



POSGRADOS

MAESTRÍA EN PRODUCCIÓN Y OPERACIONES INDUSTRIALES

RPC-SO-30-NO.506-2019

OPCIÓN DE TITULACIÓN:
ARTÍCULOS PROFESIONALES DE ALTO NIVEL

TEMA:
MEJORA DE LA PRECISIÓN EN LAS
PROYECCIONES DE VENTAS A
TRAVÉS DEL USO DE MODELOS
ESTADÍSTICOS EN UNA
CORPORACIÓN DE GUAYAQUIL

AUTOR(ES)
AXEL ALEXIS VÉLIZ DELGADO

DIRECTOR:
ANA FABIOLA TERÁN ALVARADO

GUAYAQUIL – ECUADOR
2023



Autor(es):



Axel Alexis Véliz Delgado
Ingeniero Industrial
Candidato a Magíster en Producción y Operaciones Industriales por
la Universidad Politécnica Salesiana – Sede Guayaquil.
axlevelizdelgado@hotmail.com

Dirigido por:



Ana Fabiola Terán Alvarado
Ingeniera Industrial
Magister en administración de empresas
ateran@ups.edu.ec

Todos los derechos reservados.

Queda prohibida, salvo excepción prevista en la Ley, cualquier forma de reproducción, distribución, comunicación pública y transformación de esta obra para fines comerciales, sin contar con autorización de los titulares de propiedad intelectual. La infracción de los derechos mencionados puede ser constitutiva de delito contra la propiedad intelectual. Se permite la libre difusión de este texto con fines académicos investigativos por cualquier medio, con la debida notificación a los autores.

DERECHOS RESERVADOS

2023 © Universidad Politécnica Salesiana.

GUAYAQUIL– ECUADOR – SUDAMÉRICA

Axel Alexis Véliz Delgado

MEJORA DE LA PRECISIÓN EN LAS PROYECCIONES DE VENTAS A TRAVÉS DEL USO DE
MODELOS ESTADÍSTICOS EN UNA CORPORACIÓN DE GUAYAQUIL

DEDICATORIA

Dedico este trabajo a mi familia por haberme apoyado en toda esta travesía, a mi mamá por siempre estar ahí para mí en todo momento y en especial a mi abuela y mi abuelo que están en los cielos y me alentaron a seguir siempre adelante en cualquier circunstancia en la que me encontrara, a mi mamá que me alentó a seguir adelante a pesar de las dificultades y a mis amigos más cercanos que me sacaron más de una risa en los momentos difíciles.

AGRADECIMIENTO

Agradezco a mi abuelo, abuela y a mi mamá por estar siempre a mi lado en todo momento apoyándome en cada etapa de este proceso en el que poco a poco fui avanzando hasta el final, también agradezco a mis amigos y compañeros del trabajo que me brindaron la ayuda que necesitaba.

TABLA DE CONTENIDO

Resumen	7
Abstract	8
1. Introducción	9
2. Determinación del Problema.....	11
3. Marco teórico referencial.....	13
4. Materiales y metodología.....	16
5. Resultados y discusión.....	31
6. Conclusiones.....	39
Referencias	42

MEJORA DE LA
PRECISIÓN EN LAS
PROYECCIONES DE
VENTAS A TRAVÉS DEL
USO DE MODELOS
ESTADÍSTICOS EN UNA
CORPORACIÓN DE
GUAYAQUIL

AUTOR(ES):

AXEL ALEXIS VÉLIZ DELGADO

RESUMEN

El pronóstico ha llegado a ser fundamental dentro de las compañías para tomar decisiones estratégicas en las áreas de ventas, producción, transporte y recursos a través de toda la cadena de suministros. Aunque los pronósticos de series de tiempo tradicionales prevalecen en los ámbitos de investigación y aplicación en las distintas actividades económicas a las que se dediquen las empresas, estos tienen varias limitaciones. Entonces, el uso de modelos estadísticos adecuados es esencial para mejorar la precisión de los pronósticos y es lo que se demuestra en este artículo con la finalidad de mejorar la precisión, proponiendo así, modelos adecuados con los que se puedan trabajar en cada una de las empresas estudiadas. Estos fueron evaluados y en los resultados de la investigación se evidenció que algunos de los métodos propuestos presentaron un rendimiento superior que los métodos tradicionales utilizados actualmente por las empresas participantes en el estudio.

Palabras clave:

Pronósticos, desempeño, estadística, series de tiempo, análisis de datos.

ABSTRACT

Forecasting has become fundamental within companies to take strategic decisions in sales, production, transportation and resources areas throughout the entire supply chain. Although, traditional time series forecasts are prevalent in research and application in the different economic activities in which companies focus, they have several limitations. Therefore, the use of adequate statistic modeling is essential to improve the accuracy of the forecasts and this was demonstrated in this article in order to improve the accuracy of the forecasts, thus proposing appropriate models to work in each of the studied companies. The models were evaluated and results showed that some of the proposed methods presented superior performance over currently used traditional methods.

Keywords:

Forecast, performance, simulation, time series, data analysis.

1. INTRODUCCIÓN

La globalización, la digitalización y el surgir de nuevas tecnologías en los últimos años, han acelerado los cambios en los gustos y necesidades de los consumidores, dificultando la tarea de predecir las tendencias de consumo por lo que conocer bien aquellas expectativas del mercado hace que sea muy importante realizar pronósticos más cercanos a la realidad. Los pronósticos de ventas son una parte importante en la administración de los negocios, dado que proveen de información relevante que puede ser usada para hacer buenas decisiones estratégicas, estos son definidos como la manera de predecir el futuro lo más preciso posible, usando toda la información disponible, incluyendo los datos históricos y el conocimiento de eventos futuros que afecten al pronóstico [1].

Actualmente, muchas empresas han tenido que realizar cambios en sus sistemas predictivos para aumentar la efectividad en las ventas debido a la necesidad de pronosticar a futuro y las empresas que se mencionan en este artículo no son la excepción. Sin embargo, las expectativas sobre la evolución del mercado han estado desajustadas con la realidad por lo que es necesario experimentar con mejores métodos que sirvan para suministrar los productos y servicios adecuados satisfaciendo las necesidades de los clientes. Los pronósticos de ventas precisos pueden aumentar las ganancias junto a la mejora de la competitividad en el mercado y también evitar inventarios deficientes, mal manejo de materiales, pérdidas en ventas, insatisfacción del cliente [2].

Las empresas que son sujetos de este estudio fueron creadas por una familia ecuatoriana hace más de 15 años, con tres líneas de negocios dedicadas respectivamente a la comercialización en el sector médico (Compañía A), industrial (Compañía B) y productivo (Compañía C) ganando un reconocimiento en el mercado, razón por la que realiza innovaciones en cada línea de productos de tiempo en tiempo. Actualmente, cuentan con algunas sucursales que permiten una distribución más sencilla al resto de provincias. A la actualidad desean mejorar el

área de compras dada la falta de un buen método que sirva para pronosticar adecuadamente las demandas de los distintos mercados a los que se dirigen cada una de las empresas.

2. DETERMINACIÓN DEL PROBLEMA

En las empresas se detectó una dificultad en obtener las cantidades correctas para realizar las compras de las cantidades adecuadas de productos y solventar la demanda del mercado, esto impulsa la necesidad de mejorar la manera en la que se está pronosticando el comportamiento de los clientes con respecto al tiempo planteando así la pregunta al problema:

¿Cómo mejorar la precisión de las proyecciones de ventas utilizando modelos estadísticos?

Los problemas específicos a resolver son:

- ¿Cuál es el porcentaje de cumplimiento de la meta proyectada previo a la experimentación con modelos estadísticos?
- ¿Qué modelos estadísticos para mejorar la precisión de proyecciones serán los más adecuados para cada empresa de la corporación?
- ¿Se puede mantener un mismo modelo estadístico para una empresa conforme al paso del tiempo?

La ejecución de este estudio, se debió a que actualmente las empresas de este grupo corporativo han usado una proyección basada en los históricos de ventas con un promedio móvil simple, sin embargo, el mercado ha resultado bastante volátil en estos últimos tiempos por lo que aquel método tiene una precisión menor a la que modelos estadísticos podrían llegar, causando grandes inconvenientes para cumplir con las metas de ventas.

Este trabajo tuvo como objetivo el fin de mejorar las proyecciones de ventas, por medio de un análisis de datos y pruebas con modelos estadísticos.

El enfoque estuvo en mejorar los pronósticos de las empresas dentro de la corporación, para poder utilizar de manera adecuada los recursos necesarios en la operación y tener lo suficiente en inventario acorde a las necesidades del mercado.

Como objetivo general se definió: “Mejorar la precisión en las proyecciones de ventas en las empresas dentro de la corporación con el fin de comprar los productos adecuados para la comercialización y la materia prima para la producción, por medio de la comparación de modelos estadísticos.”

Y los objetivos específicos con los que se abarcaron el problema fueron:

- Verificar la eficiencia de la metodología actual aplicada para realizar las proyecciones de ventas en las empresas de la corporación, analizando los resultados proyectados y reales.
- Aplicar diferentes modelos estadísticos comparando los datos de pronósticos de ventas de las empresas de la corporación.
- Seleccionar el modelo estadístico adecuado que mejore la precisión en las proyecciones de ventas por cada empresa de la corporación.

Como hipótesis general se definió: “Al implementar un modelo estadístico, la empresa mejorará la precisión de sus proyecciones”

Y las hipótesis específicas con los que se abarcaron el problema fueron:

- La precisión de las proyecciones de ventas es mayor al usar modelos estadísticos en comparación con los métodos tradicionales.
- El tamaño de la muestra influye en la precisión de las proyecciones de ventas al usar modelos estadísticos.
- La frecuencia de actualización de los datos utilizados en los modelos estadísticos tiene un efecto significativo en la precisión de las proyecciones.

