



Aprendizaje supervisado en el pronóstico de la demanda de artículos en diferentes minimarkets

Supervised learning in forecasting demand for items in different minimarkets

Aurora Rosa Neciosup Obando^{1*}, Isidoro Valderrama Ramos², Edinson Huaylla Salazar³

¹Departamento de Estadística, Facultad de Ciencias Físicas de Matemáticas, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n, Ciudad Universitaria, Trujillo, Perú.

²Departamento de Ingeniería Química, Facultad de Ingeniería Química, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n, Ciudad Universitaria, Trujillo, Perú.

³Escuela Académico Profesional de Estadística, Facultad de Ciencias Físicas de Matemáticas, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n, Ciudad Universitaria, Trujillo, Perú

Resumen

Esta investigación no experimental tiene como propósito pronosticar la demanda para una gestión eficaz del inventario en los almacenes de productos en supermercados, con el fin de evitar la escasez o el sobreabastecimiento que puede tener repercusiones económicas o afectar la imagen de la empresa. Para ello se utilizaron datos históricos, los mismos que fueron procesados en Python, previamente particionados en grupos de entrenamiento y prueba. Se aplicó para el análisis el modelo Holt-Winters calibrado a fuerza bruta, para predecir el tamaño de un pedido urgente, que reduce el riesgo de falta de existencias en el almacén local en un 93.4%. Luego al comparar este modelo con una Red Neuronal (LSTM), resultó que, Holt-Winter obtuvo mejor rendimiento con un error de estimación de 0.0095. Concluyendo que el modelo de Holt-Winters mejora la precisión en los pronósticos de la demanda y puede ser también de gran utilidad en este sector comercial que resultaría muy beneficiado.

Palabra clave: Aprendizaje supervisado, Holt-Winters, inventario, supermercados, Python

Abstract

The purpose of this non-experimental research is to forecast the demand for effective inventory management in supermarket product warehouses, in order to avoid shortages or oversupply that can have economic repercussions or affect the company's image. For this, historical data was used, which was processed in Python, previously partitioned into training and test groups. The Holt-Winters model calibrated to brute force was applied for the analysis to predict the size of an urgent order, which reduces the risk of out-of-stock in the local warehouse by 93.4%. Then, when comparing this model with a Neural Network (LSTM), it turned out that Holt-Winter obtained better performance with an estimation error of 0.0095. Concluding that the Holt-Winters model improves the accuracy of demand forecasts and can also be very useful in this commercial sector that would benefit greatly.

Keyword: Supervised learning, Holt-Winters, inventory, supermarkets, Python

1. Introducción

Al llegar la enfermedad del coronavirus (Covid-19) al Perú, inicialmente produjo en el comprador incertidumbre y pánico por hacer compras masivas (Li et al., 2021; Pang et al., 2021; Di Crosta et al., 2021). Esta realidad que comenzó en los mercados y supermercados al adquirir alimentos no perecibles, productos de primera necesidad e higiene, pensando en tener reservas personales o familiares, en la creencia que este virus pudiese estar presente poco tiempo en su comunidad; pero conforme pasan los meses, el consumidor convencional, entiende que todos debemos saber convivir con esta pandemia (Anastasiadou et al., 2020; Goswami et al., 2021; Wang et al., 2020).

Por otra parte, lo que busca mayormente el consumidor convencional en tiempos de pandemia, son los alimentos procesados frescos, pero envasados; algunos, buscan harina de trigo para preparar pan o suplementos similares, para ya no comprar el pan fabricado tradicionalmente en la panadería del barrio, por temor a la manipulación de cómo lo elaboran y cómo se vende el mismo producto. Los embutidos, por ejemplo, como salchicha, hot-dog, jamonada y otros, son escogidos; los derivados lácteos como queso y leche en su presentación envasada, son los nuevos productos preferidos, como también las bebidas envasadas son las de mayor demanda en la compra de productos.

