	viii
Lista de Anexos.....	x
1 Introducción	1
2 Objetivos	2
2.1 Objetivo General.....	2
2.2 Objetivos Específicos.....	2
3 Revisión de Literatura	3
3.1 Estudios Relativos a Predicción de Ventas/Precios/Demanda.....	3
3.1.1 Predicción Industria Farmacéutica.....	3
3.1.2 Demanda en Industria Plástica.....	3
3.1.3 Demanda en Sector Alimentario	3
3.2 Definiciones	4
3.2.1 Serie Temporal.....	4
3.2.2 Series Temporales (Estacionarias/ No Estacionarias).....	5
3.2.3 Contraste de Raíz Unitaria	8
3.2.4 ARIMA	10
3.2.5 Holt Winters.....	14
3.2.6 Prophet	16
3.2.7 LSTM (Long Short – Term Memory).....	18
3.2.8 Métricas de Evaluación de Modelos	21
4 Minería de Datos	23

5	Desarrollo y Resultados	24
5.1	Recolección de Datos.....	24
5.2	Exploración de Datos	25
5.3	Modelado de Datos	27
5.3.1	Modelo ARIMA y SARIMA	27
5.3.2	Modelo Holt Winters	31
5.3.3	Modelo Prophet.....	31
5.3.4	Modelo LSTM	32
5.4	Resultados	33
5.4.1	Modelo ARIMA.....	33
5.4.2	Modelo SARIMA.....	35
5.4.3	Modelo Holt Winters	36
5.4.4	Modelo Prophet.....	38
5.4.5	Modelo LSTM	39
5.4.6	Comparación de Resultados.....	40
6	Conclusiones	42
7	Referencias.....	43
8	Anexos.....	46

Lista de Tablas

Tabla 1	Términos del modelo holt Winters	16
Tabla 2	Resultados prueba de raíz unitaria	27
Tabla 3	Tabla Resumen modelos ARIMA.....	29
Tabla 4	Tabla de Resultados ARIMA(1,2,1)	29
Tabla 5	Resumen modelo SARIMA	31
Tabla 6	Datos transformados mediante estandarización	32
Tabla 7	Evaluación ARIMA(1,2,1).....	34
Tabla 8	Evaluación ARIMA(1,2,1)(1,0,1).....	36
Tabla 9	Evaluación Holt Winters.....	37
Tabla 10	Evaluación Prophet Facebook.....	38
Tabla 11	Evaluación LSTM.....	40
Tabla 12	Comparación de Todos los Modelos.....	41

Lista de Figuras

Imagen 1.	Serie con tendencia positiva (Universidad de Valladolid, s.f.).....	4
Imagen 2.	Estacionalidad Mensual (Universitat de Barcelona, s.f.).....	5
Imagen 3.	Ejemplo de ciclos anuales (Universidad de Valladolid, s.f.)	5
Imagen 4.	Ejemplo de serie de tiempo estacionaria. Fuente propia.....	6
Imagen 5.	Serie de tiempo NO estacionaria. Fuente Propia	6
Imagen 6.	Proceso de prueba de raiz unitaria	9
Imagen 7.	Función de Autocorrelación Parcial (PACF o FAP).....	12
Imagen 8.	Función de autocorrelación FAS o ACF MA(2).....	13
Imagen 9.	LSTM modelo.....	20
Imagen 10.	Ciclo de Vida de Minería de Datos (IBM, 2021).....	23
Imagen 11.	Conexión al Servidor de Droguería Proconsumo. Fuente propia	24
Imagen 12.	Query de datos de entrenamiento.....	25
Imagen 13.	Query datos de prueba	25
Imagen 14.	Verificación de datos	25
Imagen 15.	Gráfica de Ventas de la empresa.....	26
Imagen 16.	Análisis descriptivo de datos	26
Imagen 17.	Serie original con su gráfica de autocorrelación.....	27
Imagen 18.	Serie diferenciada 2 veces.....	28
Imagen 19.	Autocorrelación Parcial	28
Imagen 20.	Descomposición de la serie de tiempo.....	30
Imagen 21.	Preparación para modelo prophet	32
Imagen 22.	ARIMA(1,2,1) vs ventas reales	33
Imagen 23.	Diagnóstico modelo ARIMA(1,2,1)	34

Imagen 24.	Modelo ARIMA(1,2,1) vs Ventats Reales total serie	34
Imagen 25.	Modelo ARIMA(1,2,1)(1,0,1) vs Ventas Reales	35
Imagen 26.	Diagnóstico ARIMA(1,2,1)(1,0,1).....	35
Imagen 27.	Modelo ARIMA(1,2,1)(1,0,1) vs Ventas Reales total serie	36
Imagen 28.	Modelo Holt Winters vs Ventas Reales	37
Imagen 29.	Modelo Holt Winter vs Ventas Reales total serie	38
Imagen 30.	Modelo Prophet vs Ventas Reales	38
Imagen 31.	Modelo Prophet vs Ventas Reales total serie.....	39
Imagen 32.	Modelo LSTM vs Ventas Reales	39
Imagen 33.	Modelos vs Ventas Reales	40

Lista de Anexos

1. Anexo	Enlace para ver código de Python y HTML del mismo.....	46
----------	--	----

1 Introducción

En la actualidad las empresas en el mundo realizan pronósticos sobre sus ventas utilizando diferentes técnicas, en ciertas ocasiones utilizan promedio de los últimos N periodos o colocan incrementales porcentuales entre otras técnicas simples. Un pronóstico de ventas es una técnica que permite calcular las proyecciones de ventas de una manera rápida y confiable, utilizando como fuentes de datos, ya sea las transacciones de inventarios o la facturación de ventas realizadas. (Toro Ocampo, Mejía Giraldo, & Salazar Isaza, 2004). La predicción de ventas es en pocas palabras el número total de bienes o servicios que se venderán en un período determinado. Esto se realiza con el fin de formar un plan de acción a seguir durante los próximos periodos para prepararse con inventarios, ingresos y la operatividad de la empresa.

Un buen pronóstico ayuda a tomar decisiones financieras y sobre el futuro de la empresa. Ahora, ¿qué es un buen pronóstico de ventas? A lo largo de este documento se desarrollan y explican los distintos modelos de series de tiempo que utilizan algoritmos de aprendizaje automático que procesan la información de ventas reales obtenidas directamente de la empresa, en donde los datos se parten en datos de entrenamiento y datos de prueba.

Una vez que se aplican los modelos es importante realizar las comparaciones de las métricas pertinentes para poder tomar una decisión sobre cuál es el mejor modelo y si el modelo es bueno. A lo largo del documento también se muestran y explican las métricas utilizadas para ver el desempeño de los modelos. Es importante destacar que este proyecto se realiza en base a los datos proporcionados de la empresa, esto no quiere decir que el modelo que mejor se ajuste en este caso será el mismo que se ajuste en otras empresas del sector. En este trabajo se busca la validación interna de la empresa sobre la externa.

2 Objetivos

2.1 Objetivo General

El objetivo general de este Trabajo de Fin de Master es encontrar un modelo de series de tiempo que ayude a Droguería Proconsumo a realizar pronósticos de ventas mensuales con el mínimo error posible para que ayude a tomar decisiones en cara al futuro.

2.2 Objetivos Específicos

- Identificar las fuentes de datos y extraer la información requerida para implementar los modelos de series de tiempo.
- Realizar un procesamiento de datos para asegurar que los datos se adapten a los modelos de series de tiempo que se seleccionen.
- Ejecutar los modelos para obtener nuestras predicciones de ventas y analizar los resultados para identificar al que mejor se adapte al negocio con la mayor precisión posible en donde el error es mínimo.

3 Revisión de Literatura

En esta sección se mencionan algunos artículos relativos a este trabajo de estudio así mismo se definen y describen las técnicas a utilizar, así como otros conceptos de interés.

3.1 Estudios Relativos a Predicción de Ventas/Precios/Demanda

3.1.1 Predicción Industria Farmacéutica

El artículo “Sales Prediction Using ARIMA, Facebook’s Prophet and XGBoost Model of Machine Learning” corresponde a un estudio de la cadena de farmacias Rossmann. Se han elegido el Modelo ARIMA, el Modelo Prophet de Facebook y el Modelo XGBoost para la comparación. Se utiliza el Error Cuadrático Medio de la Raíz (RMSE) para cuantificar la precisión de la predicción debido a la naturaleza de nuestra situación. Los resultados nos muestran que el Modelo ARIMA se desempeña mejor en comparación con los otros. (Ratre & Jayaraj, 2023)

3.1.2 Demanda en Industria Plástica

El artículo llamado “A comparison study between time series model and ARIMA model for sales forecasting of distributor in plastic industry” busca comparar la aplicación de tres métodos de pronóstico en la cantidad del volumen de ventas para el distribuidor de plástico: el método de series temporales ARIMA, el método de promedio móvil y el método exponencial de Holt y Winter. Al final del estudio se concluyó que el modelo ARIMA muestra mejores resultados que otros modelos al compararlo con otros métodos utilizando el MAPE (Error Porcentual Absoluto Medio). (Udom & Phumchusri, 2014)

3.1.3 Demanda en Sector Alimentario

El artículo “Forecasting of demand using ARIMA model” busca demostrar cómo los datos históricos de la demanda pueden ser utilizados para pronosticar la demanda futura y cómo estos pronósticos afectan a la cadena de suministro. La información histórica de la demanda se utilizó para desarrollar varios modelos autorregresivos de media móvil integrada (ARIMA) utilizando el procedimiento de series temporales de Box-Jenkins y se seleccionó el modelo adecuado según cuatro criterios de rendimiento: criterio de Akaike, criterio de Schwarz-Bayesiano, máxima verosimilitud y error estándar. El modelo seleccionado correspondió al ARIMA (1, 0, 1). (Fattah, Ezzine, Aman, Moussami, & Lachhab, 2018)

3.2 Definiciones

3.2.1 Serie Temporal

Una serie temporal (o simplemente una serie) es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes momentos. Matemáticamente una serie temporal univariante se puede representar de la siguiente manera:

$$y_1, y_2, \dots, y_N; (y_t)_{t=1}^N; (y_t: t = 1, \dots, N)$$

En donde Y_t es la observación que se produce en el momento n y N es el número total de observaciones de la serie completa. (Mauricio, 2007).

Es importante tener en cuenta que una serie temporal se puede descomponer en:

- ✚ **Tendencia:** Es el comportamiento de una serie de tiempo. Esta puede ser positiva, negativa o puede no existir. En la imagen inferior se muestra una tendencia positiva.

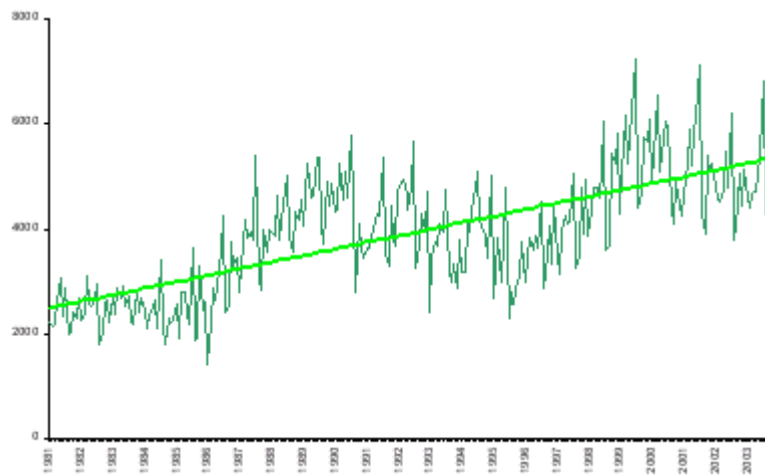


Imagen 1. Serie con tendencia positiva (Universidad de Valladolid, s.f.)

- ✚ **Estacionalidad:** Es cuando podemos observar el mismo comportamiento regular para un periodo de tiempo, por ejemplo, meses o varios meses/ trimestres, en diferentes años. En la imagen inferior se muestra un patrón de crecimiento en el mismo punto en todos los años.

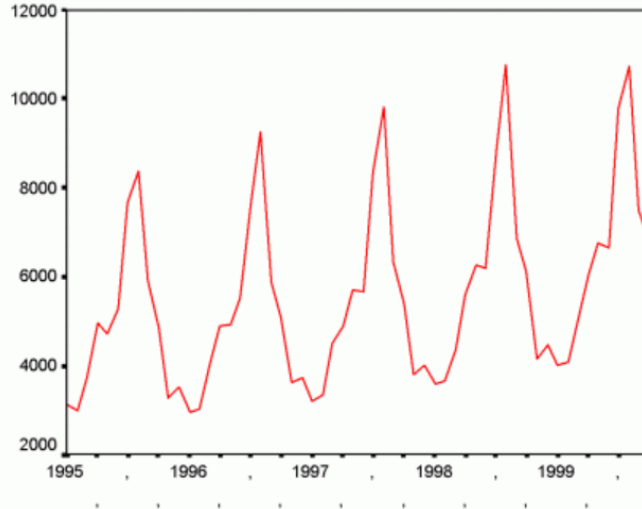


Imagen 2.Estacionalidad Mensual (Universitat de Barcelona, s.f.)