Cada uno de estos fue la guía para encontrar la solución al problema.

3. MARCO TEÓRICO REFERENCIAL

En la actualidad, formular pronósticos es un elemento esencial para el éxito de las empresas de todos los sectores puesto que, con estos datos, se puede tener una idea de la cantidad que se venderá en los siguientes periodos del año. Además, impactan las decisiones en producción, transporte y personal en todos los horizontes dentro de la cadena de suministro [3] [4] [5]. Si esta predicción se realiza mal, esto resultará en pérdidas de ventas junto a la pérdida de reputación y de ingresos [6] [7].

Uno de los factores principales para una administración correcta en ventas es la habilidad de pronosticar para uno, dos trimestres e incluso para el año entero. De hecho, la administración adecuada del inventario juega un rol importante en la administración operacional de las empresas comerciales. En un periodo de planificación, el exceso de inventario ocupa grandes presupuestos de operaciones aumentando los costos y riesgos, muy poco inventario no satisficará a los clientes y afectará a la reputación de la empresa resultando en la pérdida de clientes y baja competitividad [8].

Tener un pronóstico perfecto es virtualmente imposible debido a que hay muchos factores en el ambiente de los negocios que son difíciles de predecir con exactitud. Entonces, en lugar de buscar el pronóstico perfecto, se debe establecer la práctica de renovar pronósticos continuamente encontrando el método disponible más adecuado dentro de lo razonable [9].

Para aumentar la ventaja competitiva, las compañías requieren una estricta administración de compras, producción y distribución, esto es manifestado por R. Majhi et. al. que utilizaron un "PSO basado en un modelo adaptable *ARMA (Auto Regressive Moving Average)*" donde se demostró que el modelo con este método dio resultados muy precisos y menor valor de errores con la realidad en las ventas a diferencia de otros métodos con los que se comparó [10]. Por otro lado, X. Dairu

y Z. Shilong utilizaron el modelo optimizado *XGBoost* para realizar un modelo de predicciones en la cadena de compañías *Walmart* donde se notó una mejora en la rapidez del análisis de datos y precisión en las predicciones [11], y en la misma cadena comercial Z. Qiao creó el modelo "*LightGBM*" el cual resultó en el que más precisión tuvo de los tres que investigó [12]. Otro grupo de estudio en la misma cadena de tiendas conformado por J. Chen et. Al. propusieron un nuevo modelo basado en redes neuronales con orígenes en sitios web, superando a los algoritmos de regresión lineal y de *SVM* [13].

E. Spiliotis, et. Al. Aseguran que los pronósticos probabilísticos para las ventas de productos son indispensables para una efectiva planificación de almacenamiento y manejo de inventario [14], sin embargo, no se suelen aplicar los mismos modelos para cada industria, como es el caso de J. Marcoux y S. Selouani los cuales se enfocaron en la industria de los videojuegos y crearon el sistema subespacial-conexionista híbrido de pronósticos utilizando información de dos desarrolladores de videojuegos donde aumentó la precisión de los pronósticos de ventas [15]. Por otro lado M. Shengdong et. Al. Consiguieron utilizar adecuadamente el modelo *SICM* para la industria agrícola el cual resultó en una reducción significativa de los errores en los pronósticos [16].

A pesar de todo esto, L. Berry, P. Helman y M. West buscaron un modelo que se pueda adaptar a cualquier tipo de industria concluyendo que el "Modelo mixto de conteo dinámico (*DCMMs*)" era uno que se podía adaptar a distintas industrias utilizando variaciones para el desarrollo del método, este era un nuevo marco de trabajo del modelamiento del estado espacial Bayesiano en series de tiempo transacciones-ventas heterogéneas [17].

Las grandes empresas acumulan cantidades enormes de información sobre las ventas a través de sus máquinas en los diferentes puntos de venta, sin embargo, la información a nivel de inventario suele ser escasa dado que los productos en cada tienda pueden cambiar rápidamente debido al aumento del mercado competitivo [18]. Muchos de estos cambios se suelen dar por las temporadas o eventos festivos por lo que existe una alta volatilidad de inventario. Normalmente, cuando una

empresa tiene la intención de implementar un sistema de pronósticos existe una tendencia a replicar los conceptos, metas y principios de otras compañías para acelerar el proceso de implementación, pero existe el riesgo de que metas irreales y errores de medición sean adoptados sin tomar en cuenta la diferencia del entorno, por lo que V. Sohrabpour, et. al. desarrollaron una predicción utilizando inteligencia artificial logrando una correlación del 98% entre la predicción y las ventas reales que también resultó en una mejora para los costos de ventas [19].

C. Zhang et. al., al igual que otros autores debido a la competencia implacable a la cual las empresas deben enfrentarse y que de ello dependa su posición en la industria, ha tomado en cuenta que la precisión en las proyecciones se ha vuelto el eje principal en las compañías, y en su trabajo generaron un modelo a partir de la información compilada por un motor de búsqueda basándose en una combinación de “componente principal de análisis (PCA), Red neuronal de propagación trasera (BPNN) y el algoritmo de optimización ‘fruit fly’ mejorado (DSFOA) junto a la información de búsquedas en la web y el histórico de ventas como los datos de entrada para el experimento. Como resultado de esto, los pronósticos tuvieron un acierto promedio del 97% en 14 objetos distintos los cuales usaron como muestra para llevar a cabo la ejecución del modelo [20].

Los estudios descritos en los párrafos anteriores aportan a la presente investigación, dado que, los resultados se vinculan con los objetivos planteados, evidenciándose que, realizar pronósticos de la manera tradicional usando solamente los históricos de datos anteriores, llevará a resultados imprecisos por lo que es necesario encontrar un método que optimice las proyecciones que se tengan de las líneas de productos.

4. MATERIALES Y METODOLOGÍA

En este estudio se utilizaron como muestras los datos de ventas de los 20 productos más vendidos por cada una de las tres compañías en los distintos sectores en los que participan cada una, los cuales representan un 82,90% (compañía A), 73,71% (compañía B) y 75,49% (compañía C) de las ventas anuales. Cada una se desempeña en un sector de mercado específico, la compañía A en el sector médico, la compañía B al sector industrial - mecánico y la compañía C al sector industrial - químico. Los datos que se tomaron para este estudio constan desde enero 2019 hasta octubre 2022 y se pronosticaron las ventas en unidades desde noviembre 2022 hasta octubre 2023.

Los datos mencionados se utilizaron debido a que la empresa se reorganizó y los datos anteriores a las fechas usadas tenían una estructura diferente a la actual, dificultando la identificación de los datos relevantes, la comparación de los datos entre sí y la integración de los datos en un único sistema.

El modelo de pronóstico actual usado en todas las empresas ha sido un promedio móvil simple con una meta propuesta al 90% respecto a la venta de los productos escogidos, esto se midió comparando el valor vendido contra el proyectado.

$$f_t = \frac{1}{n} \sum_{i=1}^n d_{t-i} \quad (1)$$

Donde:

- f_t : Pronóstico del siguiente periodo
- n : número de periodos
- d_{t-i} : demanda de los periodos anteriores

Se observa en la **Tabla 1** “Eficiencia de método actual en compañía A” su efectividad en el paso del tiempo.

Tabla 1 Eficiencia de método actual en compañía A

Mes	Precisión - 2021	Cumplimiento meta
Ene	88.83%	No
Feb	85.52%	No
Mar	87.48%	No
Abr	90.67%	Si
May	78.49	No
Jun	73.75%	No
Jul	69.64%	No
Ago	69.32%	No
Sep	75.10%	No
Oct	79.16%	No
Nov	71.33%	No
Dic	82.37%	No

Tal como se puede ver, los pronósticos realizados con el promedio móvil simple no son del todo precisos y han causado pérdidas en ventas puesto que no tomaron en cuenta variables como la estacionalidad y la demanda del mercado. Ahora, con el fin de realizar las simulaciones de los pronósticos con una mejor facilidad y precisión a la hora de calcular datos se utilizó el programa *RStudio*, el cual permite realizar cálculos complejos de forma rápida y sencilla, siendo esto algo esencial en un mercado dinámico.

Para definir los modelos a usar, se graficaron los datos de ventas de las tres empresas con sus productos más destacados determinando el comportamiento de los mismos. Estos datos se cargaron al programa para su tratamiento y uso.

Fig. 1 Ventas de compañía A (unidades)

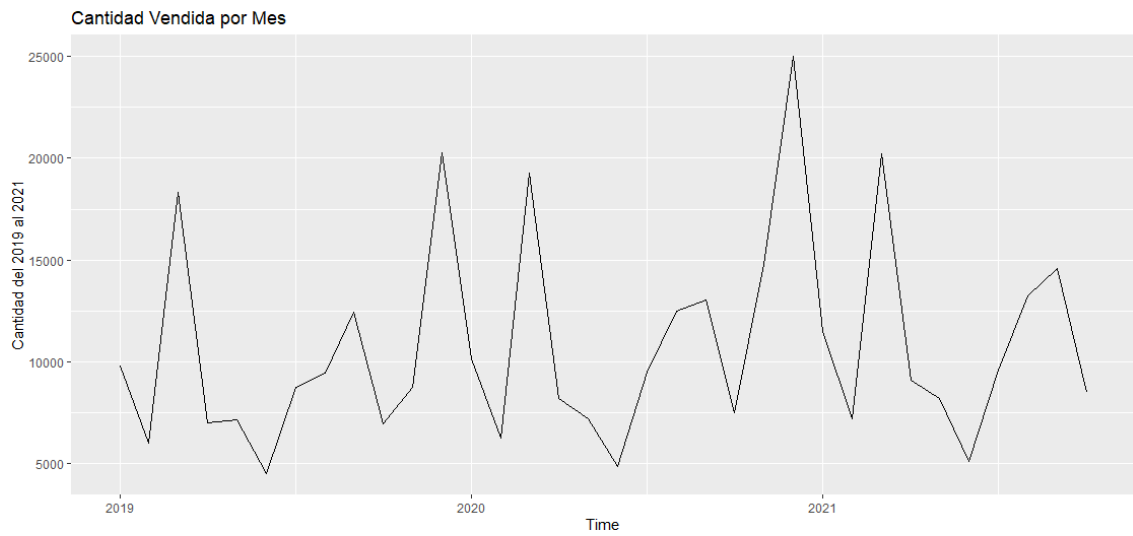


Fig.2 Ventas de compañía B (unidades)