Este comportamiento antes expuesto se evidencia también en el consumidor convencional peruano y por ende en el departamento de la Libertad, que ha dado un giro al adquirir sus productos de primera necesidad. Actualmente, muestra mayor interés en productos enlatado y envasados, leyendo rótulos y etiquetas para saber su contenido y fecha de vencimiento de los productos, poniendo especial atención a los octógonos, que poco antes de la pandemia se normaron por ley peruana obligatoria, que adviertan al consumidor de los alimentos con alto contenido de grasas, azúcares y sodio. De igual forma se observa también una disminución en el consumo del cebiche y busca otras alternativas como carne para cocinar y freír o enlatadas.

Los cambios ocurridos en el comportamiento del consumidor post pandemia en el Perú, provocan una incertidumbre en los supermercados, por lo que es importante realizar investigaciones como esta, donde se estudie y modele el comportamiento de los clientes para poder realizar pronósticos que permitan llevar a cabo una mejor gestión del inventario, lo que podría evitar la escasez de productos con mayor demanda o exceso de productos con poca demanda, lo que podría repercutir en la insatisfacción de los clientes en desmedro de su fidelidad; así como ocasionar costos de almacenamiento innecesarios, pérdidas y desperdicios a causa de productos vencidos, afectando los ingresos de la empresa.

Lo expuesto anteriormente respecto a la inquietud relacionada con la variabilidad de la demanda de artículos en los supermercados motivó la realización de este estudio en los supermercados del departamento de La Libertad con el objetivo de desarrollar un modelo de aprendizaje supervisado basado en Holt-Winters, calibrado con métodos de fuerza bruta, para pronosticar la demanda de artículos en diferentes supermercados.

Trabajos previos

En esta sección se presentan investigaciones realizadas acerca del estudio de la demanda, de productos o servicios, en particular aquellas que hayan empleado modelos de predicción, como es el caso del modelo de serie de tiempo de Holt-Winters.

López (2021) realiza la comparación entre la metodología de Box-Jenkins y el método de Holt-Winters con el fin de determinar la capacidad predictiva, llevándose a cabo este estudio a partir de una serie de tiempo estacional y con tendencia, procedente de la serie de tráfico de pasajeros a nivel nacional vía aérea en Perú en el periodo enero 2005-diciembre 2019. Se da a conocer mediante este trabajo la teoría de ambas técnicas utilizadas y los fundamentos básicos de las series de tiempo como métodos de predicción. Finalmente, con ayuda de los errores de pronóstico MSE, RMSE, MAD y MAPE se logró encontrar que la mejor metodología que predice a la serie de tiempo tráfico de pasajeros a nivel nacional vía aérea es el modelo de Box-Jenkins.

Pérez (2012) evaluó el comportamiento histórico de la demanda de los productos de consumo masivo en una cadena de suministros con una bodega y múltiples puntos de venta partiendo del análisis de la clasificación existente de los ítems para proponer una nueva con mayor cantidad de elementos de decisión; con base en lo anterior y mediante simulación, se eligió el método de pronóstico (promedio móvil, suavización exponencial simple, método de Croston, método de Winters) más apropiado para cada elemento y para cada una de las categorías considerando como factor de decisión el que obtuvo el menor coeficiente de variación; concluyendo que, debido a las particularidades en la rotación y al comportamiento heterogéneo de la demanda en cada sitio de distribución, es conveniente realizar una clasificación por cada punto y definir modelos de pronósticos de manera individual.

Ricce et al. (2021) propone un modelo básico de red neuronal artificial (RNA) como base para pronosticar la demanda de pedidos de urgencia y mejorar el pronóstico. Construyó un modelo RNA para pronosticar la ocurrencia y volumen de un pedido de urgencia. Este pronóstico permite solicitar, a tiempo, la cantidad suficiente de producto al centro de distribución en Norteamérica, con la finalidad de reducir el riesgo de rotura de stock en el almacén local en un 91.1%. El modelo fue validado a través de simulaciones, utilizando el método de transformada inversa y los beneficios del modelo fueron confirmados a través de un flujo de caja al comparar los escenarios con y sin el uso del modelo, obteniéndose una tasa de retorno de la inversión superior al 20%.