- ✚ Ciclos: Este lo que muestra son los comportamientos recurrentes, que no necesariamente tienen por qué ser exactamente periódicos, con un periodo superior a un año.

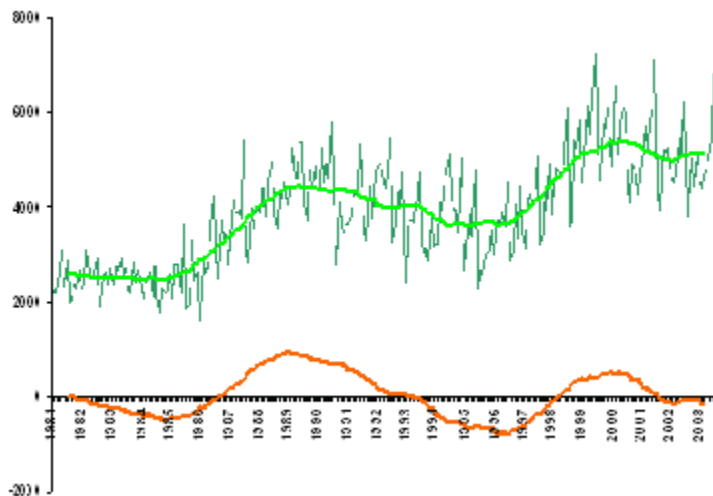


Imagen 3.Ejemplo de ciclos anuales (Universidad de Valladolid, s.f.)

- ✚ Componente Errático: Se le conoce como ruido y este componente explica la variabilidad aleatoria de la serie

3.2.2 Series Temporales (Estacionarias/ No Estacionarias)

Las series temporales se pueden clasificar en (Villavicencio):

- + **Estacionarias.** Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.

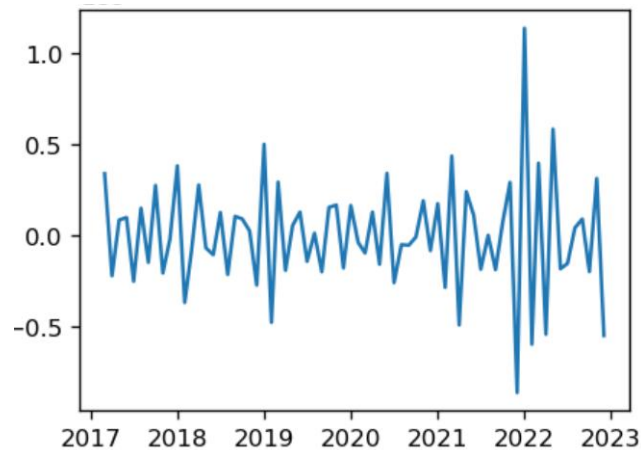


Imagen 4. Ejemplo de serie de tiempo estacionaria. Fuente propia.

- + **No estacionarias.** Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

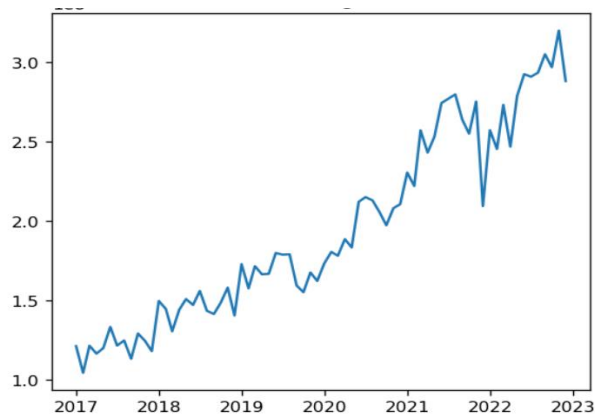


Imagen 5. Serie de tiempo NO estacionaria. Fuente Propia

Nos interesa mucho que las series que trabajemos sean estacionarias. Esto tiene una importante implicación a la hora de predecir. Pues, de tener estacionariedad, estaríamos prediciendo que las características estadísticas de nuestra serie de tiempo serán las mismas en el futuro como en el pasado.

Otra consecuencia importante es que “estacionarizando” una serie de tiempo podemos obtener información significativa como media, varianza y autocorrelaciones (correlaciones entre distintos valores de la serie de tiempo); dichos valores son útiles para describir el comportamiento futuro únicamente si la serie es estacionaria. Por ejemplo, si hay presencia de tendencia positiva, la media y varianza muestral aumentarán y se subestimarán a la media y varianza en periodos posteriores. Y si la media y varianza no están bien definidas, tampoco lo estarán las autocorrelaciones. Por ende, cuando se trabaja con series de tiempo no es recomendable ajustar una regresión, pues los datos no son estáticos (estacionarios).

A grandes rasgos verificar esta propiedad en el estudio de las series de tiempo tiene su importancia en los siguientes aspectos (Ramírez, 2017).

- ✚ Facilidad de Análisis: Entre todos los procesos estocásticos, los procesos estacionarios resultan ser un caso particular mucho más fácil de analizar.
- ✚ Mejor Entendimiento: Aunque en el punto anterior recalamos que puede ser de ayuda para analizar de forma más sencilla algún fenómeno, no es de sorprenderse que también nos permite modelar fenómenos de mayor complejidad. Por ejemplo, la estacionariedad es un supuesto común a la hora de realizar estimación de tendencia, previsión financiera (forecasting), inferencia, entre otros.
- ✚ Ubicuidad: Es decir la presencia de la estacionariedad no sólo en el problema que estemos analizando, sino en el campo de estudio.

En series de tiempo estacionarias nos encontramos con 2 tipos:

- ✚ Estacionariedad Fuerte. Un proceso estocástico es fuertemente estacionario si cumple la siguiente propiedad.

$$F_X(x_{t_1 + \tau}, \dots, x_{t_n + \tau}) = F_X(x_{t_1}, \dots, x_{t_n}) \text{ Con } X \in Z \text{ y } T \in N$$

Lo que quiere decir que la probabilidad de que el proceso tome cierto valor en cada estado, permanece igual con el transcurso del tiempo. Como podemos ver es evidente la fuerza de esta propiedad.

- ✚ Estacionariedad Débil. Un proceso estocástico es débilmente estacionario si:
 - $E[X_t] = \mu; \forall t \in T$ es decir, media constante

- $E[X_{2t}] < \infty$; $\forall t \in T$ es decir que el segundo momento sea finito en cualquier tiempo, esto garantiza que la varianza sea finita.
- $Cov(X_{t_1}, X_{t_2}) = Cov(X_{t_1+h}, X_{t_2+h})$; $\forall t \in N, \forall h \in Z$ la varianza de cada estado del proceso es la misma.

3.2.3 Contraste de Raíz Unitaria

3.2.3.1 Dickey Fuller

La Prueba Dickey Fuller es una prueba que permite determinar la presencia de raíz unitaria en series de tiempo, a partir de esto se demuestra la existencia estadística de la propiedad de estacionariedad o no estacionariedad. Esta prueba es una de las más utilizadas para este fin, siendo importante porque para algunos modelos como los modelos ARIMA se requiere que la serie de tiempo sea estacionaria (Alas, 2021).

Si consideramos un proceso estocástico de la forma: Donde: Y_t es la serie de tiempo y u_t es un término de error de ruido blanco.

$$Y_t = \phi Y_{t-1} + u_t, -1 \leq \phi \leq 1$$

$$H_0: \beta = 0 \text{ (es equivalente a } \phi = 1)$$

$$H_1: \beta < 0 \text{ (es equivalente a } \phi < 1)$$

En la H_0 existe raíz unitaria, por lo tanto, no hay estacionariedad. En la H_1 no existe raíz unitaria, por lo tanto, hay estacionariedad.

3.2.3.2 Dickey and Pantula

El orden de diferenciación a menudo se determina en la práctica mediante la inspección visual de los gráficos de la ACF (Autocorrelation Function) y los datos. El procedimiento habitual es diferenciar si la ACF decae lentamente. Posteriormente, se aplican pruebas como la Prueba de Dickey-Fuller para determinar la presencia de una raíz unitaria, como se mencionó anteriormente. Estas pruebas ayudan a tomar decisiones menos subjetivas.

Normalmente, las personas primero prueban la presencia de una raíz unitaria en la serie. Si la hipótesis de una raíz unitaria no es rechazada, entonces se realizan diferencias para verificar la presencia de una segunda raíz unitaria y así sucesivamente. La manera correcta de probar las raíces unitarias es comparar una hipótesis nula de k raíces unitarias con una alternativa de $k-1$ raíces unitarias. (Dickey & Pantula, 1987)

Las serie de tiempo de precios, ventas, demanda lo usual es que tenga como máximo 2 diferencias, es decir un $I(2)$. Es por ellos que el proceso para encontrar el orden de diferenciación se utiliza el siguiente proceso:

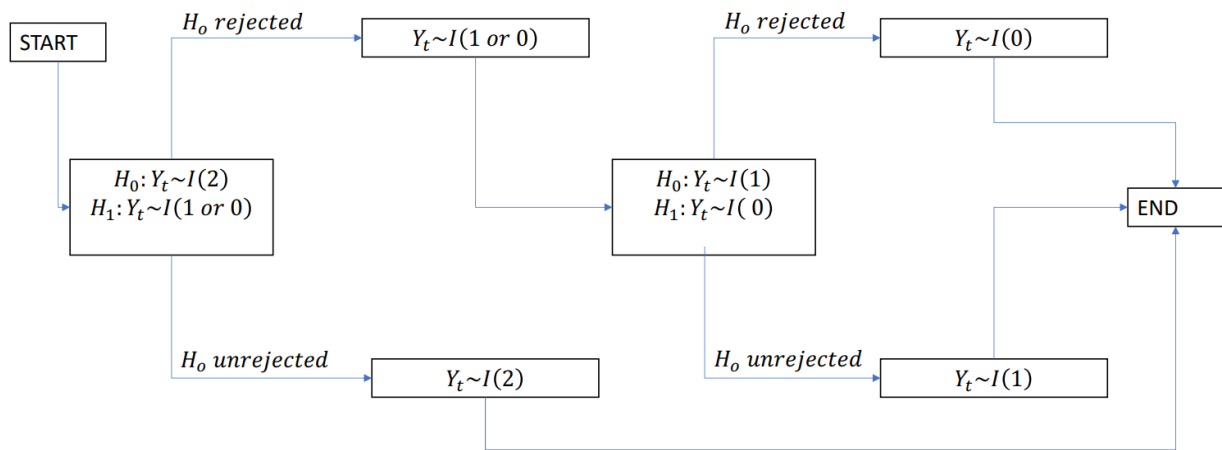


Imagen 6. Proceso de prueba de raíz unitaria

En resumen lo ideal es tomar la primera diferencia y testear la presencia de 2 raíces unitarias, en caso de no rechazar la H_0 el proceso termina y significa que se debe diferenciar 2 veces la serie para volverla estacionaria.

3.2.3.3 OCSB

La hipótesis nula del test OCSB (Out of Sample Box) es que una serie temporal contiene una raíz unitaria estacional. Esto se prueba mediante un método de regresión similar al test de Dickey-Fuller. La regresión de prueba a menudo debe incluir términos de autocorrelación para asegurar que los errores en el modelo se comporten como ruido blanco.

Cuando se incluyen rezagos estacionales y el método se establece en 'OLS' (Mínimos Cuadrados Ordinarios), la regresión de prueba se calcula utilizando Mínimos Cuadrados Ordinarios y solo se incluyen los rezagos estacionales. Si, en lugar de 'OLS', el método se establece en 'ML' (Máxima Verosimilitud), se calcula un modelo AR estacional, lo que implica que los rezagos no estacionales de alto orden se incluirán de manera indirecta también. Para aumentaciones estacionales, el método 'ML' es considerablemente más lento que 'OLS'. (R Project, s.f.)

Para probar esta hipótesis nula, se realiza una regresión de tipo Dickey-Fuller en la serie temporal. La Prueba de Dickey-Fuller es una prueba estadística utilizada para evaluar si una serie temporal es estacionaria o no estacionaria. Si la serie es estacionaria, significa que no tiene una tendencia sistemática que cambie con el tiempo.

3.2.4 ARIMA

Un modelo ARIMA significa modelo Autorregresivo Integrado de Medias Móviles. Definimos un modelo como autorregresivo si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación).

3.2.4.1 Funciones de Autocorrelación (ACF y PACF)

La autocorrelación y la autocorrelación parcial son medidas de asociación entre valores de series actuales y pasadas e indican cuáles son los valores de series pasadas más útiles para predecir valores futuros. Con estos datos se puede determinar el orden de los procesos en un modelo ARIMA (IBM, 2021).

- ✚ Función de autocorrelación (ACF o FAS). En el retardo k , es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia.
- ✚ Función de autocorrelación parcial (PACF o FAP). En el retardo k , es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia, teniendo en cuenta los valores de los intervalos intermedios.