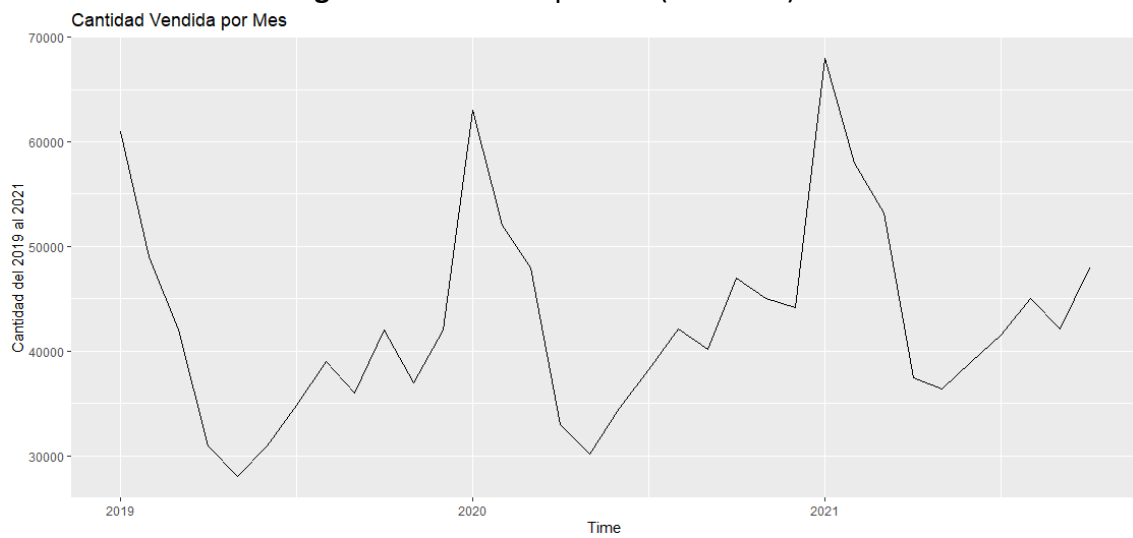
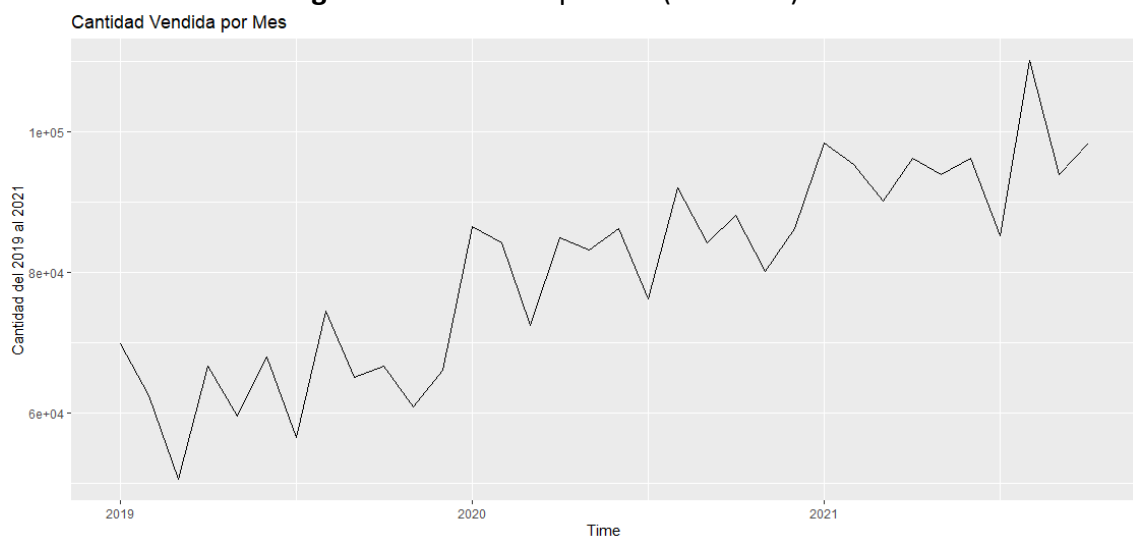


Fig. 3 Ventas de compañía C (unidades)



Dentro del último año se observó en la figura 1 que las ventas de la compañía A tuvo un crecimiento del 9,17%, en la figura 2 se observó que la compañía B tuvo un crecimiento de ventas del 19,63% y en la figura 3 se observó un crecimiento de ventas del 7,22% en la compañía C, además de la identificación de posibles patrones estacionales.

Para determinar si existe estacionalidad en los datos, se los transformó para deshacer la tendencia, esto se realiza tomando la diferencia entre un mes y el siguiente.

$$Dif = (y - y_{-1}) \tag{2}$$

Donde:

- Dif: Diferencia entre el valor actual y el valor anterior
- y: valor actual
- y-1: valor anterior

Fig. 4 Ventas de compañía A sin tendencia

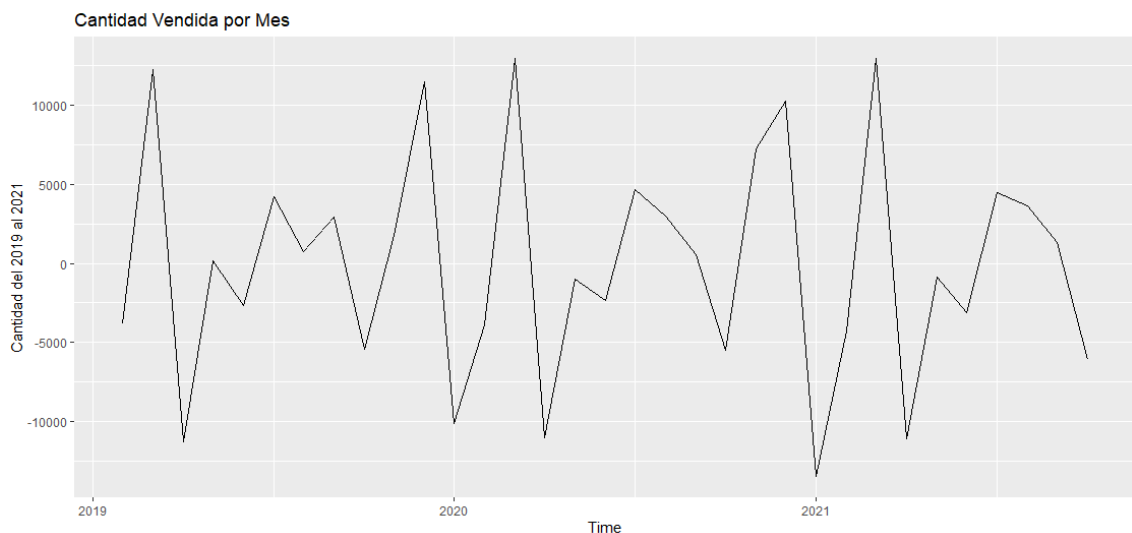


Fig. 5 Ventas de compañía B sin tendencia

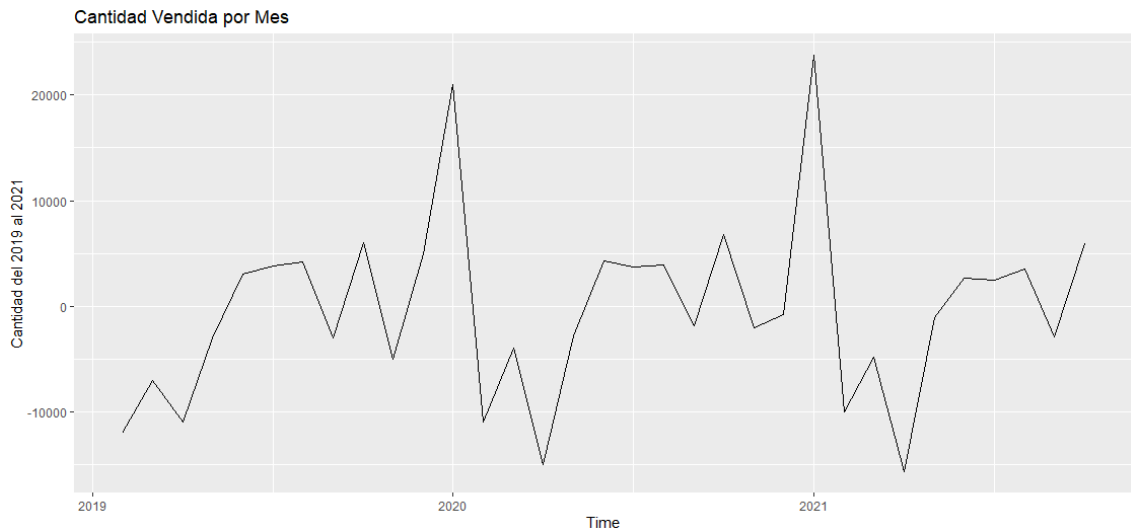
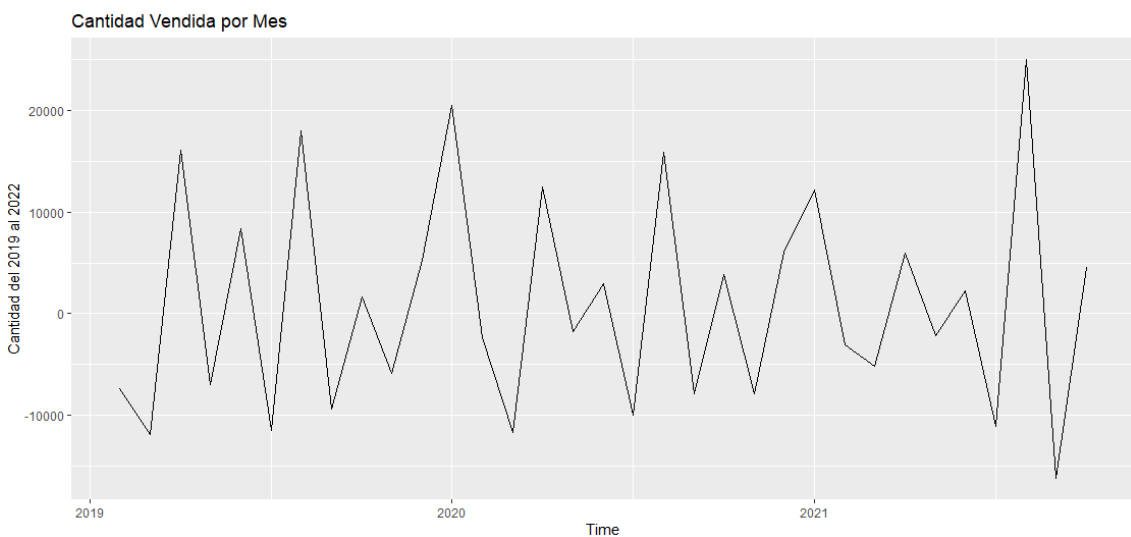