Jara (2021) desarrolló un modelo de estimación para la gestión de repuestos mediante herramientas de análisis de datos para mejorar los tiempos de respuesta ante incidencias. Este obtuvo el menor porcentaje de error para los repuestos con más nivel de rotación dentro de la bodega, determinado que Holt-Winters es más preciso.

Mor et al. (2021) pronosticaron la demanda para productos lácteos de vida cota, usando los modelos Holt-Winters, promedio móvil y regresión múltiple, con la finalidad de formular estrategias adecuadas en el proceso productivo, concluyendo que la elección de los modelos está influenciada por el conocimiento, la accesibilidad a los datos y el entorno donde se hace el pronóstico. Asimismo, Molina (2020) identificó al modelo de Holt-Winters como un buen modelo de ajuste para el pronóstico de la demanda de productos.

Acosta et al. (2013) utilizan redes neuronales artificiales para predecir el desempeño de varios métodos estadísticos tradicionales de pronóstico para así seleccionar el de mejor potencial. Para llevar a cabo la validación, se emplearon dieciocho series de tiempo reales, correspondientes a actividades económicas del estado de Tamaulipas. Resultando que el método de selección propuesto es suficientemente confiable para devenir un recurso de fácil aplicación para personas con poco conocimiento estadístico.

2. Metodología

La investigación con diseño no experimental (Hernández & Mendoza, 2018), desarrollada a partir de datos secundarios, de accesibilidad abierta, publicados en la página de la Cámara de Comercio de la Libertad.

La muestra de estudio estuvo conformada por la serie diaria histórica de la demanda de artículos de diferentes minimarkets del departamento de La Libertad en el periodo 2013-2017.

Para el análisis de los datos se aplicaron técnicas de modelación de series de tiempo utilizando el modelo de Redes Neuronales y el modelo Holt-Winters calibrado, utilizando datos históricos de la demanda. Para el caso de Holt-Winters se trató de calibrar mediante programación, con la técnica de fuerza bruta para conseguir un mejor ajuste del modelo. Posteriormente se encontraron cada uno de los modelos a comparar utilizando el software Python.

En lo que respecta al proceso ético seguido en la investigación, se cumplió desde el inicio puesto que el proyecto del estudio fue aprobado por las comisiones de investigación y ética de la Universidad Nacional de Trujillo. En cuanto al manejo de los datos fue realizado con transparencia y honestidad manteniendo el anonimato de los supermercados incluidos en la investigación.

3. Resultados

3.1 Paquetes importados

```
In [87]: 1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from datetime import date
5 from datetime import datetime
6 %matplotlib inline
7 import datetime
8 import scipy.stats
9 import statsmodels.graphics.tsaplots as sgt
10 import statsmodels.tsa.stattools as sts
11 from statsmodels.tsa.seasonal import seasonal_decompose
12 import seaborn as sns
13 sns.set()
```

```
In [75]: 1 demanda = pd.read_csv("demanda.csv", delimiter = ';', decimal = ",", header = 0)
```

```
In [76]: 1 fechas = pd.date_range(start = "1982-01-01", periods = 476, freq = 'MS')
2 ts = pd.Series(demanda.demanda.values, index = fechas)
3 ts
```

```
Out[76]: 1982-01-01    93.2
1982-02-01    96.0
1982-03-01    95.2
1982-04-01    77.1
1982-05-01    70.9
...
2021-04-01    127.0
2021-05-01    151.0
2021-06-01    130.0
2021-07-01    119.0
2021-08-01    153.0
Freq: MS, Length: 476, dtype: float64
```

```
In [77]: 1 demanda.isna().sum()
```