3.2.4.2 Modelos Autorregresivos

Los modelos autorregresivos se abrevian con la palabra AR tras la que se indica el orden del modelo: AR(1), AR(2),... etc. El orden del modelo expresa el número de observaciones retasadas de las series temporal analizada que intervienen en la ecuación. Así, por ejemplo, un modelo AR(1) tendría la siguiente expresión (de Arce & Mahía, 2001):

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + a_t$$

El término de error de los modelos de este tipo se denomina generalmente ruido blanco cuando cumple las tres hipótesis básicas:

- ✚ Media Nula
- ✚ Varianza constante
- ✚ Covarianza nula entre errores correspondientes a observaciones diferentes

En términos generales un modelo AR(P) se puede escribir de la siguiente forma:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t$$

Pudiéndose escribir de la siguiente forma simplificada:

$$\phi_p(L)Y_t = \phi_0 + a_t; \text{ donde } \phi_p(L) \text{ es el operador polinomial de retardos}$$

Cuando trabajamos con modelos AR(p) es importante conocer cuántos autorregresores vamos a utilizar. Para esto nos vamos a centrar en la Función de Autocorrelación Parcial de la serie. En el ejemplo de la imagen inferior se tiene un AR(1), es importante destacar que la función de autocorrelación de la serie debe reducir su valor a lo largo de los retardos para que el modelo previsional resulte estable y estacionario aparte para que la memoria de la variable no genere resultados divergentes en un horizonte temporal relativamente corto.

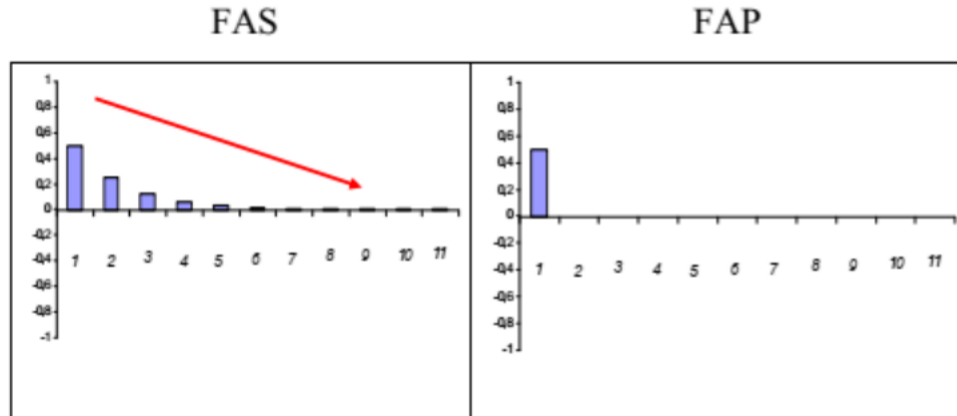


Imagen 7. Función de Autocorrelación Parcial (PACF o FAP)

3.2.4.3 Modelo de Medias Móviles

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas MA, seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA(q) respondería a la siguiente expresión (de Arce & Mahía, 2001):

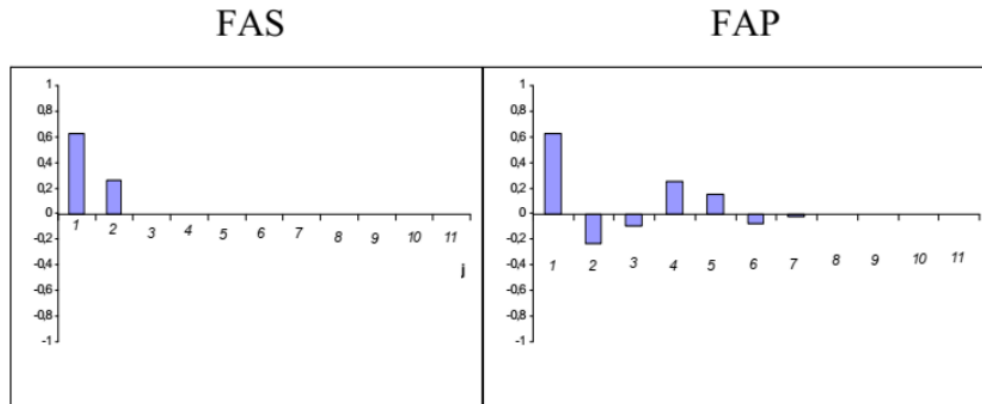
$$Y_t = \mu + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + \theta_q a_{t-q}$$

También se puede simplificar de la siguiente forma:

$$Y_t = \theta_q (L) a_t + \mu$$

Cuando trabajamos con modelos MA(q) es importante conocer cuantas medias móviles vamos a utilizar. Para esto nos vamos a centrar en la Función de Autocorrelación de la serie. En el ejemplo de la imagen inferior vemos que tenemos un MA(2), es importante destacar que la función de autocorrelación parcial de la serie debe reducir su valor a lo largo de los retardos para que el modelo previsional resulte estable e invertible y la memoria de la variable no genere resultados divergentes en un horizonte temporal relativamente corto.

Imagen 8. Función de autocorrelación FAS o ACF MA(2)



3.2.4.4 Modelos ARIMA

Un modelo arima se define como $ARIMA(p,d,q)$, en donde p es el componente $AR(p)$, q es el componente $MA(q)$ y d es el componente de diferenciación requerido para que nuestra serie sea estacionaria. Matemáticamente lo podemos representar con la siguiente ecuación:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d Y_t = \delta + (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) u_t$$

Donde U_t se supone un ruido blanco con media cero y varianza constante. Definida esta función genérica, el objetivo principal es estimar el vector formado por los parámetros correspondientes a la parte autorregresiva y de medias móviles (incluido, si fuera necesario, el término independiente) así como la varianza residual.

Un modelo estacional puede ajustarse a datos que presentan algún tipo de estacionalidad. Eso suele ser el caso si la variable cuyo proceso estocástico queremos estudiar y predecir está relacionada con períodos cortos y secuenciales, como trimestres, meses, semanas o días. Un patrón estacional, indicado como 's', afecta cíclicamente (ciclos cortos, es decir) a la variable de interés (Rubino, 2022). Una serie estacional la podemos representar conjuntamente de la siguiente manera $ARIMA(p,d,q) \times (P,D,Q)s$, donde (P,D,Q) es la componente estacional de $AR(P)$, $MA(Q)$ y la diferenciación D .

El modelo en su forma más genérica lo podemos representar de la siguiente manera:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - \Phi_1 L^s - \dots - \Phi_p L^{ps})(1 - L)^d(1 - L^s)^D Y_t = \mu + (1 - \theta_1 L - \dots - \theta_q L^q)(1 - \Theta_1 L^s - \dots - \Theta_Q L^{Qs}) u_t$$

Donde U_t se supone un ruido blanco con media cero y varianza constante. Las condiciones de estacionariedad e invertibilidad del polinomio característico también se aplican en el SARIMA.

3.2.5 Holt Winters

El método Holt-Winters es una ampliación perfeccionada del enfoque de la suavización exponencial, mientras que el procedimiento de suavización proporciona una impresión general, movimientos a largo plazo en la información y permite la elaboración de pronósticos a corto plazo. Este método permite también el estudio de tendencia a futuro mediante la elaboración de pronósticos a mediano y largo plazo. El método de alisado exponencial lineal y estacional de tres parámetros de Winters es una extensión del modelo de Holt, que pudiera reducir el error de pronóstico. El método aditivo de Holt-Winters, se usa para series temporales con variación estacional constante, y el método multiplicativo de Holt-Winters, se aplica en series temporales con variación estacional creciente. El método de Holt-Winters, se basa en tres ecuaciones suavizadas: una para el nivel, una para la tendencia y una para la estacionalidad, se usa tres parámetros de suavización para cada ecuación: α , β y γ . Para la construcción del modelo tanto aditivo como multiplicativo existen tres etapas importantes (Real, 2021):

1. **Inicialización:** La razón de conseguir valores iniciales para los métodos de suavización exponencial es que los modelos son ecuaciones recursivas, por lo tanto, se necesita empezar por algún lado. Para inicializar el método de pronóstico de Holt-Winters, se debe inicializar los índices de niveles, tendencia, y las estacionalidades.
2. **Optimización:** Se habrá logrado obtener los parámetros óptimos cuando estos reduzcan al MSE al mínimo. Es posible utilizar alguna otra medida de error de predicción como el MAE o MAPE, pero el MSE tiende a ser más fácil para trabajar este método.

3. **Ajuste y Pronóstico:** La fase de ajuste es un proceso de cálculos antes de llegar a la etapa de pronóstico, en ambas etapas se utilizan las ecuaciones presentadas anteriormente correspondientes a cada modelo, y finalmente se halla los pronósticos que por lo general es el objetivo principal de estudiar series de tiempo.

Como mencionamos anteriormente tenemos 2 tipos de modelos de holt winters, el aditivo y multiplicativo. El modelo matemático del holt winters aditivo es el siguiente (Minitab, 2023):

$$L_t = \alpha (Y_t - S_{t-p}) + (1 - \alpha) [L_{t-1} + T_{t-1}]$$

$$T_t = \gamma [L_t - L_{t-1}] + (1 - \gamma) T_{t-1}$$

$$S_t = \delta (Y_t - L_t) + (1 - \delta) S_{t-p}$$

$$y_t = L_{t-1} + T_{t-1} + S_{t-p}$$

El modelo multiplicativo de Holt Winters se representa mediante la siguiente ecuación:

$$S_t = \delta (Y_t / L_t) + (1 - \delta) S_{t-p}$$

$$y_t = (L_{t-1} + T_{t-1}) S_{t-p}$$

En donde:

Término	Descripción
L_t	El nivel en el tiempo t , α es la ponderación para el nivel
T_t	Tendencia en el tiempo t ,
γ	La ponderación para la tendencia
S_t	Componente estacional en el tiempo t
δ	La ponderación para el componente estacional
p	Período estacional
Y_t	El valor de los datos en el tiempo t
y_t	El valor ajustado o el pronóstico de un periodo por anticipado, en el tiempo t

Tabla 1 Términos del modelo holt Winters

3.2.6 Prophet

En febrero 2017, Facebook liberó una herramienta (de código abierto) disponible para Python y R, llamada Prophet. En grandes líneas, es una librería que permite construir modelos de ajuste y pronósticos de series. Permite manejar sets de datos en donde haya observaciones nulas, faltantes, grandes valores atípicos, cambios de tendencia importantes o mismo, tendencias no lineales. Es decir, resulta extremadamente versátil para utilizarlo sobre todo tipo de serie de tiempo sobre la cual se quieran realizar pronósticos (Sáez, 2018).

Facebook Prophet es un algoritmo de código abierto para generar modelos de series temporales que utiliza algunas ideas antiguas con algunas nuevas innovaciones. Es particularmente eficiente en la modelación de series temporales que tienen múltiples estacionalidades y no enfrenta algunas de las limitaciones de otros algoritmos. En su núcleo se encuentra la suma de tres funciones de tiempo más un término de error: crecimiento $g(t)$, estacionalidad $s(t)$, días festivos $h(t)$ y error e (Krieguer, 2021).

El modelo Prophet se puede representar matemáticamente bajo la siguiente fórmula tomando en consideración los 3 criterios mencionados anteriormente (la estacionalidad, el crecimiento, días festivos y error):

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Aquí, $g(t)$ es la función de tendencia que modela cambios no periódicos en el valor de la serie temporal, $s(t)$ representa cambios periódicos (por ejemplo, estacionalidad semanal y anual), y $h(t)$ representa los efectos de los días festivos que ocurren en horarios potencialmente irregulares durante uno o más días. El término de error $\epsilon(t)$ representa cualquier cambio idiosincrático que no es acomodado por el modelo (Taylor & Letham, 2017). Algunas de las ventajas de este modelo sobre otros son las siguientes:

- ✚ Flexibilidad: Podemos incorporar fácilmente estacionalidades con múltiples períodos y permitir que el analista haga diferentes suposiciones sobre las tendencias.
- ✚ A diferencia de los modelos ARIMA, las mediciones no necesitan estar espaciadas regularmente y no es necesario interpolar valores faltantes, por ejemplo, al eliminar valores atípicos.
- ✚ El ajuste es muy rápido, lo que permite al analista explorar de manera interactiva muchas especificaciones del modelo, por ejemplo, en una aplicación Shiny (paquete de R que permite crear páginas web interactivas).
- ✚ El modelo de pronóstico tiene parámetros fácilmente interpretables que pueden ser modificados por el analista para imponer suposiciones en el pronóstico. Tendencia Lineal con Puntos de Cambio

Para problemas de pronóstico que no presentan un crecimiento de saturación, una tasa de crecimiento constante por segmentos proporciona un modelo parsimonioso y a menudo útil. Aquí, el modelo de tendencia es:

$$g(t) = (k + a(t)^\tau \delta)t + (m + a(t)^\tau \gamma)$$

donde, k es la tasa de crecimiento, δ tiene los ajustes de tasa, m es el parámetro de desplazamiento, y γ se establece para hacer que la función sea continua.