Fig. 6 Ventas de compañía C sin tendencia

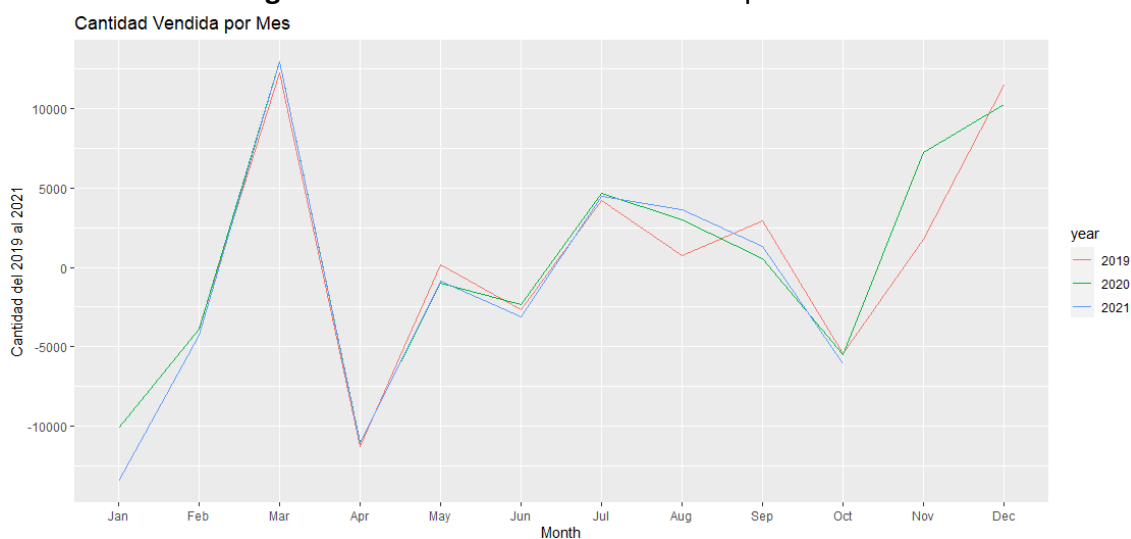


En la figura 4, figura 5 y figura 6 se identificó que existieron picos altos y bajos de ventas similares en cada año en cada una de las compañías indicando una posible estacionalidad y para comprobar si esto es correcto, se realizó un análisis de la estacionalidad de los datos a través de una gráfica que mostraba los datos de todos los años donde se comprobó la existencia de una estacionalidad fija.

Tabla 2 Estacionalidad de ventas en compañía A (unidades)

Mes	2019	2020	2021	2022
Ene	9820	10140	11500	12320
Feb	6010	6250	7236	8100
Mar	18305	19250	20215	22001
Abr	7000	8201	9100	9505
May	7175	7200	8210	9800
Jun	4502	4852	5100	6501
Jul	8745	9520	9610	10256
Ago	9470	12501	13250	11020
Sep	12428	13025	14580	14200
Oct	6965	7520	8510	9310
Nov	8760	14760	15141	-
Dic	20261	25014	32140	-

Fig. 7 Estacionalidad en ventas de compañía A

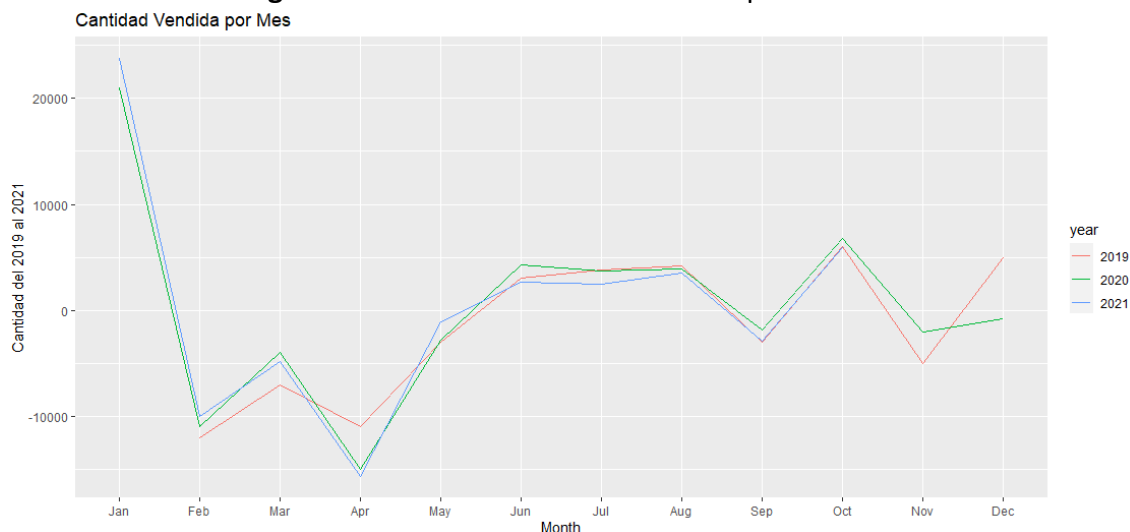


En la tabla 2 y la figura 7 se comprobó que existe estacionalidad en el conjunto de datos de la compañía A, donde tiene picos altos de demanda entre los meses de febrero a marzo y los meses de noviembre a diciembre, incluso los picos de baja demanda en el mes de abril debido al comportamiento del mercado.

Tabla 3 Estacionalidad de ventas en compañía B (unidades)

Mes	2019	2020	2021	2022
Ene	61000	63000	68000	125000
Feb	49000	52000	58000	120140
Mar	42000	48000	53200	110003
Abr	31000	33021	37503	112520
May	28000	30210	36400	109521
Jun	31000	34500	39021	114201
Jul	34800	38200	41500	98521
Ago	39000	42100	45012	130214
Sep	36000	40215	42102	118201
Oct	42010	47000	48000	122141
Nov	37000	45000	47000	-
Dic	42000	44200	46000	-

Fig. 8 Estacionalidad en ventas de compañía B

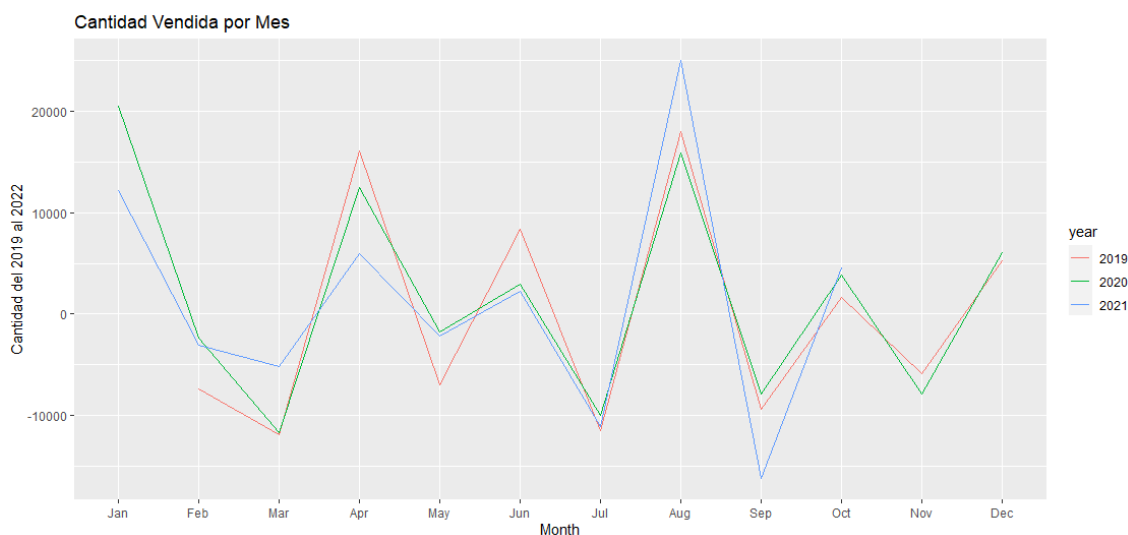


En la tabla 3 y figura 8 para la compañía B también se comprobó la existencia de una estacionalidad dado que se evidencian picos altos de la demanda en los meses de enero y octubre, además de los picos bajos de la demanda en los meses de febrero y abril, esto se observó con una recurrencia en los 3 años donde se analizaron los datos.