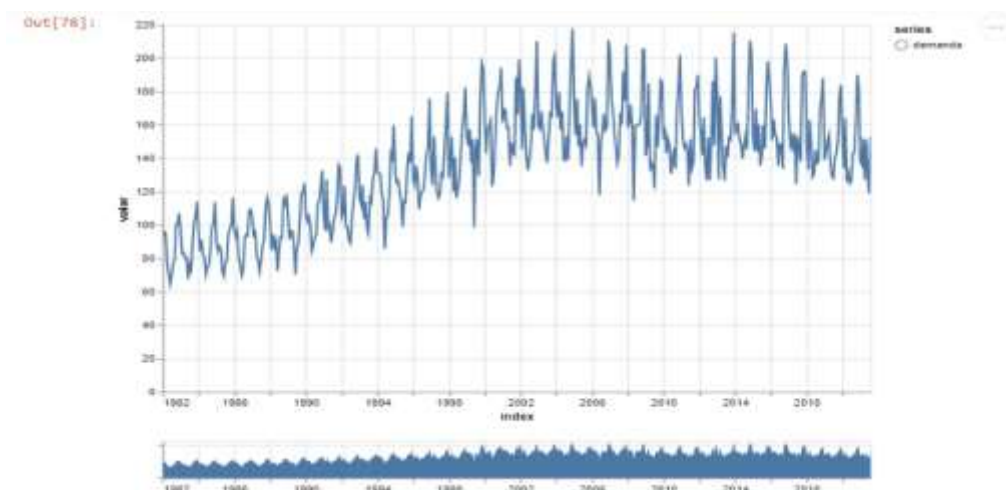
```
Out[77]: demanda    0
dtype: int64
```

3.2 Gráfico de la demanda usando la biblioteca “altair”

```
In [78]: 1 ts_df = ts.to_frame(name = "demanda").reset_index()
2
3 ts_plot = plot_ts(ts_df, True, True)
4 ts_plot
```

3.3 Representación gráfica de la serie temporal

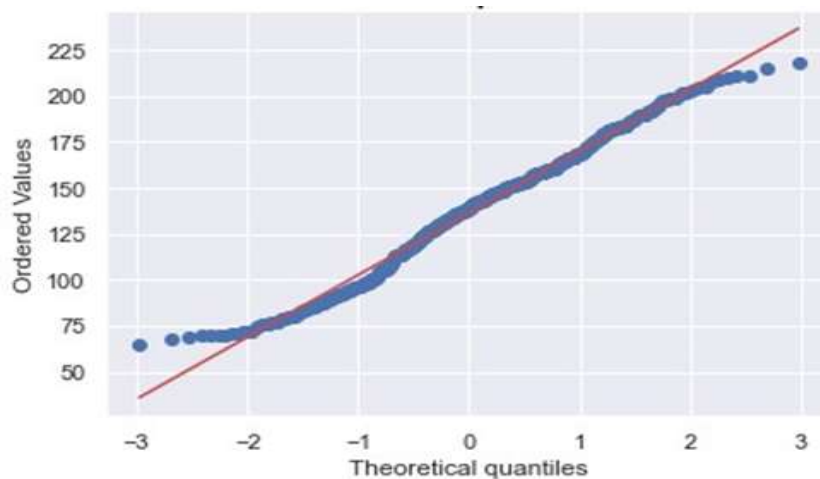
Figura 1. Serie temporal de la demanda de artículos en los minimarkets



3.4 Prueba de normalidad

```
In [89]: 1 scipy.stats.probplot(demanda.demanda, plot=plt)
2 plt.title("QQ plot", size=24)
3 plt.show()
```