3.2.6.1 Estacionalidad

Las series temporales empresariales a menudo presentan una estacionalidad de múltiples períodos como resultado de los comportamientos humanos que representan. Por ejemplo, una semana laboral de 5 días puede producir efectos en una serie temporal que se repiten cada semana, mientras que los calendarios de vacaciones y los recesos escolares pueden producir efectos que se repiten cada año. Para ajustar y pronosticar estos efectos, debemos especificar modelos de estacionalidad que sean funciones periódicas de t .

Confiamos en las series de Fourier para proporcionar un modelo flexible de efectos periódicos. Sea P el período regular que esperamos que tenga la serie temporal (por ejemplo, $P = 365.25$ para datos anuales o $P = 7$ para datos semanales, cuando escalamos nuestra variable de tiempo en días). Podemos aproximar efectos estacionales suaves arbitrarios con:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

Una serie de Fourier estándar. Ajustar la estacionalidad requiere estimar los $2N$ parámetros $\beta = [a_1, b_1, \dots, a_N, b_N]$. Esto se realiza construyendo una matriz de vectores de estacionalidad para cada valor de t en nuestros datos históricos y futuros (Taylor & Letham, 2017).

3.2.7 LSTM (Long Short – Term Memory)

Antes de definir un modelo LSTM es importante conocer las redes neuronales recurrentes (RNN). Una RNN es un tipo de red neuronal artificial que utiliza datos secuenciales o datos de series temporales. Las redes neuronales recurrentes utilizan datos de entrenamiento para aprender. Se distinguen por su "memoria", ya que obtienen información de entradas anteriores para influir en la entrada y salida actuales. (IBM, s.f.)

Long Short Term Memory(LSTM) son una extensión de las redes neuronales recurrentes, que básicamente amplían su memoria para aprender de experiencias importantes que han pasado hace mucho tiempo. Las LSTM permiten a las RNN recordar sus entradas durante un largo período de tiempo. Esto se debe a que LSTM contiene su información en la memoria, que puede considerarse similar a la memoria de un ordenador, en el sentido que una neurona de una LSTM puede leer, escribir y borrar información de su memoria.

En una neurona LSTM hay tres puertas a estas “celdas” de información, estas puertas determinan si se permite o no una nueva entrada, se elimina la información porque no es importante o se deja que afecte a la salida en el paso de tiempo actual. (Torres, 2019):

- ✚ Puerta de olvido (forget gate): La puerta de olvido decide qué información de la celda de memoria anterior se debe olvidar o mantener en el nuevo estado de la celda de memoria. Su objetivo es eliminar información irrelevante u obsoleta. Se representa bajo la siguiente formula:

$$f_t = \sigma_g(W_f(x_t) + U_i(h_{t-1} + b_f))$$

- ✚ Puerta de entrada (input gate): Su función principal es decidir qué nueva información debe ingresar a la celda de memoria en un paso de tiempo dado. Se representa bajo la siguiente formula:

$$i_t = \sigma_g(W_i(x_t) + U_i(h_{t-1} + b_i))$$

- ✚ Puerta de salida (output gate): La puerta de salida decide qué información de la celda de memoria se debe mostrar como salida en el estado oculto en el paso de tiempo actual. Se representa bajo la siguiente formula:

$$o_t = \sigma_g(W_o(x_t) + U_o(h_{t-1} + b_o))$$

El LSTM tiene una entrada $x(t)$, que es la secuencia de entrada directamente. $h(t-1)$ es la entrada del paso de tiempo anterior del LSTM. $o(t)$ es la salida del LSTM para este paso de tiempo. Las matrices de pesos (W_f , W_i , W_o , U_f , U_i , U_o) y los sesgos (b_f , b_i , b_o) no dependen del tiempo, se mantienen constantes.

A continuación, se muestra el funcionamiento del LSTM. (Rastogi, 2020)

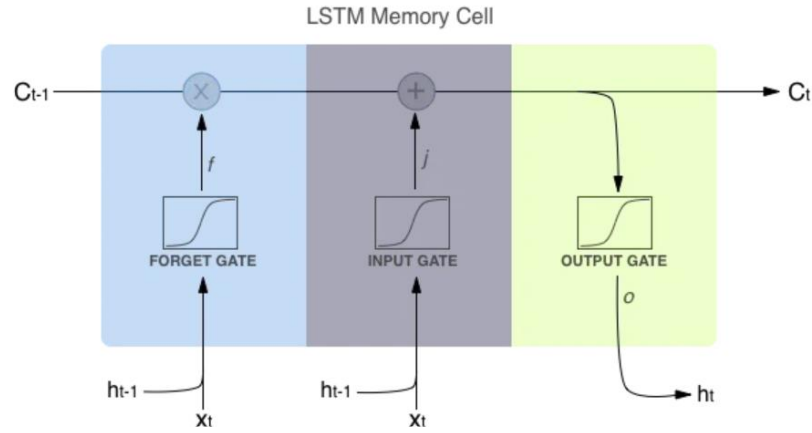


Imagen 9. LSTM modelo

Finalmente, un modelo LSTM para un único paso de tiempo viene dado por las siguientes ecuaciones:

$$C'_t = \sigma_c(W_c(x_t) + U_c(h_{t-1} + b_c))$$

$$C_t = f_t * C_{t-1} + i_t * C'_t$$

$$h_t = o_t * \sigma_c(C_t)$$

En donde C_t es el estado de la celda, h_t es el estado oculto, σ_c es la tangente hiperbólica para introducir no linealidad en las operaciones, similar a la función sigmoide, pero con un rango que va desde -1 hasta 1; finalmente σ_g es la sigmoide que de igual se utiliza para introducir no linealidad y tiene un rango de 0 a 1. Las variables $f(t)$, $i(t)$, y $c'(t)$ son utilizadas dentro de la red LSTM para realizar cálculos internos y que estos cálculos se utilizan para generar las variables $c(t)$ y $h(t)$.

3.2.8 Métricas de Evaluación de Modelos

3.2.8.1 Error Absoluto Medio (MAE)

En MAE, el error se calcula como un promedio de diferencias absolutas entre los valores objetivo y las predicciones. El MAE es una puntuación lineal, lo que significa que todas las diferencias individuales se ponderan por igual en el promedio. Matemáticamente, se calcula utilizando la fórmula que se presenta a continuación (Sitio Big Data, 2018):

$$MAE = \frac{1}{N} \sum_{m=1}^N |Y_{i+m} - \bar{Y}_m|$$

Donde Y_i es el resultado real esperando y \bar{Y}_i es la predicción del modelo.

3.2.8.2 Error Cuadrático Medio (MSE)

Es quizás la métrica más simple y común para la evaluación de regresión, pero también es probablemente la menos útil. Se define por la ecuación:

$$MSE = \frac{1}{N} \sum_{m=1}^N (Y_{i+m} - \bar{Y}_m)^2$$

El MSE básicamente mide el error cuadrado promedio de nuestras predicciones. Para cada punto, calcula la diferencia cuadrada entre las predicciones y el objetivo y luego promedia esos valores.

3.2.8.3 Error cuadrático medio (RMSE)

El RMSE es solo la raíz cuadrada de MSE. La raíz cuadrada se introduce para hacer que la escala de los errores sea igual a la escala de los objetivos. Ahora, es muy importante entender en qué sentido RMSE es similar a MSE y cuál es la diferencia.

Primero, son similares en términos de sus minimizadores, cada minimizador de MSE es también un minimizador para RMSE y viceversa, ya que la raíz cuadrada es una función que no disminuye. Por ejemplo, si tenemos dos conjuntos de predicciones, A y B, y decimos que el MSE de A es mayor que el MSE de B, entonces podemos estar seguros de que RMSE de A es mayor que RMSE de B.

3.2.8.4 Error de Porcentaje Medio Absoluto (MAPE)

El Error Porcentual Absoluto Medio, se calcula utilizando el error absoluto en cada período dividido por los valores observados que son evidentes para ese período. Luego, se promedian esos porcentajes fijos. El MAPE indica cuánto error existe al predecir en comparación con el valor real (Khair, Fahmi, Al Hakim, & Rahim4, 2017). Un MAPE menor a 1% se considera bien, entre 1% y 5% es aceptable y mayor a 5% se considera malo. Se representa de la siguiente forma:

$$MAPE = \frac{100}{n} \sum_{m=1}^n \frac{|Y_{i+m} - \bar{Y}_m|}{Y_{i+m}}$$

3.2.8.5 Error Escalados Absoluto Medio (MASE)

En la predicción de series temporales, el Error Absoluto Medio Escalado (MASE, por sus siglas en inglés) es una medida para determinar la eficacia de las predicciones generadas mediante un algoritmo, al comparar las predicciones con el resultado de un enfoque de predicción ingenua. La predicción ingenua se genera en cualquier paso al igualar la predicción actual con la salida del paso anterior en el tiempo. (Ahuja, 2021)

El MASE se puede representar bajo la siguiente fórmula:

$$MASE = \frac{MAE}{MAE_{naive}}$$

Donde el MAE Naive se calcula bajo la siguiente ecuación:

$$MAE_{naive} = \frac{1}{1-n} |y_i - y_{i-1}|$$

Se considera aceptable un MASE menor a 1 ya que de lo contrario significa que la previsión ingenua es mejor que el algoritmo.

4 Minería de Datos

Como se mencionó anteriormente, el objetivo principal es realizar un pronóstico de ventas que se ajuste a las necesidades de la empresa. Para el análisis y modelado de los datos se utilizó la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), considerada como una de las más utilizadas y recomendadas para proyectos de minería de datos. Esta metodología consiste en un modelo de proceso con seis fases que describe el ciclo de vida de la ciencia de datos.

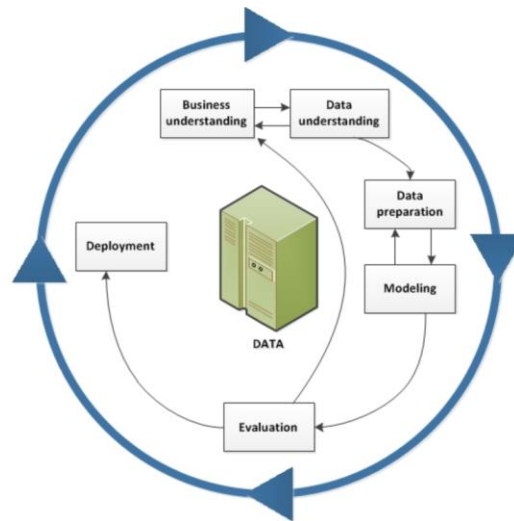


Imagen 10. Ciclo de Vida de Minería de Datos (IBM, 2021)

De forma general, la metodología CRISP-DM divide el proceso de minería de datos en seis fases principales: 1) comprensión del negocio, 2) comprensión de los datos, 3) preparación de los datos, 4) fase de modelado, 5) evaluación, y 6) implantación. La primera fase pretende conocer los objetivos del negocio, valorar la situación actual y realizar un plan del proyecto. La segunda fase busca conocer los datos iniciales de los que se dispone. En la tercera fase se preparan los datos recogidos en la fase anterior, para que en la cuarta fase se aplique una técnica de modelado que se adapte a los objetivos del negocio y genere información nueva y relevante. En la quinta fase se analizan los resultados obtenidos valorando su calidad, para que, por último, se indiquen los planes de uso del proyecto.

5 Desarrollo y Resultados

5.1 Recolección de Datos

Droguería Proconsumo es una empresa grande que cada vez la apuesta más a la rama tecnológica, es por ello que almacenen su información (ventas, inventarios, etc.) en almacenes de datos, específicamente PostgreSQL. El hecho que tengan almacenada su información en este programa facilita su extracción ya que se puede conectar fácilmente con otras herramientas de análisis de datos como R y Python. En el anexo 1 se encuentra el código de Python en Jupyter Notebook y en HTML para su facilidad de lectura.