Tabla 4 Estacionalidad de ventas en compañía C (unidades)

Mes	2019	2020	2021	2022
Ene	69800	86600	98520	73021
Feb	62400	84200	95410	65014
Mar	50500	72512	90215	60124
Abr	66600	85010	96210	40125
May	59600	83250	94012	38120
Jun	68000	86201	96252	42020
Jul	56500	76250	85212	44000
Ago	74500	92123	110252	48542
Sep	65100	84212	94012	46852
Oct	66700	88102	98521	49025
Nov	60800	80215	90412	-
Dic	66100	86320	95200	-

Fig. 9 Estacionalidad en ventas de compañía C



Se evidenció en la tabla 4 y figura 9 que existe una estacionalidad en la compañía C, donde se evidencian picos altos de demanda en los meses de abril, junio y agosto, pero también picos bajos en los meses de marzo, julio y septiembre debido al comportamiento del mercado.

Con el fin de entrenar los modelos y evitar sobreajustes, se realizó una división de datos para realizar un entrenamiento con cada uno de los modelos dentro del código definiendo que este se realice con un 80% de entrenamiento y 20% de prueba.

Fig. 10 Prueba de validación cruzada dentro del código

```
# Split Data 80/20
cat("Split data (Entrenamiento 80% - Test 20%)\n"); cat("\n")
splits <- initial_time_split(.Data, prop = 1 - (.MesesPrediccion/nrow(.Data)))
```

Para pronosticar la demanda futura se seleccionan modelos de series temporales y los modelos más apropiados que se seleccionaron para estos casos son:

Auto Arima. – El modelo *ARIMA (AutoRegressive Integrated Moving Average)* se basa en una ecuación matemática que combina componentes de regresión autorregresiva (AR), diferenciación (I) y media móvil (MA) para usarla con el fin de modelar la tendencia y estacionalidad de una serie temporal. Este modelo selecciona automáticamente los parámetros de esta ecuación que mejor se ajustan a los datos [21].

Los hiperparámetros utilizados son:

- p: el número de retardos AR. Valor utilizado = 1
- d: el número de diferencias. Valor utilizado = 0
- q: el número de retardos MA. Valor utilizado = 1
- P: el número de retardos AR de la estacionalidad. Valor utilizado = 1
- D: el número de diferencias de la estacionalidad. Valor utilizado = 0
- Q: el número de retardos MA de la estacionalidad. Valor utilizado = 1

XG Boosted Auto Arima: Este modelo utiliza el algoritmo de *XGboost* para mejorar el rendimiento del modelo, este es un algoritmo de aprendizaje automático que combina las ventajas del modelo ARIMA con las ventajas del algoritmo XGBoost. El modelo ARIMA se utiliza para identificar la tendencia y la estacionalidad de la serie

de tiempo, mientras que el algoritmo XGBoost se utiliza para realizar el pronóstico [22].

Los hiperparámetros de XGBoosted Auto ARIMA son los siguientes:

- p: el número de retardos AR. Valor utilizado = 1
- d: el número de diferencias. Valor utilizado = 0
- q: el número de retardos MA. Valor utilizado = 1
- P: el número de retardos AR de la estacionalidad. Valor utilizado = 1
- D: el número de diferencias de la estacionalidad. Valor utilizado = 0
- Q: el número de retardos MA de la estacionalidad. Valor utilizado = 1
- nrounds: el número de iteraciones del algoritmo XGBoost. Valor utilizado = 100
- eta: el tamaño del paso del algoritmo XGBoost. Valor utilizado = 0.1
- gamma: el parámetro de regularización L2 del algoritmo XGBoost. Valor utilizado = 0
- alpha: el parámetro de regularización L1 del algoritmo XGBoost. Valor utilizado = 0
- colsample_bytree: la proporción de las variables que se utilizan para cada árbol. Valor utilizado = 1
- subsample: la proporción de los datos que se utilizan para cada árbol. Valor utilizado = 1

ETS (Suavizamiento exponencial triple). – Es un método estadístico que combina tres medidas sucesivas de una serie temporal con el fin de producir una estimación más precisa y suave de la tendencia subyacente asignando pesos decrecientes a medidas más antiguas para dar a entender que la información reciente es la más relevante suavizando las fluctuaciones aleatorias [23]:

$$l_x = \alpha * (y_x - s_{x-L}) + (1 - \alpha) * (l_{x-1} + b_{x-1}) \text{ Nivel} \quad (4)$$

Donde:

- l_x : valor suavizado de la serie de tiempo en el periodo x

- α : coeficiente de suavizado que controla la velocidad a la que se adapta el modelo a nuevos datos
- y_x : valor real en la serie de tiempo x
- s_{x-L} : valor suavizado de la tendencia en el periodo x-L
- l_{x-1} : valor suavizado de la tendencia en el periodo x-1
- b_{x-1} : valor suavizado de la estacionalidad en el periodo x-1

$$b_x = \beta * (l_x - l_{x-1}) + (1 - \beta) * b_{x-1} \text{ Estacionalidad} \quad (5)$$

Donde:

- b_x : valor suavizado de la estacionalidad en el periodo x
- β : coeficiente de suavizado que controla la velocidad a la que se adapta el modelo a nuevos datos
- l_x : valor suavizado de la tendencia en el periodo x
- l_{x-1} : valor suavizado de la tendencia en el periodo x-1
- b_{x-1} : valor suavizado de la estacionalidad en el periodo x-1

$$s_x = \gamma * (y_x - l_x) + (1 - \gamma) * s_{x-L} \text{ Tendencia} \quad (6)$$

Donde:

- S_x : valor suavizado de la tendencia en el periodo x
- γ : coeficiente de suavizado que controla la velocidad a la que se adapta el modelo a los nuevos datos
- y_x : el valor real de la serie de tiempo en el período x
- l_x : el valor suavizado de la tendencia en el período x
- s_{x-L} : el valor suavizado de la tendencia en el período x-L

$$\hat{y}_{x+m} = l_x + m * b_x + s_{x-L+(m-1)modL} \text{ Pronóstico} \quad (7)$$

Donde:

- \hat{y}_{x+m} : valor pronosticado para el periodo x+m

- I_x : valor suavizado de la tendencia en el periodo x
- m : número de periodos que se pronostican
- b_x : valor suavizado de la estacionalidad en el periodo x
- s_{x-L} : el valor suavizado de la tendencia en el período $x-L$
- $modL$: indica el módulo de la diferencia entre $x-L$ y $m-1$. Esto se utiliza para garantizar que el valor suavizado de la estacionalidad se calcula para el período correcto.

Los hiperparámetros utilizados son:

- α : Valor utilizado =0.1
- β : Valor utilizado =0.06
- γ : Valor utilizado =12

Prophet y boosted Prophet. – Este modelo propuesto por *Meta* se basa en un modelo de tendencia no lineal utilizando una combinación de regresión y análisis estadístico para generar pronósticos a corto y largo plazo. Se basa en modelos que pueden integrar los antecedentes de una investigación y conocimiento estadístico para realizar predicciones de alta precisión con parámetros simples e intuitivos, pero también puede recibir impacto de características en estacionalidad y feriados [24].

Los hiperparámetros de Prophet son los siguientes:

- ***growth (crecimiento)***: el tipo de crecimiento que se utiliza para modelar la tendencia. Los valores posibles son "*linear*", "*logistic*", "*quadratic*", "*trigonometric*" y "*flat*". Valor utilizado = "*linear*"
- ***changepoint_prior_scale (escala de probabilidad previa de cambio de punto)***: el parámetro de regularización que controla la probabilidad de que ocurra un cambio de punto en la tendencia. Valor utilizado = 0.05
- ***seasonality (estacionalidad)***: el tipo de estacionalidad que se utiliza para modelar la serie de tiempo. Los valores posibles son "*additive*" y "*multiplicative*". Valor utilizado = "*additive*"

- **Seasonality period (período de estacionalidad):** el período de la estacionalidad. Valor utilizado = 7
- **holidays (días festivos):** una lista de días festivos que se utilizan para modelar la estacionalidad. Valor utilizado = NULL
- **Holidays prior scale (escala de probabilidad previa de días festivos):** el parámetro de regularización que controla la probabilidad de que ocurran días festivos en la serie de tiempo. Valor utilizado = 0.05
- **Yearly seasonality (estacionalidad anual):** un booleano que indica si se debe usar una estacionalidad anual. Valor utilizado = TRUE
- **Weekly seasonality (estacionalidad semanal):** un booleano que indica si se debe usar una estacionalidad semanal. Valor utilizado = TRUE
- **Daily seasonality (estacionalidad diaria):** un booleano que indica si se debe usar una estacionalidad diaria. Valor utilizado = FALSE
- **Holidays included (días festivos incluidos):** un booleano que indica si se deben incluir los días festivos en el conjunto de datos de entrenamiento. Valor utilizado = FALSE

MARS. – Se basa en una combinación de regresión lineal y *splines* no lineales para generar pronósticos, siendo similar a un modelo de regresión lineal generalizado, pero tiene la ventaja de ser capaz de manejar relaciones no lineales entre las variables. Simula el modelo usando funciones básicas (BF) que se definen en la forma de pares basándose en un nudo que define una región de inflexión.