Figura 2. QQ-plot para visualizar la normalidad de los residuos



3.5 Prueba de Dickey-Fuller para evaluar la estacionariedad de la serie

```
In [52]: adf=adfuller(demanda['demanda'], maxlag=1)
print('El T-Test es:', adf[0])

El T-Test es: -5.471435034738571

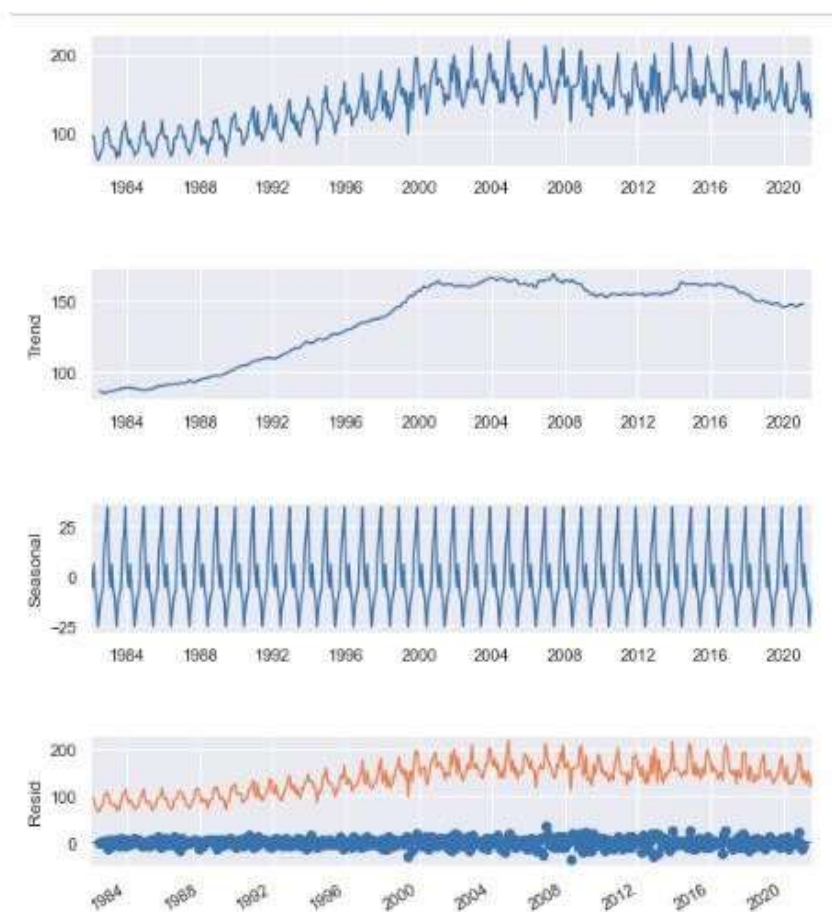
In [53]: print('El P-Value es:', adf[1])
print('Valores críticos:', adf[4])

El P-Value es: 2.393774859655648e-06
Valores críticos: {'1%': -3.4442214494657586, '5%': -2.8676569002144148, '10%': -2.5700280720682227}
```

La serie con respecto a la demanda de productos en minimarkets es estacionaria, por ser $p < .01$, hay evidencia para rechazar la hipótesis nula que la serie tiene raíz unitaria, por lo que podría considerarse es estacionaria.

3.6 Descomposición de la serie

Figura 3. Descomposición de la serie temporal de la demanda en tendencia, estacionalidad y residuos