La extracción de datos se llevó a cabo mediante el lenguaje de programación Python utilizando una librería llamada “sqlalchemy” importando el objeto “create engine”. En este objeto se debe indicar los siguientes parámetros como una cadena de texto para poder conectarse a la base de datos:

- 🚩 Usuario
- 🚩 Contraseña
- 🚩 Nombre del Servidor
- 🚩 Nombre de la base de datos

El comando se puede apreciar de la siguiente manera (se muestran “xxx” por seguridad de la empresa):

```
engine = create_engine("xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx")
#por seguridad de la empresa he quitado los datos de conexión al servidor
```

Imagen 11. Conexión al Servidor de Droguería Proconsumo. Fuente propia

Una vez conectados al servidor creamos los “queries” para poder extraer las ventas. Tal como se estableció en el objetivo general, las ventas se extraen de manera mensual desde el 2017-01-01 hasta el 2023-07-31; en donde la idea es predecir la venta del 2023. La extracción se realiza de la siguiente manera:

1. Los datos de entrenamiento van desde el 2017-01-01 hasta el 2023-12-31 bajo el siguiente query llamado “ventas_proco”:

```
# Query de Ventas este será nuestro train_set
query= """
SELECT
    DATE_TRUNC('month',fecha_transaccion) AS month_dates,
    CAST(SUM(COALESCE(venta_bruta, 0)) - (SUM(descuento) - SUM(COALESCE(descuento_proveedor, 0))) AS NUMERIC(18,2)) AS net_sales
FROM mstr_dim_zsales_sap
WHERE fecha_transaccion >= '2017-01-01' AND fecha_transaccion <= '2022-12-31'
AND customer_id NOT IN ('CA00000002','CA00000005')
GROUP BY month_dates
"""
ventas_proco = pd.read_sql(query, engine)
ventas_proco = ventas_proco.set_index("month_dates")
```

Imagen 12. Query de datos de entrenamiento

2. Seguidamente se procede a extraer los datos de prueba llamado “ventas_test”:

```
query= """
SELECT
    DATE_TRUNC('month',fecha_transaccion) AS month_dates,
    CAST(SUM(COALESCE(venta_bruta, 0)) - (SUM(descuento) - SUM(COALESCE(descuento_proveedor, 0))) AS NUMERIC(18,2)) AS net_sales
FROM mstr_dim_zsales_sap
WHERE fecha_transaccion >= '2023-01-01' AND fecha_transaccion <= '2023-07-31'
AND customer_id NOT IN ('CA00000002','CA00000005')
GROUP BY month_dates
"""
ventas_test = pd.read_sql(query, engine) #este sera nuestro test_set
ventas_test = ventas_test.set_index("month_dates")
```

Imagen 13. Query datos de prueba

- Una vez que tenemos ambos conjuntos de datos los unimos y llamamos “ventas_total” para realizar un análisis exploratorio para ver media de los datos, desviación estándar y una gráfica para tener una idea de cómo se comportan los datos.

5.2 Exploración de Datos

En el apartado de recolección de datos ya logramos extraer nuestra información deseada, ahora se debe realizar un análisis exploratorio de los datos y comprobar formatos de los datos.

- ✚ Verificar que tenemos datos en formato correctos: tenemos nuestras fechas en formato de fecha, no tenemos valores vacíos y el total de datos corresponde a 79 meses:

Imagen 14. Verificacion de datos

```
ventas_total.info() #verificar que tenemos Las fechas en formato correcto

<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 79 entries, 2017-01-01 to 2023-07-01
Data columns (total 1 columns):
#   Column      Non-Null Count  Dtype
---  -
0   net_sales  79 non-null    float64
dtypes: float64(1)
memory usage: 1.2 KB
```

- Análisis visual de como se ve el total de ventas de la empresa desde el 2017 hasta la fecha. A simple vista podemos ver que si tenemos una tendencia positiva y puede ser que tengamos estacionalidad.

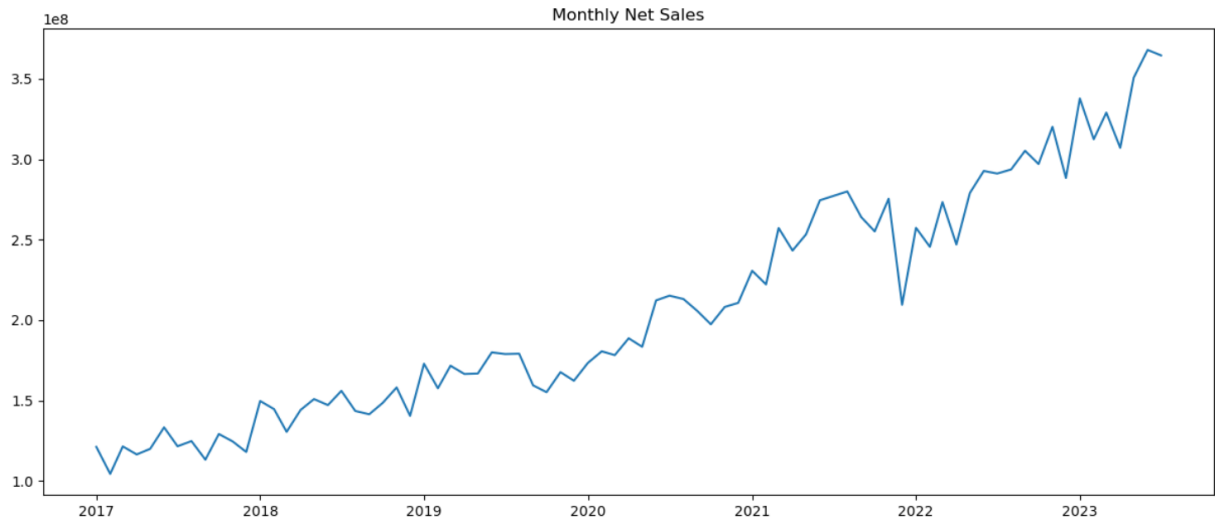


Imagen 15. Gráfica de Ventas de la empresa

- Vemos valores de mínimos, máximos, desviación estándar y cuartiles

```
ventas_proco.describe()
```

net_sales	
count	7.200000e+01
mean	1.947304e+08
std	5.967801e+07
min	1.043884e+08
25%	1.465247e+08
50%	1.790024e+08
75%	2.485617e+08
max	3.201240e+08

Imagen 16. Análisis descriptivo de datos

5.3 Modelado de Datos

5.3.1 Modelo ARIMA y SARIMA

5.3.1.1 Modelo ARIMA

La implementación de un modelo ARIMA requiere que la serie sea estacionaria, por ende, es lo primero que se debe validar y en caso que no sea, la serie se debe diferenciar. Siguiendo el planteamiento de Dickey and Pantula se hace la prueba de ADF en donde la hipótesis nula es $I(2)$ vs la alternativa de $I(1)$ o $I(0)$. En la tabla inferior se puede ver los resultados de la prueba de Dickey Fuller aumentada.

ADF Test	Resultado
ADF statistic	-2.0487
p_value	0.2657

Tabla 2 Resultados prueba de raiz unitaria

El resultado de la serie original es que no es estacionaria, y requiere una diferenciación de orden 2, es decir $I(2)$, ya que NO se rechaza la H_0 y el proceso termina. Esto confirma el análisis visual que se puede obtener viendo la serie original con la ACF que claramente muestra que no es estacionara, tal como se aprecia en la imagen inferior.

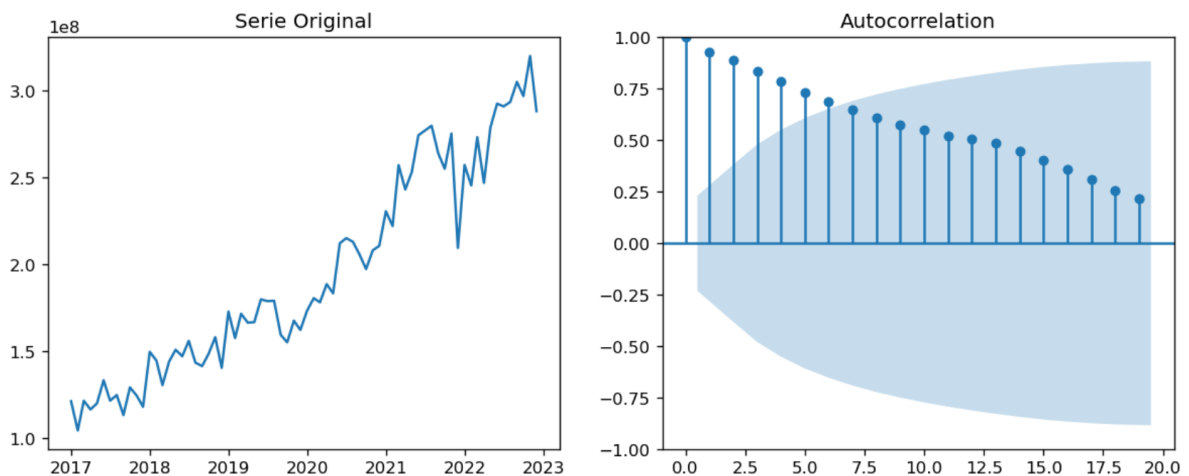


Imagen 17. Serie original con su grafica de autocorrelación

A continuación, se muestra la serie diferenciada 2 veces

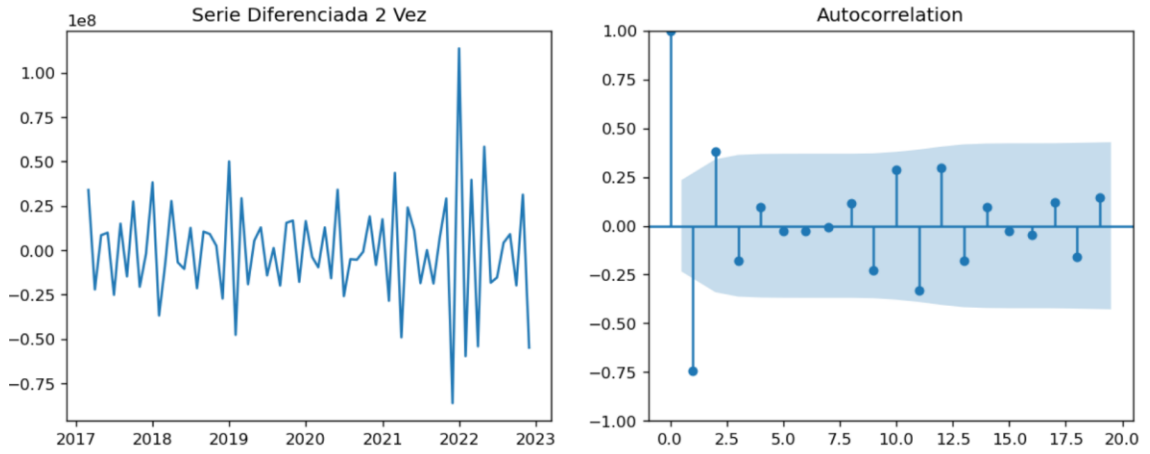


Imagen 18. Serie diferenciada 2 veces

El siguiente paso consiste en encontrar los parámetros $AR(p)$ y $MA(q)$ de la serie, para ello se analiza el gráfico de Autocorrelación y Autocorrelación Parcial. En la imagen 17 se tienen 2 o 3 puntos que cruzan la frontera del nivel de confianza por tanto se debe probar con 2 y 3 medias móviles ($MA(1)$, $MA(2)$). La autocorrelación parcial indica 2 puntos que cruzan la frontera según la imagen inferior.

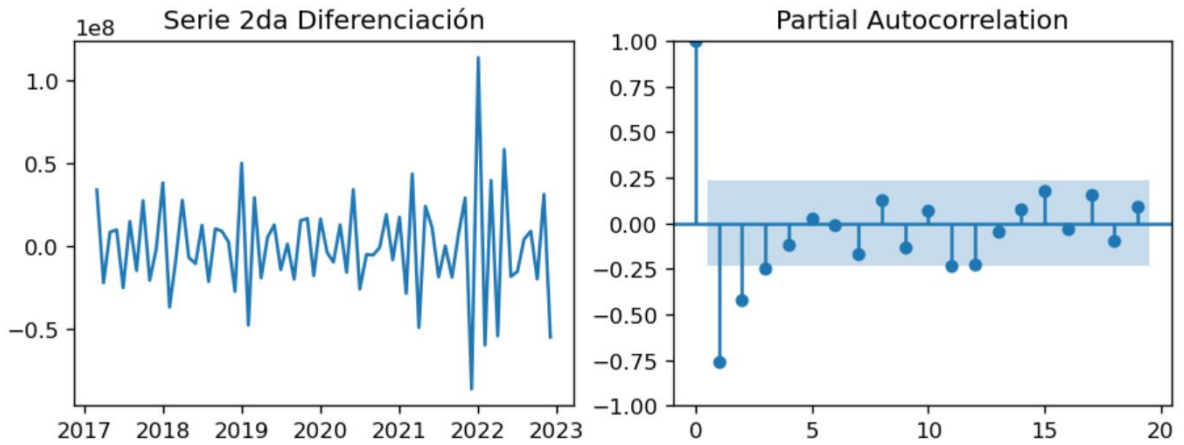


Imagen 19. Autocorrelación Parcial

El siguiente paso una vez definidos los parámetros ARIMA(p,d,q) es realizar pruebas y comparar ambos AIC, al final se utiliza el que tenga un valor más bajo de AIC. En la siguiente tabla se muestran los resultados de los modelos ARIMA.