Los hiperparámetros de *MARS* son los siguientes:

- **degree (grado):** el grado de los splines que se utilizan. Los valores posibles son 1, 2, 3, etc. Valor utilizado = 2
- **interaction.depth (profundidad de interacción):** la profundidad de las interacciones que se permiten. Los valores posibles son 1, 2, 3, etc. Valor utilizado = 2
- **nprune (n-recorte):** el número de términos que se eliminan del modelo durante el proceso de poda. Valor utilizado = 10

- **penalty (penalización):** el tipo de penalización que se utiliza para evitar el sobreajuste. Los valores posibles son "none", "l1" y "l2". Valor utilizado = l2
- **γ (lambda):** el parámetro de penalización. Valor utilizado = 0.1

Random Forest. – Se basa en un conjunto de árboles de decisión para generar pronósticos, el algoritmo construye varios árboles de decisión utilizando diferentes subconjuntos de datos de entrenamiento y luego combina los resultados de estos árboles para generar una predicción final presentando pronósticos más acertados [25].

$$y = \frac{1}{B} \sum_{b=1}^B t_b(x) \quad (8)$$

Donde:

y: Pronóstico

B: Muestras de entrada (*Bootstrap*) del set de entrenamiento

t_b : Árboles usando las impurezas de Gini en las muestras

X: ejemplos no visualizados

Los hiperparámetros de Random forest son los siguientes:

- **n_{tree} (número de árboles):** el número de árboles que se utilizan en el conjunto. Valor utilizado = 200
- **m_{try} (número de variables a considerar):** el número de variables que se consideran para cada división en cada árbol. Valor utilizado = la raíz cuadrada del número de variables
- **maxdepth (profundidad máxima):** la profundidad máxima de los árboles. Valor utilizado = 5
- **minsplit (mínimo de divisiones):** el número mínimo de observaciones necesarias para dividir un nodo. Valor utilizado = 2
- **minbucket (mínimo de observaciones por nodo):** el número mínimo de observaciones necesarias para que un nodo sea un nodo hoja. Valor utilizado = 1

- **Random state (estado aleatorio):** la semilla que se utiliza para generar los números aleatorios. Valor utilizado = NULL

Cada uno de ellos tiene sus ventajas y desventajas, por lo que seleccionar el modelo más adecuado depende de la naturaleza de los datos, el objetivo específico de la predicción y sobre todo el valor de error por cada uno de ellos para conocer qué tan precisos son respecto al conjunto de datos. Para esto se tomó en consideración seleccionar el MAE y el RMSE para comparar la efectividad de los modelos.

El Error Absoluto Medio (MAE) es un buen indicador para medir la precisión de un pronóstico, siendo un número absoluto el resultado de este indicador.

$$MAE = \frac{1}{n} \sum_n |e_t| \quad (9)$$

Donde:

- MAE: Error absoluto medio
- n: número de periodos
- e_t : error

La Raíz Cuadrada Media del Error (RMSE) está definido como la raíz cuadrada de un error pronosticado al cuadrado [26].

$$RMSE = \sqrt{\frac{1}{n} \sum_n e_t^2} \quad (10)$$

5. RESULTADOS Y DISCUSIÓN

Se estimaron 8 modelos para cada empresa con periodos de prueba 3, 6 y 12 meses, tomando en cuenta los errores de cada operación con el fin de conocer cuál de los métodos se alinearon más a los mercados de las compañías.

Compañía A:

Fig. 11 Modelos ejecutados a 3 meses en la compañía A.

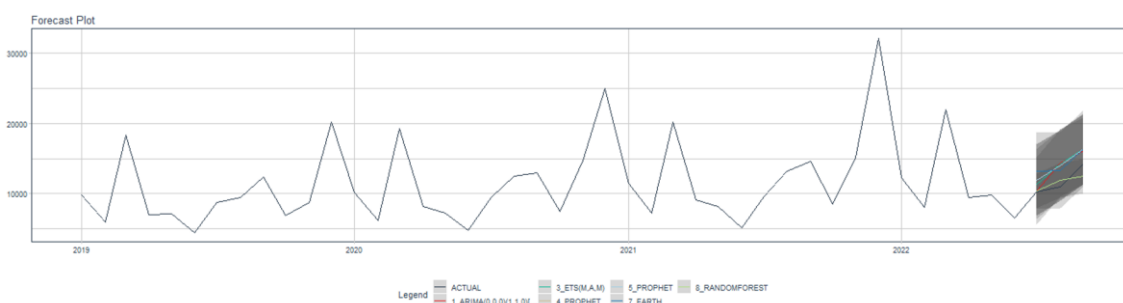


Tabla 5 Errores obtenidos en pronóstico a 3 meses en compañía A

Modelo	MAE	RMSE
ARIMA	1791.61	2216.69
XGBOOSTED ARIMA	NA	NA
ETS	2138.53	2302.02
PROPHET	2294.22	2355.77
BOOSTED PROPHET	2294.22	2355.77
MARS	2510.75	2533.71
RANDOM FOREST	929.44	1134.30

Tal como se puede ver en la tabla 5 y la figura 11, se evidenció que el modelo XGBOOSTED ARIMA no fue el más adecuado para este tipo de series dado que en la ejecución de los cálculos se dio una división para 0, retornando así un resultado como NA por lo que no se recomienda su uso para este tipo de pronóstico a corto plazo. En su contraparte, el modelo que menor error tuvo fue el de *Random Forest* donde se observó que la predicción realizada con este modelo estadístico se ajustó

a la estacionalidad que tiene la gráfica de ventas por lo que resulta adecuado para la predicción de datos en el periodo de 3 meses.

Fig. 12 Pronóstico a 6 meses en la compañía A.

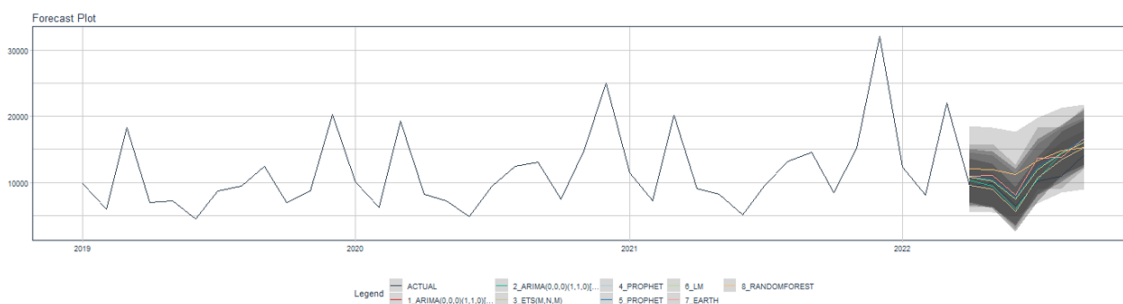


Tabla 6 Errores obtenidos en pronóstico a 6 meses en compañía A

Modelo	MAE	RMSE
ARIMA	1175.34	1624.25
XGBOOSTED ARIMA	1178.80	1629.99
ETS	989.43	1237.44
PROPHET	1866.15	2069.55
BOOSTED PROPHET	1866.15	2069.55
MARS	2156.65	2291.93
RANDOM FOREST	2912.01	3134.81

Se evidencia en la tabla 6 y la figura 12 que a los 6 meses el modelo que menos se ajustó fue el de *Random Forest* con un alto margen de error como resultado, en su contraparte el que más se ajustó fue el *ETS* dado que este asumió que la serie temporal tuvo una tendencia y estacionalidad multiplicativa además de un error aleatorio no estacional.

Fig. 13 Pronóstico a 12 meses en la compañía A.

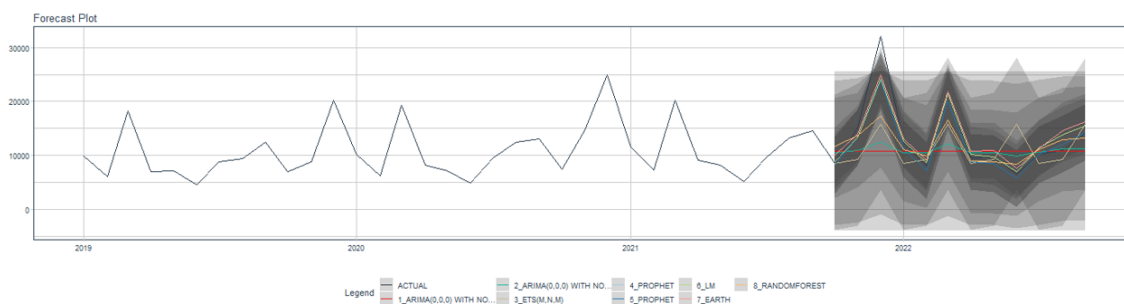


Tabla 7 Errores obtenidos en pronóstico a 12 meses en compañía A

Modelo	MAE	RMSE
ARIMA	4512.50	7345.48
XGBOOSTED ARIMA	4045.48	6692.75
ETS	4122.43	6129.33
PROPHET	1535.20	2613.69
BOOSTED PROPHET	1535.20	2613.69
MARS	1838.59	2581.24
RANDOM FOREST	2901.61	4780.62

Se evidencia en la tabla 7 y la figura 13 que a los 12 meses el error se minimizó con el modelo *Prophet* y *boosted prophet*, sin embargo, se pudo apreciar en la gráfica que el modelo *ETS* ni *Random Forest* tuvieron valores menores respecto a los errores que se encontraron en los pronósticos de 3 y 6 meses por lo que se supuso que será poco preciso pronosticar a largos periodos de tiempo con estos modelos estadísticos.

Compañía B:

Fig. 14 Pronóstico a 3 meses en la compañía B.

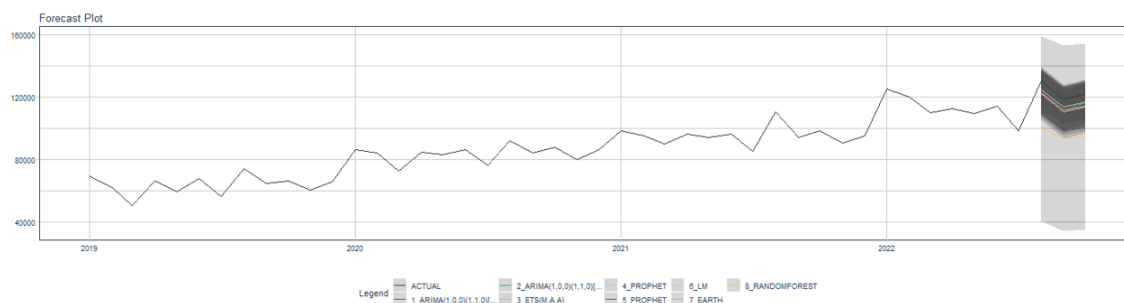


Tabla 8 Errores obtenidos en pronóstico a 3 meses en compañía B

Modelo	MAE	RMSE
ARIMA	6286.19	6332.63
XGBOOSTED ARIMA	6743.94	6787.24
ETS	7944.42	7951.34
PROPHET	6700.62	6725.66
BOOSTED PROPHET	6700.62	6725.66
MARS	7328.82	7353.95
RANDOM FOREST	27433.64	27542.67

Se evidencia en la tabla 8 y la figura 14 que el modelo que más se ajustó a la predicción de los 3 meses fue el de ARIMA con el menor valor en el error y el modelo que menos se ajustó en este caso fue el de *Random Forest* teniendo los valores más alto de medidas de error.