3.7 Modelo de Holt-Winters calibrado

```

class HW_Prediccion(Prediccion):
    def __init__(self, modelo, alpha, beta, gamma):
        super().__init__(
            (modelo)self.__
            alpha = alpha self.
            beta = beta self.__
            gamma = gamma

    @property
    def alpha(self):
        return self.__alpha

    @property
    def beta(self):
        return self.__beta

    @property
    def gamma(self):
        return self.__gamma

    def forecast(self, steps = 1):
        res = self.modelo.forecast(steps)
        return(res)

class HW_calibrado(Modelo):
    def __init__(self, ts, test, trend = 'add', seasonal = 'add'):
        super().__init__(ts)
        self.__test = test
        self.__modelo = ExponentialSmoothing(ts, trend = trend, seasonal = seasonal)

    @property
    def test(self):
        return self.__test

    @test.setter
    def test(self, test):
        if(isinstance(test, pd.core.series.Series)):
            if(test.index.freqstr !=
                None):self.__test = test
            else:
                warnings.warn('ERROR: No se indica la frecuencia de la serie de tiempo.')
        else:
            warnings.warn('ERROR: El parámetro ts no es una instancia de serie de tiempo
                .')

    def fit(self, paso =
        0.1):error =
        float("inf") alpha =
        paso
        while alpha <= 1:beta =

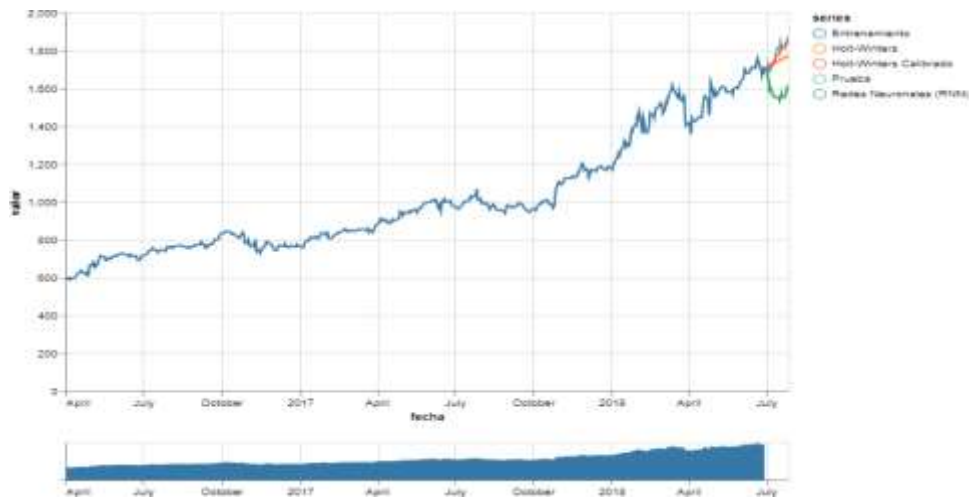
```

```

0
while beta
  <= 1:
    gamma
    = 0
    while gamma <= 1:
      model_fit = self.modelo.fit(smoothing_level = alpha, smoothing_slope =
beta, smoothing_seasonal = gamma)
      pred      = model_fit.forecast(len(self.test))
      mse = sum((pred - self.test)**2)
      if mse < error:
        res_alpha = alpha
        res_beta = beta
        res_gamma = gamma
        res = model_fit
        gamma += paso
      beta += paso
      alpha += paso
return(HW_Prediccion(res, res_alpha, res_beta, res_gamma))
modelo_calibrado = HW_calibrado(AirPassengers_train, AirPassengers_test)
modelo_calibrado_fit = modelo_calibrado.fit(0.05)

modelo_calibrado_fit.alpha
    
```

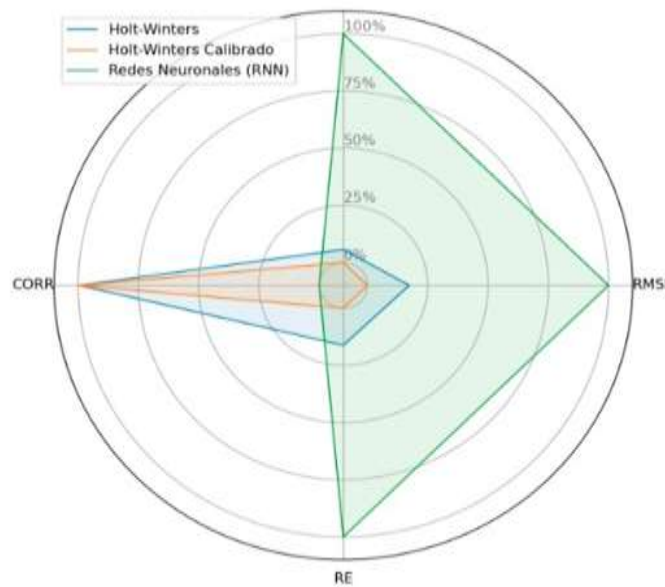
Figura 4. Modelado y predicción en base a datos de entrenamiento y prueba