Modelo	AIC
ARIMA(2,2,2)	2,546.66
ARIMA(2,2,1)	2,547.01
ARIMA(1,2,1)	2,545.74

Tabla 3 Tabla Resumen modelos ARIMA

El modelo finalmente seleccionado es ARIMA(1,2,1), a con continuación, se muestra el resumen del modelo.

```

=====
SARIMAX Results
=====
Dep. Variable:          net_sales  No. Observations:          72
Model:                 ARIMA(1, 2, 1)  Log Likelihood             -1269.871
Date:                  Fri, 01 Sep 2023  AIC                        2545.742
Time:                  13:30:39        BIC                        2552.487
Sample:                01-01-2017      HQIC                       2548.421
                    - 12-01-2022
Covariance Type:      opg
=====
              coef    std err          z      P>|z|      [0.025    0.975]
-----
ar.L1          -0.2437     0.069     -3.549     0.000     -0.378     -0.109
ma.L1          -1.0000     0.114     -8.805     0.000     -1.223     -0.777
sigma2         2.708e+14    4.2e-16    6.46e+29    0.000    2.71e+14    2.71e+14
=====
Ljung-Box (L1) (Q):                1.49  Jarque-Bera (JB):                5.51
Prob(Q):                             0.22  Prob(JB):                       0.06
Heteroskedasticity (H):              1.19  Skew:                            -0.45
Prob(H) (two-sided):                 0.67  Kurtosis:                        4.04
=====

```

Tabla 4 Tabla de Resultados ARIMA(1,2,1)

5.3.1.2 Modelo SARIMA

El modelo SARIMA toma en consideración la estacionalidad que se tiene en la serie de tiempo. A simple vista parece que sí según la imagen 15, pero se debe comprobar descomponiendo la serie. En la imagen inferior se detalla la descomposición que muestra que si hay estacionalidad:

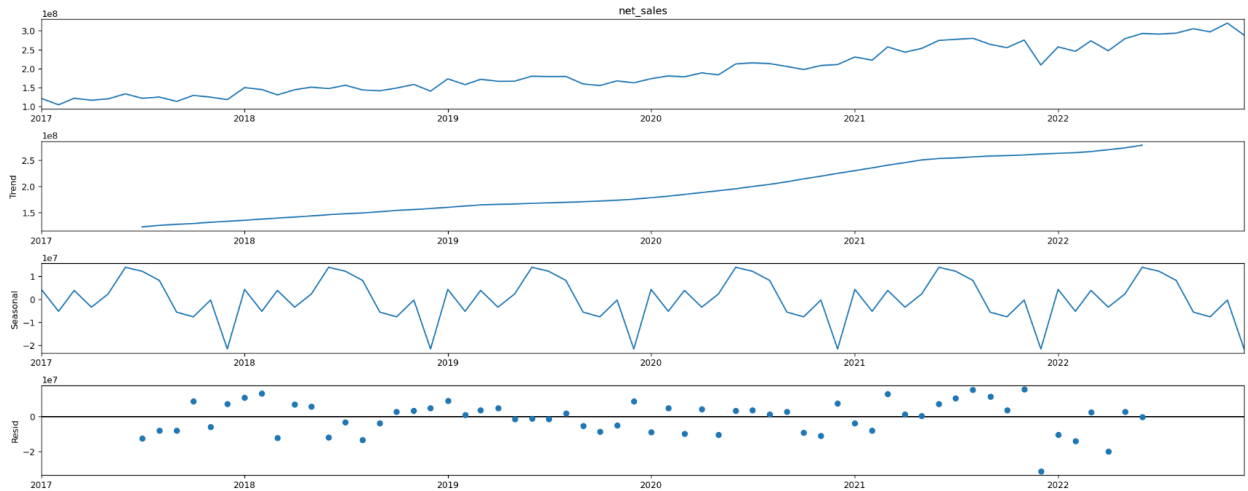


Imagen 20. Descomposición de la serie de tiempo

El método de descomposición utilizado para la serie de tiempo es aditivo y se descompone utilizando medias móviles.

Para encontrar los parámetros P, D y Q vamos a utilizar la función de “auto_arima” en donde selecciona el modelo que mejor se ajusta a los datos reduciendo el AIC. Es importante destacar que para encontrar el orden de diferenciación D por defecto utiliza el test “OCSB”. El resumen del modelo según la tabla inferior es $ARIMA(1,2,1) \times (1,0,1)$.

Dep. Variable:	y	No. Observations:	72			
Model:	SARIMAX(1, 2, 1)x(1, 0, 1, 12)	Log Likelihood	-1260.628			
Date:	Mon, 11 Sep 2023	AIC	2531.257			
Time:	12:41:04	BIC	2542.499			
Sample:	01-01-2017	HQIC	2535.722			
	- 12-01-2022					
Covariance Type:	opg					
	coef	std err	z	P> z 	[0.025	0.975]
ar.L1	-0.2767	0.086	-3.211	0.001	-0.446	-0.108
ma.L1	-0.8192	0.092	-8.892	0.000	-1.000	-0.639
ar.S.L12	0.7740	0.178	4.341	0.000	0.425	1.123
ma.S.L12	-0.4749	0.263	-1.806	0.071	-0.990	0.040
sigma2	2.708e+14	1.55e-15	1.74e+29	0.000	2.71e+14	2.71e+14
Ljung-Box (L1) (Q):	0.46	Jarque-Bera (JB):	22.31			
Prob(Q):	0.50	Prob(JB):	0.00			
Heteroskedasticity (H):	1.42	Skew:	-0.83			
Prob(H) (two-sided):	0.41	Kurtosis:	5.21			

Tabla 5 Resumen modelo SARIMA

5.3.2 Modelo Holt Winters

Al trabajar con el modelo de Holt Winters es importante definir parámetros de tendencia y estacionalidad como multiplicativos o aditivos. Para definir estos parámetros se analiza la imagen 19 de la descomposición de la serie. Como se observa una tendencia línea creciente entonces se puede concluir que la tendencia es aditiva, en el caso de la estacionalidad es multiplicativa ya que son cambios estacionales crecientes. Los periodos para considerar estacionalidad son mensuales por tanto el valor del parámetro es de 12.

5.3.3 Modelo Prophet

El modelo Pophet de Facebook requiere cambios en los nombres de las variables, en donde estas se deben llamar “ds” a las fechas y “y” a los valores a analizar. En la imagen se puede observar el cambio en el nombre de las variables.

	ds	y
0	2017-01-01	1.212200e+08
1	2017-02-01	1.043884e+08
2	2017-03-01	1.214885e+08
3	2017-04-01	1.164841e+08
4	2017-05-01	1.199715e+08

Imagen 21. Preparación para modelo prophet

5.3.4 Modelo LSTM

La preparación de los datos para un LSTM requiere de una transformación de los datos de entrenamiento (ventas_proco) y datos de prueba (ventas_test). Esto consiste en una transformación que consiste en la estandarización, que resta la media y la divide por la desviación, esto se debe realizar ya que se tienen diferentes magnitudes entre los datos es decir un amplio rango. A continuación, en la tabla inferior se muestra en ejemplo de los datos de entrenamiento transformados:

Dato Transformado	Dato Real Entrenamiento
0.07801967	121,219,995.74
0	104,388,373.03
0.07926412	121,488,468.21
0.05606741	116,484,112.24
0.07223262	119,971,522.85

Tabla 6 Datos transformados mediante estandarización

Otro aspecto importante que se debe destacar es la activación que se hace en el modelo, que sea ReLu (Rectified Linear Unit). La activación ReLu significa que cualquier valor positivo se mantiene igual, mientras que los valores negativos se establecen en cero. Esta simplicidad no lineal es útil para el aprendizaje profundo, ya que permite que el modelo aprenda representaciones más complejas y detecte características importantes en los datos. Finalmente el algoritmo de adaptación que se utiliza es el Adam (Adaptive Moment Estimation) se utiliza por su capacidad para adaptar automáticamente la tasa de aprendizaje para cada parámetro del modelo. Esto es beneficios cuando los datos cambian con el tiempo.

5.4 Resultados

En esta sección se presentan los resultados de los modelos realizados, sus métricas de evaluación y la comparación de resultados.

5.4.1 Modelo ARIMA

En la sección anterior se concluyó que el modelo ARIMA a utilizar es el ARIMA(1,2,1), ya que es el que tiene el valor más bajo del AIC. Una vez realizada la predicción esta se puede apreciar bajo la siguiente gráfica que compara la predicción (color azul) y la venta real (color rojo). Al ser un modelo sencillo que no considera la estacionalidad a partir de mi segunda predicción se vuelve una línea recta que solo captura la tendencia.

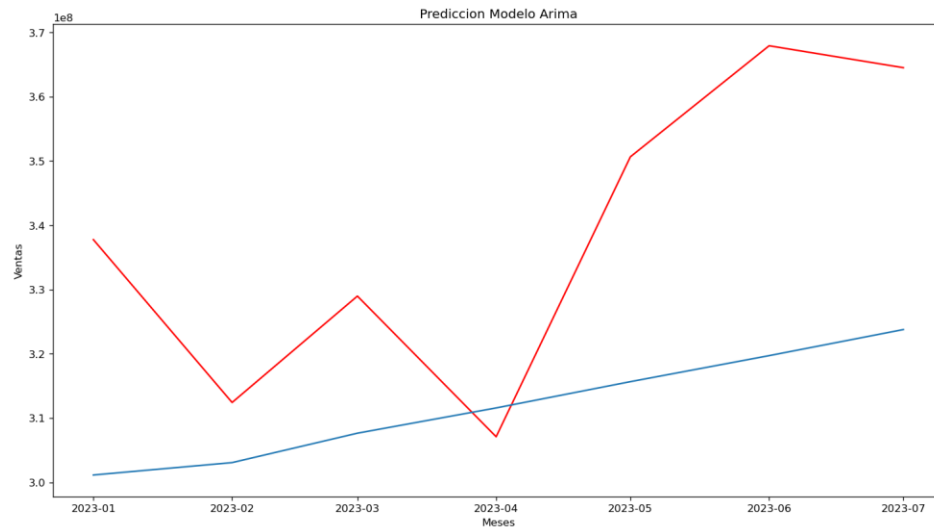


Imagen 22. ARIMA(1,2,1) vs ventas reales

En la imagen inferior se tiene un diagnóstico del modelo. Arriba a la izquierda: los errores residuales parecen fluctuar alrededor de una media de cero y tienen una varianza uniforme. Arriba a la derecha: la gráfica de densidad sugiere una distribución normal con media cero. Abajo a la izquierda: todos los puntos deben estar perfectamente alineados con la línea roja. Cualquier desviación significativa implicaría que la distribución está sesgada. Abajo a la derecha: el correlograma, también conocido como gráfico ACF, muestra que los errores residuales no están auto correlacionados.

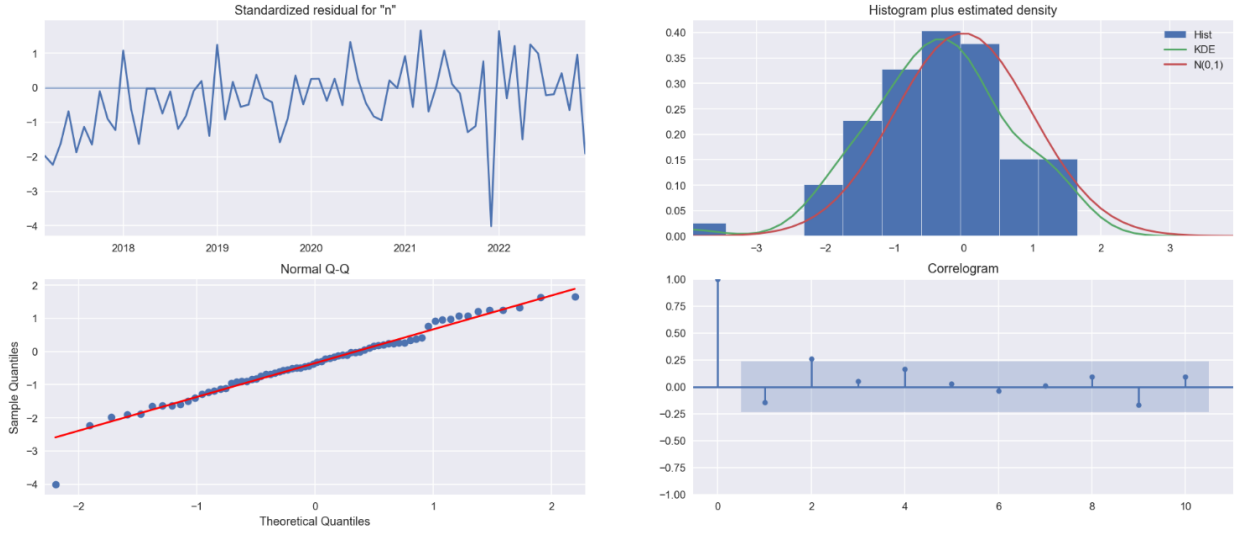


Imagen 23. Diagnóstico modelo ARIMA(1,2,1)