Fig. 15 Pronóstico a 6 meses en la compañía B.

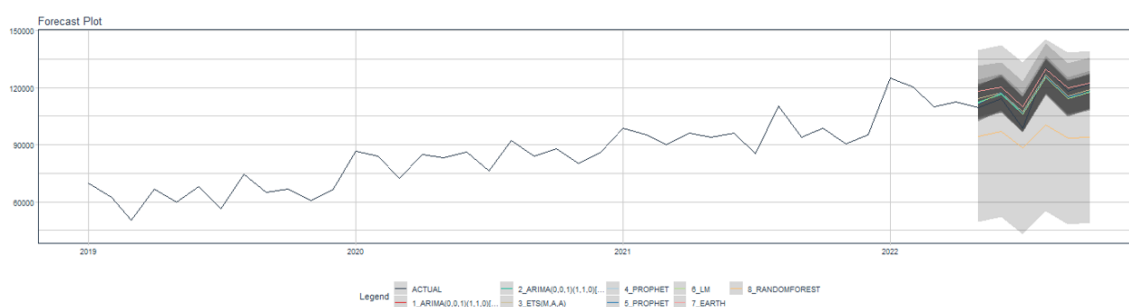


Tabla 9 Errores obtenidos en pronóstico a 6 meses en compañía B

Modelo	MAE	RMSE
ARIMA	4216.92	4530.78
XGBOOSTED ARIMA	4216.92	4530.69
ETS	4281.99	4795.17
PROPHET	4155.79	4549.84
BOOSTED PROPHET	4155.79	4549.84
MARS	4758.80	6465.15
RANDOM FOREST	20870.13	22068.91

Se evidencia en la tabla 9 y la figura 15 que a los 6 meses los modelos más adecuados fueron el *Prophet* y el *boosted Prophet* puesto que presentan menor cantidad de errores a diferencia de los demás y el menos adecuado de usarse fue el *Random Forest* siendo así el que presentó un mayor valor de errores.

Fig. 16 Pronóstico a 12 meses en la compañía B.

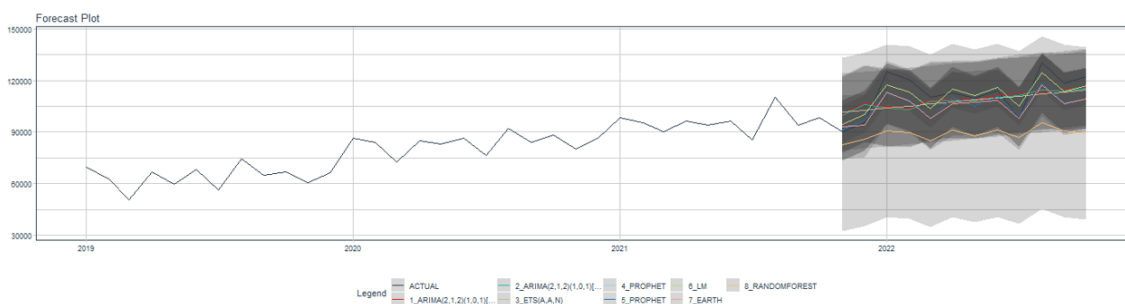


Tabla 10 Errores obtenidos en pronóstico a 12 meses en compañía B

Modelo	MAE	RMSE
ARIMA	9006.77	11119.69
XGBOOSTED ARIMA	9189.34	11062.10
ETS	9366.08	11111.85
PROPHET	7101.83	8720.80
BOOSTED PROPHET	7101.83	8720.80
MARS	7752.35	9136.25
RANDOM FOREST	23347.10	25045.42

Se evidenció en la tabla 10 y la figura 16 que el modelo con el menor valor de error fue el *Prophet* y *boosted prophet* los cuales presentaron un menor valor de error, sin embargo, los resultados generales de error fueron mayores a diferencia de cuando se pronosticó a 3 o 6 meses y se mostró que el modelo *Random Forest* fue uno de los menos adecuados para trabajarlo en este conjunto de datos.

Compañía C

Fig. 17 Pronóstico a 3 meses en la compañía C.

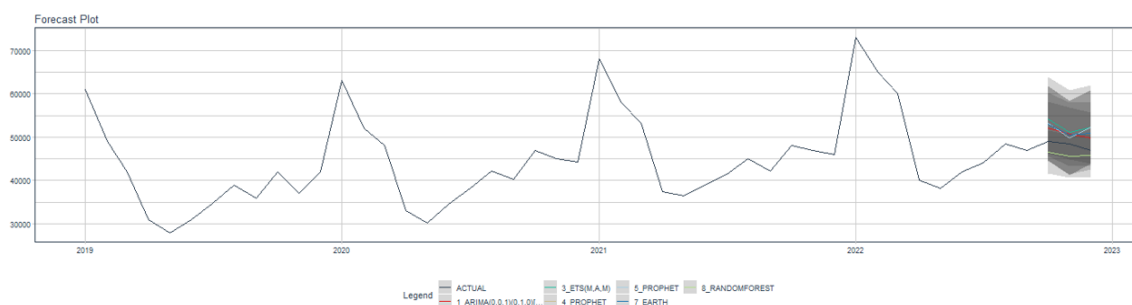


Tabla 11 Errores obtenidos en pronóstico a 3 meses en compañía C

Modelo	MAE	RMSE
ARIMA	2747.45	2767.05
XGBOOSTED ARIMA	NA	NA
ETS	4343.33	4520.54
PROPHET	3581.44	3957.88
BOOSTED PROPHET	3581.44	3957.88
MARS	3292.12	3403.56
RANDOM FOREST	2207.44	2312.38

Se evidencia en la tabla 11 y la figura 17 que el error se minimizó con el modelo *Random Forest* y se pudo observar que el modelo menos adecuado fue el *XGBoosted ARIMA*.

Fig. 18 Pronóstico a 6 meses en la compañía C.

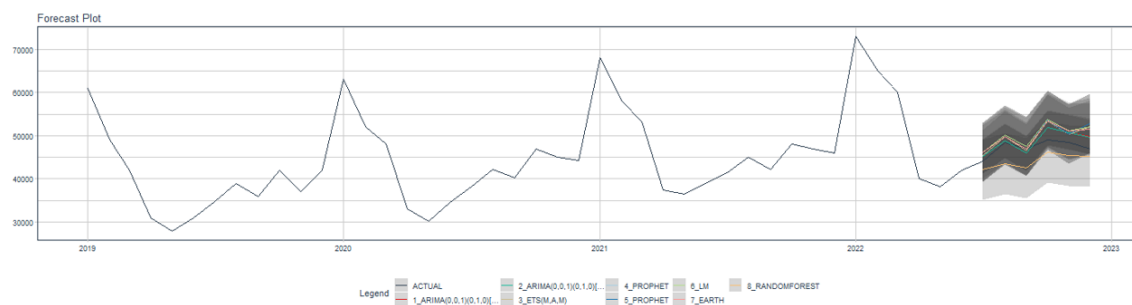


Tabla 12 Errores obtenidos en pronóstico a 6 meses en compañía C

Modelo	MAE	RMSE
ARIMA	1728.20	1982.87
XGBOOSTED ARIMA	1671.20	1911.01
ETS	2349.47	2936.62
PROPHET	2784.23	3340.61
BOOSTED PROPHET	2784.23	3340.61
MARS	2534.04	2992.46
RANDOM FOREST	3201.08	3420.68

Se evidencia en la tabla 12 y la figura 18 que el modelo más adecuado fue el *XGBoosted Arima*, también se observó que el modelo que menos se ajustó fue el de *Random Forest* el cual tuvo un valor alto de error.

Fig. 19 Pronóstico a 12 meses en la compañía C.

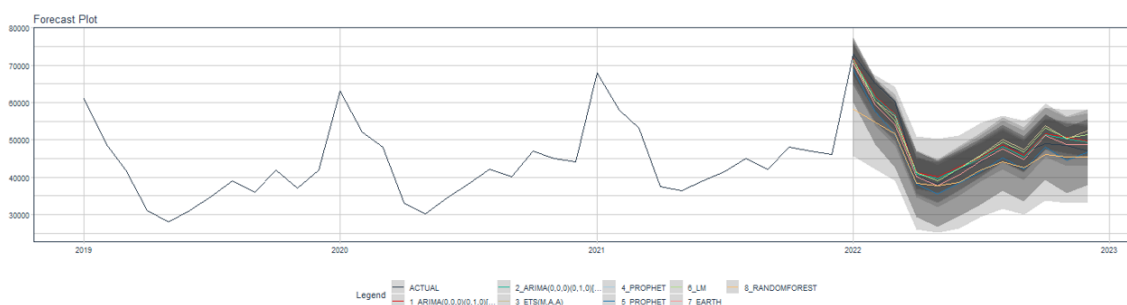


Tabla 13 Errores obtenidos en pronóstico a 12 meses en compañía C

Modelo	MAE	RMSE
ARIMA	1802.86	2053.32
XGBOOSTED ARIMA	1721.44	2028.00
ETS	2437.87	2963.79
PROPHET	3691.58	4382.99
BOOSTED PROPHET	3691.58	4382.99
MARS	1978.60	2765.17
RANDOM FOREST	4755.37	6260.91

Se evidencia en la tabla 13 y la figura 19 que el modelo más ajustado fue el *XGBoosted ARIMA*, y a pesar de un mayor tiempo de pronóstico, se pudo observar que los valores de error tienden a mantenerse ajustados al valor real, por el contrario, el modelo *Random Forest* es el que menos se ajustó al conjunto de datos de la empresa C en general dado que se evidenció un mayor valor de error en los diferentes tiempos donde se realizó el pronóstico.