3.8 Medición del error

	MSE	RMSE	RE	CORR
Holt-Winters	3026.631311	55.014828	0.025871	0.927642
Holt-Winters Calibrado	467.017457	21.610587	0.009481	0.934492
Redes Neuronales (RNN)	45816.024096	214.046780	0.110907	-0.507623

Figura 5. Ploteo de los errores



4. Discusión

Esta investigación tuvo como propósito determinar que el modelo Holt-Winters calibrado es mejor que el modelo de Redes Neuronales para el pronóstico de la demanda de artículos endiferentes minimarkets, de La Libertad desde 2013 al 2017, para encontrar los mejores parámetros en Holt-Winters, se utilizó la fuerza bruta, creándose la función respectiva.

Los resultados obtenidos establecen que el modelo Holt-Winters calibrado es el que representa el mejor ajuste para el pronóstico de la demanda de artículos en los minimarkets incluidos en la investigación. Estos resultados concuerdan con Jara (2021) que, en su investigación, sobre la mejora de la gestión de la demanda de repuestos en estaciones de telefonía móvil encontró que el modelo de Holt-Winters obtuvo un mejor rendimiento con respecto a los modelos ARIMA y regresión lineal.

Por otra parte, los resultados del presente estudio son coherentes con lo afirmado por Fabson et al. (2013) que confirman la efectividad del modelo Holt-Winters tanto en los pronósticos, como en el control del efecto látigo que se define como el incremento de las variaciones en la demanda en una cadena de suministro.

En concordancia con la importancia del pronóstico de la demanda en la planificación del inventario y las estrategias de mercado, se aplicaron tres modelos de pronóstico en los referidos minimarkets; según lo sugerido por investigadores como Wanchoo et al. (2019) y Acosta et al. (2013) se pueden aplicar diferentes técnicas de pronóstico con buenos resultados como, ARIMA, regresión lineal, Holt, Holt-Winters, redes neuronales, suavizado exponencial, métodos de aumento de gradiente entre otros para reducir la incertidumbre en la planificación de abastecimiento de productos.

Finalmente se sugiere ampliar el uso de estas técnicas para el pronóstico de la demanda en supermercados de La Libertad, utilizando específicamente el modelo de Holt-Winters calibrado con fuerza bruta; recomendando su aplicación en diferentes contextos y la evaluación de su rendimiento, así como el empleo de nuevas metodologías que talvez puedan brindar mejores modelos de pronóstico siempre con el objetivo de lograr una mejor planificación de inventario en beneficio tanto del cliente como del empresario.

5. Conclusiones

Como conclusión de esta investigación se logró desarrollar y establecer al modelo de aprendizaje supervisado de Holt-Winters calibrado con fuerza bruta, como el mejor modelo, que permite pronosticar con precisión la demanda de artículos en diferentes supermercados de la Libertad, al reportar un nivel de error de 0.009, muy inferior a los errores de estimación de los modelos de Holt-Winters y de redes neuronales que obtuvieron valores de 0.026, y 0.111 respectivamente. Estos resultados podrían servir de base para una mejor gestión del inventario en los supermercados, evitando la escasez y/o el sobreabastecimiento de productos repercutiendo en beneficio de la satisfacción de los clientes y de la economía de las estas empresas.