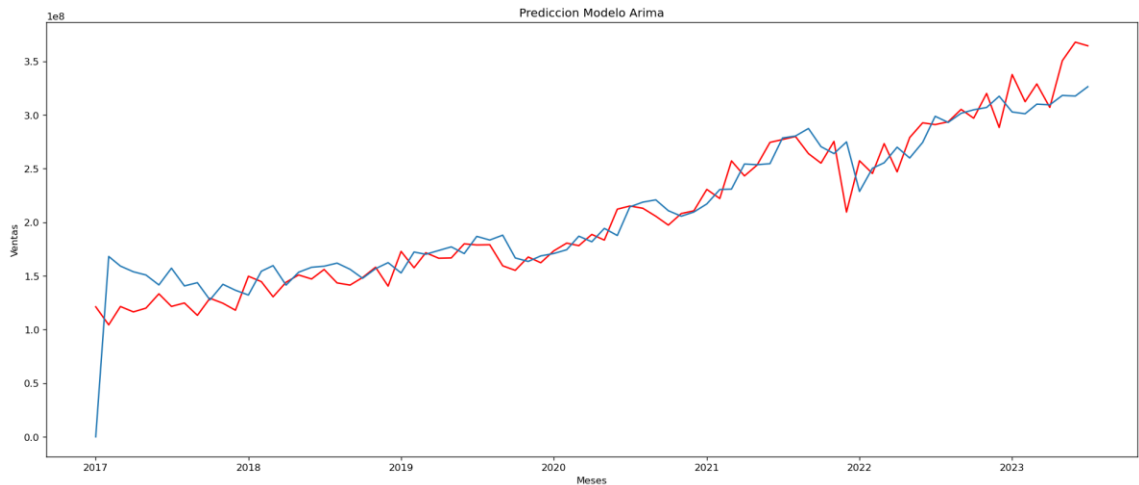
Las métricas de evaluación del modelo se representan en la tabla inferior:

Métrica	Resultado
MSE	1,016,251,030,925,030.00
MAE	27,964,437.25
RMSE	31,878,692.42
MAPE	8%
MASE	1.31

Tabla 7 Evaluación ARIMA(1,2,1)

El valor del MAPE está por encima del 5% establecido anteriormente por tanto el modelo tiene una capacidad predictiva mala. El MAE, MSE y RMSE se compara más adelante con los demás modelos. En la imagen inferior se observa la serie total y la predicción del modelo.

Imagen 24. Modelo ARIMA(1,2,1) vs Ventats Reales total serie



5.4.2 Modelo SARIMA

La descomposición de la serie muestra que existe estacionalidad, por ende, se prueba el modelo arima considerándola. Según se estableció anteriormente el modelo a seguir es $ARIMA(1,2,1) \times (1,0,1)$. En la gráfica inferior se observa cómo se ajustó mejor el modelo (verde) a los datos reales (azul).

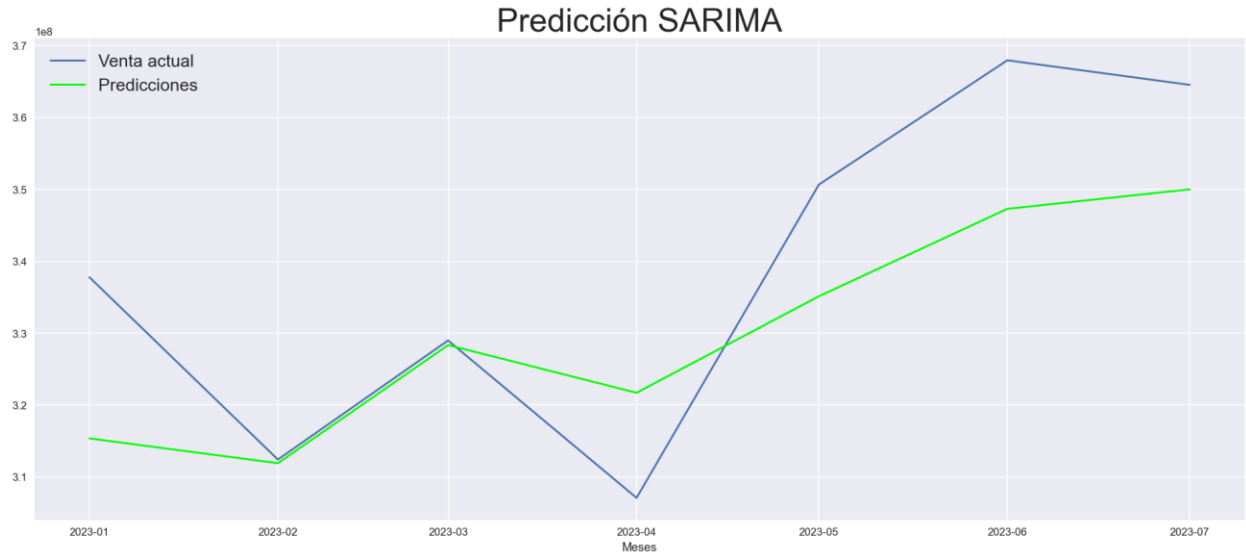


Imagen 25. Modelo $ARIMA(1,2,1)(1,0,1)$ vs Ventas Reales

En la imagen inferior se tiene un diagnóstico del modelo. Se tienen gráficos muy parecidos al modelo $ARIMA(1,2,1)$.

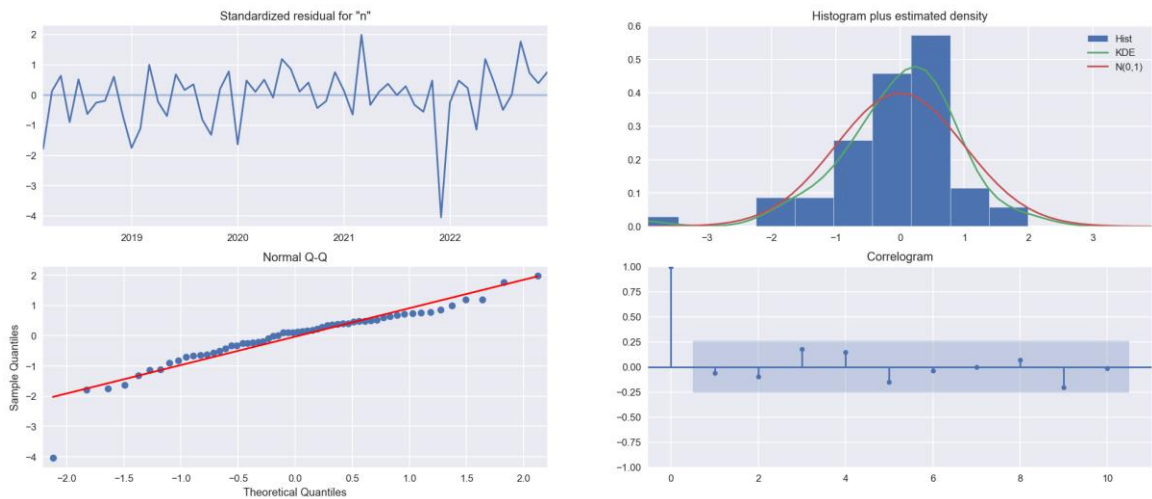


Imagen 26. Diagnóstico $ARIMA(1,2,1)(1,0,1)$

Las métricas de evaluación del modelo se representan en la tabla inferior:

Métrica	Resultado
MSE	227,862,377,627,780.00
MAE	12,697,327.74
RMSE	15,095,111.05
MAPE	3.68%
MASE	0.59

Tabla 8 Evaluación ARIMA(1,2,1)(1,0,1)

A diferencia del primer modelo arima, ahora el MAPE es de 3.68% que se encuentra en el rango aceptable. Ahora si se tiene un buen modelo predictivo.

A manera de visualización se muestra la serie completa, se puede apreciar como al inicio el modelo no se ajusta bien a los datos, pero después se comienzan a ajustar.



Imagen 27. Modelo ARIMA(1,2,1)(1,0,1) vs Ventas Reales total serie

5.4.3 Modelo Holt Winters

El modelo de suavización exponencial se muestra de la siguiente manera comparada con la serie real. A simple vista parece que el modelo aprendió bien de los datos.

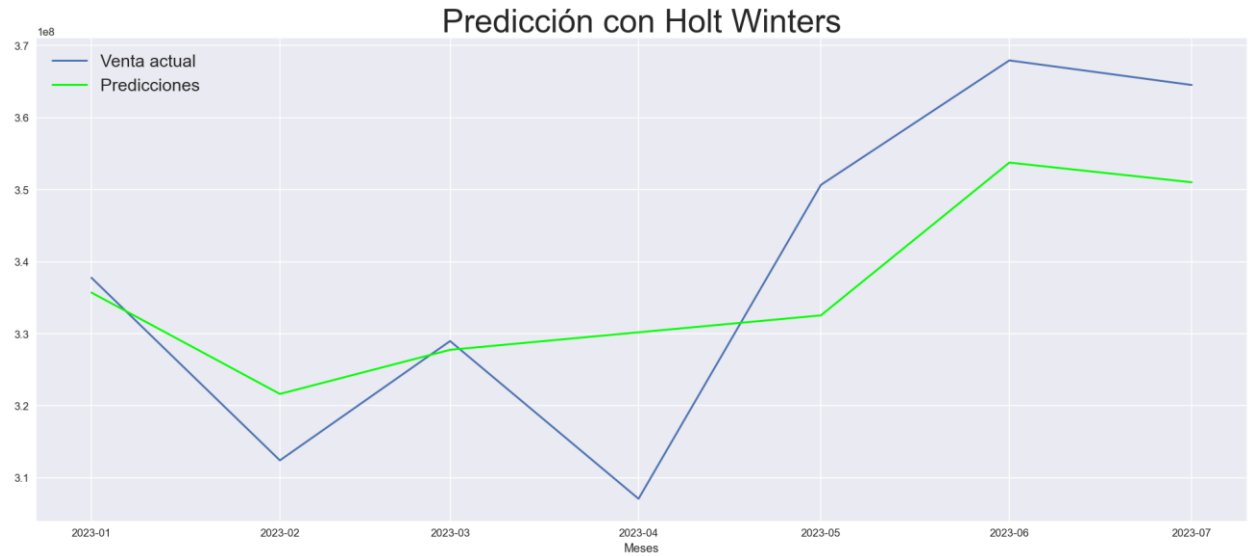


Imagen 28. Modelo Holt Winters vs Ventas Reales

Las métricas de evaluación del modelo se representan en la tabla inferior:

Métrica	Resultado
MSE	190,830,678,474,031.00
MAE	11,626,490.01
RMSE	13,814,147.76
MAPE	3.45%
MASE	0.54

Tabla 9 Evaluación Holt Winters

El modelo Holt Winters arroja un resultado muy bueno para el MAPE, por debajo del 5%, por tanto, el modelo si tiene una buena capacidad predictiva.

Luego se obtiene el comportamiento global en la imagen inferior. Se puede observar que, si logró capturar la estacionalidad, aunque con algunos puntos muy por encima o muy por debajo

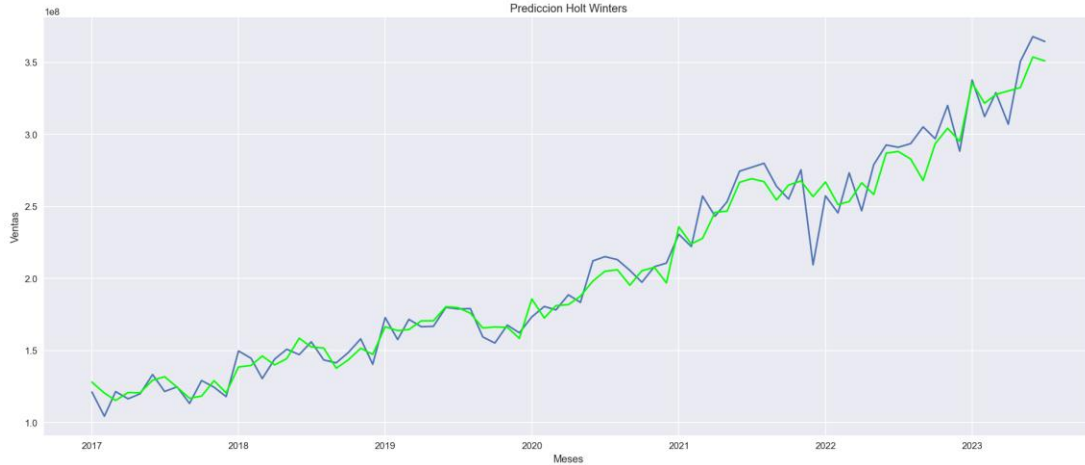


Imagen 29. Modelo Holt Winter vs Ventas Reales total serie

5.4.4 Modelo Prophet

El pronóstico del modelo de Facebook se muestra de la siguiente manera comparada con la serie de tiempo real. A simple vista captura muy bien la estacionalidad, sin embargo, queda muy por debajo en varios meses.