En el siguiente enlace se podrá encontrar el código utilizado para la simulación en R Studio:

<https://github.com/Hacxel/tesis/blob/main/README.md?plain=1>

6. CONCLUSIONES

En este artículo, se investigó el problema de pronósticos de ventas para tres empresas de la ciudad de Guayaquil las cuales se encontraban en un entorno de mercado bastante dinámico, por lo que se formuló que el problema sería encontrar modelos más adecuados para realizar pronósticos más acertados en sus ventas futuras.

Se evidenció que, de todos los modelos evaluados para las tres compañías, no existe un modelo específico que sirviera para todos los periodos establecidos en el estudio, por lo que se recomienda siempre entrenar los modelos (seguir añadiendo datos a los modelos y ejecutar pruebas) para así poder obtener el resultado más adecuado con el fin de realizar los pronósticos.

Los resultados obtenidos sugieren que la selección del modelo de pronóstico adecuado depende del periodo de tiempo específico para el que se realizan los pronósticos. Es importante considerar las características de los datos de ventas, así como los factores externos que pueden afectar la demanda, al seleccionar un modelo de pronóstico.

Para mejorar la precisión de los pronósticos, se podrían considerar estandarizar los datos de ventas ayudando a reducir la variabilidad de los datos y mejorar la precisión de los modelos, considerar factores externos que pueden afectar la demanda incluyendo la economía, la competencia y las tendencias del mercado y también usar modelos de pronóstico más complejos los cuales pueden ser más precisos que los modelos simples, pero también pueden ser más difíciles de interpretar.

En base a estos hallazgos, se propone un enfoque futuro para el pronóstico de ventas que tenga en cuenta la separación de los datos por variables relevantes lo que permitirá identificar patrones y tendencias que no serían evidentes si los datos se analizaran en su conjunto, además del uso de modelos de pronóstico más

complejos para capturar la complejidad de la relación entre las variables y mejorar la precisión de los pronósticos.

La separación de los datos por variables relevantes puede realizarse por geografía, sector de mercado, tipo de producto, tipo de cliente, etc. Por ejemplo, las ventas de un producto pueden ser diferentes en diferentes regiones geográficas, por lo que es importante analizar las ventas por regiones para comprender mejor la demanda.

El uso de modelos de pronóstico más complejos puede ayudar a capturar la complejidad de la relación entre las variables. Por ejemplo, un modelo de regresión logística puede ayudar a predecir las ventas de un producto en función de la ubicación geográfica, el sector de mercado, el tipo de producto y el tipo de cliente.

Este enfoque futuro para el pronóstico de ventas tiene el potencial de mejorar la precisión de los pronósticos al considerar las características de los datos y los factores externos que pueden afectar la demanda.

Para el caso de las compañías escogidas para el estudio, las distribuidoras tienen más estacionalidad en sus ventas, por lo que los modelos como *ARIMA*, *XG Boosted ARIMA*, *ETS* o *Prophet* son los más adecuados para realizar pronósticos a largo plazo puesto que presentaron los mejores resultados entre los modelos simulados

En comparación con otros estudios se observó que para el caso de empresas distribuidoras el modelo *Random Forest* no es el más adecuado para realizar los pronósticos a largo plazo debido a que no toma en consideración la estacionalidad de las ventas.

En conclusión, los modelos estadísticos propuestos mejoraron el rendimiento puesto que los valores de error son menores a los originalmente usados por las compañías por lo que se acepta la hipótesis planteada.

Los modelos estadísticos propuestos pueden ser usados para los pronósticos a mediano plazo, sin embargo, se debe seguir entrenándolos con los datos nuevos

para corroborar que siguen siendo aptos para los mercados y que muy probablemente necesiten cambiarse.

A medida que se aumentaba el tiempo, el número de errores crecía por lo que no es recomendable realizar pronósticos a largo plazo dentro de estas compañías de estos sectores de mercado, esto se da por el dinamismo del mercado el cual se conoce que es bastante cambiante respecto a la necesidad de los productos.

De todos los modelos, el que tuvo mayor la cantidad de errores era el Random Forest por lo que no se recomienda el uso del mismo para estos casos.

Los modelos propuestos fueron útiles para las estimaciones de pronósticos en estas empresas ayudando en la eficiencia de los procesos.

REFERENCIAS

- [1] G. A. Rob J Hyndman, *Forecasting: principles and practice*, Monash - Australia: Otexts, 2018.
- [2] W. B. G. D. Levy M., *Retailing Management*, McGraw-Hill Education, 2019.
- [3] E. A. B. D. G. Verstraete, «A leading macroeconomic indicators based framework to automatically generate tactical sales forecasts,» *Computers & Industrial Engineering*, vol. 139, nº 139, 2019.
- [4] Y. Y. Chih-Hsuan Wang, «Demand planning and sales forecasting for motherboard manufacturers considering dynamic interactions of computer products,» *Computers & Industrial Engineering*, vol. 1, nº 149, 2020.
- [5] M. M. B. M. G. P. S. P. R. Majhi, «Forecasting of Retail Sales Data Using Differential Evolution,» *IEEE*, vol. 1, 2009.
- [6] F. Ş. İlker Güven, «Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods,» *Computers & Industrial Engineering*, vol. 1, nº 149, 2020.
- [7] R. D. J. H. J. Holton Wilson, «Women count: Perceptions of forecasting in sales,» *Business Horizons*, vol. 1, nº 63, pp. 637-646, 2020.
- [8] X. L. Qingwen Sun, «The Study of Cluster Predication Method on Sales Forecast Based on Residual Error Modified GM (1, 1),» de *International Seminar on Business and Information Management*, Shijiazhuang, 2008.
- [9] R. Jacobs, *Operations and Supply Management - The Core*, McGraw Hill, 2008.
- [10] S. M. B. M. G. P. M. R. Ritanjali Majhi, «Efficient Sales Forecasting Using PSO Based Adaptive ARMA Model,» de *World Congress on Nature & Biologically Inspired Computing*, Bhubaneswar, 2009.
- [11] Z. S. Xie dairu, «Machine Learning Model for Sales Forecasting by Using XGBoost,» de *IEEE International Conference on Consumer Electronics and Computer Engineering*, Chongqing, 2021.
- [12] Z. Qiao, «Walmart Sale Forecasting Model Based On LightGBM,» de *2nd International Conference on Machine Learning, Big Data and Business Intelligence*, Edinburgh, 2020.
- [13] W. K. S. X. Z. L. Junhang Chen, «Sales Forecasting Using Deep Neural Network And SHAP techniques,» de *IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering*, Wuhan, 2021.
- [14] S. M. A. K. V. A. Evangelos Spiliotis, «Product sales probabilistic forecasting: An empirical evaluation using the M5 competition data,» *Int. J. Production Economics*, vol. 1, nº 240, 2021.
- [15] S.-A. S. Julie Marcoux, «A Hybrid Subspace-Connectionist Data Mining Approach for Sales Forecasting in the Video Game Industry,» de *World Congress on Computer Science and Information Engineering*, Canadá, 2009.

- [16] W. Y. W. F. L. O. Mu Shengdong, «Transformative computing for products sales forecast based on SCIM,» *Applied Soft Computing*, vol. 1, nº 109, 2021.
- [17] P. H. M. W. Lindsay R. Berry, «Probabilistic forecasting of heterogeneous consumer transaction–sales time series,» *International Journal of Forecasting*, vol. 1, nº 36, 2020.
- [18] R. F. Shaohui Ma, «Retail sales forecasting with meta-learning,» *European Journal of Operational Research*, vol. 1, nº 288, pp. 111-128, 2021.
- [19] P. O. R. T. A. N. Vahid Sohrabpour, «Export sales forecasting using artificial intelligence,» *Technological Forecasting & Social Change*, vol. 1, nº 163, 2021.
- [20] Y.-X. T. Z.-P. F. Chuan Zhang, «Forecasting sales using online review and search engine data: A method based on PCA–DSFOA–BPNN,» *International Journal of Forecasting*, vol. 38, nº 3, pp. 1005-1024, 2022.
- [21] R. Hyndman, «Automatic Time Series Forecasting,» *Journal of Statistical Software*, vol. 27, nº 3, pp. 1-22, 2008.
- [22] M. A.-A. Z. I. A. Goswamy, «Factors affecting injury severity at pedestrian crossing locations with rectangular rapid flashing beacons (RRFB) using XGBoost and random parameters discrete outcome models,» *Accident Analysis & Prevention*, vol. 181, 2023.
- [23] E. Carbonero, Análisis de tráfico de internet mediante el uso del suavizado exponencial de series temporales, Madrid: UAM, 2021.
- [24] Y.-I. B. Q.-h. Y. L. D. Y.-j. M. Yu-ting Huang, «Application of a hybrid model based on the Prophet model, ICEEMDAN and multi-model optimization error correction in metal price prediction,» *Resources Policy*, vol. 79, 2022.
- [25] A. N. T. V. G. Rachit Srivastava, «Solar radiation forecasting using MARS, CART, M5, and random forest model: A case study for India,» *Heliyon*, vol. 5, 2019.
- [26] N. Vandeput, Data Science for Supply Chain Forecasting 2nd edition, Lituania: CPI books GmbH, Leck, 2021.