6. Referencias

- Acosta M., Villareal, M., & Cabrera, M. (2013). Estudio de validación de un método para seleccionar técnicas de pronóstico de series de tiempo mediante redes neuronales artificiales. *Ingeniería. Investigación y Tecnología*, 14(1), 53-63.
<https://www.sciencedirect.com/science/article/pii/S140577431372225X>
- Anastasiadou, E., Anestis, M. C., Karantza, I., & Vlachakis, S. (2020). The coronavirus' effects on consumer behavior and supermarket activities: insights from Greece and Sweden. *International Journal of Sociology and Social Policy*, 40(9/10), 893-907.
- Di Crosta, A., Ceccato, I., Marchetti, D., La Malva, P., Maiella, R., Cannito, L., ... & Di Domenico, A. (2021). Psychological factors and consumer behavior during the COVID-19 pandemic. *PloS one*, 16(8), e0256095.
- Fabson, T. V. O., & Oyatoye, E. O. (2013). An assessment of the Holt-Winters model in making effective forecast for supply chain system. *International Journal of Business and Systems Research*, 7(2), 208-227.
- Goswami, S., & Chouhan, V. (2021). Impact of change in consumer behaviour and need prioritisation on retail industry in Rajasthan during COVID-19 pandemic. *Materials Today: Proceedings*, 46, 10262-10267.
- Hernández, S. & Mendoza, P. (2018). *Metodología de la Investigación*. Mc Graw Hill.
http://www.biblioteca.cij.gob.mx/Archivos/Materiales_de_consulta/Drogas_de_Abuso/Articulos/SampieriLasRutas.pdf
- Rivero José (2007) "Educación, Docencia y Clase Política en el Perú" 1ra Edición Editorial Tarea, Ayuda en Acción. Lima-Perú.

- Jara, J. (2021). *Desarrollo de un modelo de estimación de inventario tecnológico mediante analítica de datos para mejorar tiempos de respuesta ante incidentes*. (Tesis de Maestría, Universidad de las Fuerzas Armadas).
<http://repositorio.espe.edu.ec/bitstream/21000/27658/1/T-ESPE-018152.pdf>
- Li, X., Zhou, Y., Wong, Y. D., Wang, X., & Yuen, K. F. (2021). What influences panic buying behaviour? A model based on dual-system theory and stimulus-organism-response framework. *International Journal of Disaster Risk Reduction*, 64, 102484.
- López, F. (2021) *Comparación de la metodología de Box-Jenkins y el método de Holt-Winters para el pronóstico de series de tiempo univariado*. Pontificia Universidad Católica de Chile. <https://renati.sunedu.gob.pe/handle/sunedu/3157540>
- Molina, K. (2020). *Implementación de un modelo analítico para la predicción de la venta del portafolio de productos OTC de un Laboratorio Farmacéutico*. (Tesis de Maestría, Universidad de las Fuerzas Armadas).
<http://repositorio.espe.edu.ec/bitstream/21000/22561/1/T-ESPE-043875.pdf>
- Mor, R. S., Jaiswal, S. K., Singh, S., & Bhardwaj, A. (2019). Demand forecasting of the short-lifecycle dairy products. *Understanding the Role of Business Analytics: Some Applications*, 87-117.
- Pang, Q., Meng, H., Fang, M., Xing, J., & Yao, J. (2021). Social distancing, health concerns, and digitally empowered consumption behavior under COVID-19: a study on livestream shopping technology. *Frontiers in Public Health*, 9, 748048.
- Pérez, R.; Mosquera, S. & Bravo J. (2012). Aplicación de modelos de pronósticos en productos de consumo masivo. *Rev. Bio. Agro*, 10(2), Popayán July/Dec.
http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-35612012000200014
- Rice-Medina et al. (2021). *Pronóstico de la demanda altamente variable e intermitente usando un modelo básico de Red Neuronal Artificial para disminuir el riesgo de rotura de stock de una compañía que abastece productos en Sudamérica*. (Tesis de Pregrado, Universidad Peruana de Ciencias Aplicadas).
<https://repositorioacademico.upc.edu.pe/handle/10757/657936>
- Wanchoo, K. (2019). Retail demand forecasting: a comparison between deep neural network and gradient boosting method for univariate time series. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, 1-5. IEEE.
- Wang, E., An, N., Gao, Z., Kiprop, E., & Geng, X. (2020). Consumer food stockpiling behavior and willingness to pay for food reserves in COVID-19. *Food Security*, 12, 739-747.