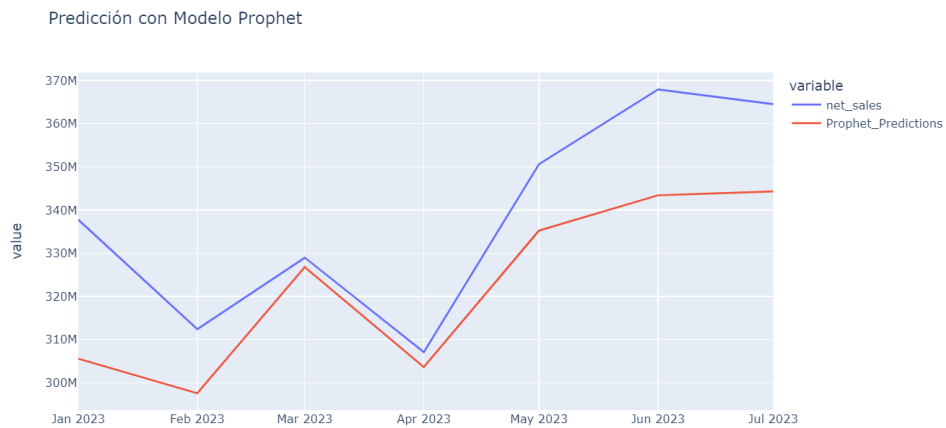


Imagen 30. Modelo Prophet vs Ventas Reales

Las métricas de evaluación del modelo se representan en la tabla inferior:

Métrica	Resultado
MSE	359,811,484,364,806.00
MAE	16,112,995.72
RMSE	18,968,697.48
MAPE	4.66%
MASE	0.75

Tabla 10 Evaluación Prophet Facebook

El modelo Prophet cuenta con un 4.66% de MAPE. Este modelo también se ajusta bien a los datos de la serie de tiempo.



Imagen 31. Modelo Prophet vs Ventas Reales total serie

5.4.5 Modelo LSTM

El pronóstico del modelo LSTM se muestra de la siguiente manera en la imagen inferior. A simple vista se puede apreciar que el modelo no aprendió bien, de hecho parece que tiene un comportamiento como el modelo ARIMA.



Imagen 32. Modelo LSTM vs Ventas Reales

Las métricas de evaluación se muestran en la siguiente tabla:

Métrica	Resultado
MSE	497,471,771,407,083.00
MAE	19,474,689.42
RMSE	22,304,075.21
MAPE	5.60%
MASE	0.91

Tabla 11 Evaluación LSTM

El LSTM tiene un MAPE de 5.60% lo que indica que es malo, sin embargo, comparado contra un modelo ingenuo es mejor, ya que tiene un MASE de 0.91.

5.4.6 Comparación de Resultados

En la imagen inferior se puede visualizar todos los modelos vs la venta real donde se puede decir que el mejor modelo es el Holt Winters.

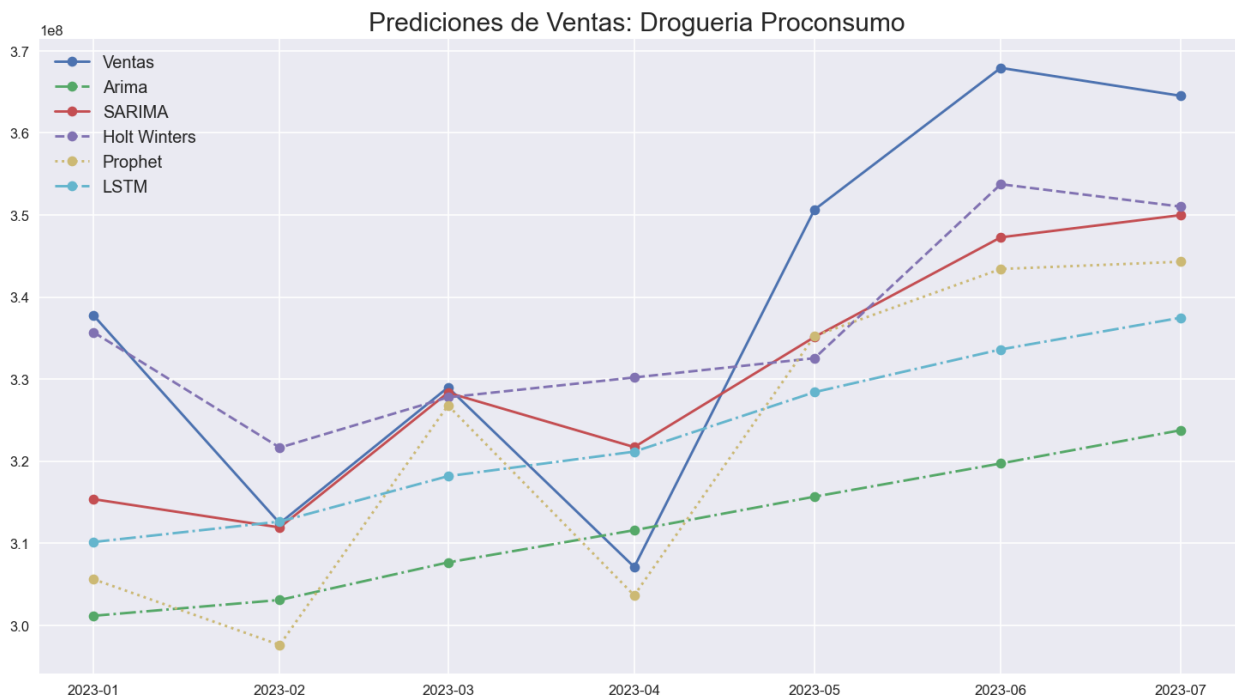


Imagen 33. Modelos vs Ventas Reales

En el cuadro inferior se muestra la comparación de las métricas de cada modelo.

Métrica	Resultado ARIMA	Resultado SARIMA	Resultado Holt Winters	Resultado Prophet	Resultado LSTM
MSE	1,016,251,030,925,030.00	227,862,377,627,780.00	190,830,678,474,031.00	359,811,484,364,806.00	497,471,771,407,083.00
MAE	27,964,437.25	12,697,327.74	11,626,490.01	16,112,995.72	19,474,689.42
RMSE	31,878,692.42	15,095,111.05	13,814,147.76	18,968,697.48	22,304,075.21
MAPE	8%	3.68%	3.45%	4.66%	5.60%
MASE	1.31	0.59	0.61	0.75	0.91

Tabla 12 Comparación de Todos los Modelos

Para comparar modelos, basta con buscar el que tenga valores más bajos para MSE, RMSE y MAE, en este caso, el mejor modelo fue el Holt Winters. Este modelo además como se mencionó anteriormente tiene un MAPE de 3.45% lo cual está muy bueno, así como un MASE de 0.61 que indica que es mejor modelo que un modelo ingenuo.

6 Conclusiones

- ✚ Cuando las empresas invierten en tecnología y son, el almacenamiento de su información y la extracción de la misma se vuelve más fácil y rápido. Droguería Proconsumo al tener un control de su información en PostgreSQL facilitó a extraer sus ventas agrupadas por mes. En otras empresas se tiene que descargar el histórico de sus ventas desde un sistema y luego agrupar la información, mientras que en este caso la identificación y extracción de los datos fue fácil, llevó segundos.
- ✚ Antes de aplicar cualquier técnica es importante entender los datos, ver su comportamiento por ejemplo si la serie de tiempo es estacionaria mediante pruebas de raíces unitarias o si los residuos se comportan de manera normal y realizar pruebas de ruido blanco. Esto es importante porque afecta el modelo predictivo, una pobre selección de parámetros tiende a arruinar el modelo y su precisión. Un ejemplo claro de entender los datos es el modelo ARIMA y el SARIMA, la capacidad de predicción aumento significativamente reduciendo en MAPE, esto se logró al descomponer la serie y observar que si tenía un componente estacional que no se toma en cuenta en el ARIMA.
- ✚ Finalmente, la selección del modelo depende del criterio del analista, en este caso se optó por seleccionar el modelo que tuviera valores de error más bajos en RMSE y MAE que son utilizados para comparar modelos. En este caso el modelo que mejor se adaptó a los datos fue el Holt Winters, dejando en segundo lugar el SARIMA(1,2,1)(1,0,1) luego el Prophet, seguidamente el LSTM y por último el ARIMA(1,2,1). El modelo prophet es un modelo prometedor ya que en este caso únicamente se realizó de la manera más sencilla, a futuro cabe destacar que se debe incluir otra parte de sus componentes que es la adaptación a los días festivos.

7 Referencias

- Ahuja, A. (11 de 01 de 2021). *Mean Absolute Scaled Error (MASE) in Forecasting*. Obtenido de <https://medium.com/@ashishdce/mean-absolute-scaled-error-mase-in-forecasting-8f3aecc21968>
- Alas, C. A. (2021). *Métodos para el Análisis Ecocómico*. San Salvador.
- de Arce, R., & Mahía, R. (2001). *MODELOS ARIMA*. Madrid.
- Dickey, D., & Pantula, S. (1987). Determining the Order of Differencing in Autoregressive Processes. *Journal of Business and Economic Statistics*, 8.
- Fattah, J., Ezzine, L., Aman, Z., Moussami, H. E., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*.
- IBM. (s.f.). *¿Qué son las redes neuronales recurrentes?* Obtenido de <https://www.ibm.com/es-es/topics/recurrent-neural-networks>
- IBM. (17 de 8 de 2021). *Conceptos básicos de ayuda de CRISP-DM*. Obtenido de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- IBM. (17 de 8 de 2021). *IBM Documentación*. Obtenido de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=data-autocorrelation-partial-autocorrelation-functions>
- Khair, U., Fahmi, H., Al Hakim, S., & Rahim4, R. (2017). Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error. *Journal of Physics*, 7.
- Krieguer, M. (9 de 2 de 2021). *Towards Data Science*. Obtenido de Time Series Analysis with Facebook Prophet: How it works and How to use it: <https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a>
- Mauricio, J. A. (2007). *Introducción al Análisis de Series Temporales*. Madrid.

- Minitab. (2023). *Métodos y fórmulas para la Método de Winters*. Obtenido de <https://support.minitab.com/es-mx/minitab/21/help-and-how-to/statistical-modeling/time-series/how-to/winters-method/methods-and-formulas/methods-and-formulas/#:~:text=El%20modelo%20aditivo%20es%3A,1%20%2B%20T%20t%E2%80%931%5D>
- R Project. (s.f.). *R Documentation*. Obtenido de OCSB test: <https://search.r-project.org/CRAN/refmans/seastests/html/ocsb.html>
- Ramírez, A. (2017). *RPubs by RStudio*. Obtenido de <https://rpubs.com/Arthurus/492144?ref=ciberseguridad.blog#:~:text=Intuitivamente%20e1%20concepto%20se%20refiere,a%20la%20hora%20de%20predecir>.
- Rastogi, M. (6 de 04 de 2020). *Towards Data Science*. Obtenido de Tutorial on LSTMs: A Computational Perspective: <https://towardsdatascience.com/tutorial-on-lstm-a-computational-perspective-f3417442c2cd#b10c>
- Ratre, S., & Jayaraj, J. (2023). Sales Prediction Using ARIMA, Facebook's Prophet and XGBoost Model of Machine Learning. *Machine Learning, Image Processing, Network Security and Data Sciences. Lecture Notes in Electrical Engineering*.
- Real, F. A. (2021). *Comparación de la metodología de Box - Jenkins y el método de Holt - Winters para el pronóstico de series de tiempo univariado*. Santiago.
- Rubino, N. (2022). *Clase Modelos Predictivos del curso "Master en Economía, Finanzas y Computación" - Modelos Autoregresivos*. Huelva.
- Sáez, J. I. (2018). *Análisis de Series de Tiempo*. Argentina.
- Sitio Big Data. (27 de 08 de 2018). *Sitio Big Data*. Obtenido de Aprendizaje automatico y las Metricas de regresión: <https://sitiobigdata.com/2018/08/27/machine-learning-metricas-regresion-mse/#:~:text=En%20MAE%2C%20el%20error%20se,por%20igual%20en%20el%20promedio>.
- Taylor, S. J., & Letham, B. (2017). Forecasting at Scale. *PeerJ Preprints*, 25.

Toro Ocampo, E., Mejía Giraldo, D., & Salazar Isaza, H. (2004). Pronóstico de Ventas Usando Redes Nueronales. *Scientia Et Technica*, 7.

Torres, J. (22 de 09 de 2019). *Redes Neuronales Recurrentes*. Obtenido de https://torres.ai/redes-neuronales-recurrentes/#Long-Short_TermMemory

Udom, P., & Phumchusri, N. (2014). A comparison study between time series model and ARIMA model for sales forecasting of distributor in plastic industry. *Journal of Engineering*, 7.

Universidad de Valladolid. (s.f.). *Universidad de Valladolid*. Obtenido de <http://www5.uva.es/estadmed/datos/series/series1.htm>

Universitat de Barcelona. (s.f.). *Universitat de Barcelona*. Obtenido de http://www.ub.edu/aplica_infor/spss/cap8-4.htm

Villavicencio, J. (s.f.). *Introducción a Series de Tiempo*. Puerto Rico.

8 Anexos

[TFM 22-23](#)

1. Anexo [Enlace para ver código de Python y HTML del mismo.](#)