

# EVALUACIÓN DE ALGORITMOS DE APRENDIZAJE AUTOMATIZADO PARA LA PROYECCIÓN DE LA DEMANDA DE CLIENTES DE UNA INSTALACIÓN HOTELERA

**Mayly Torres Álvarez<sup>1</sup>, María Estela Peralta Álvarez<sup>2</sup>, Julio Ernesto de la Rosa Melian<sup>1</sup>, Any Flor Nieves Julbe<sup>1</sup>**

<sup>1</sup>Universidad de Holguín; Cuba

<sup>2</sup>Universidad de Sevilla; España

E-mail de correspondencia: [mayly.torres@uho.edu.cu](mailto:mayly.torres@uho.edu.cu)

## RESUMEN

El turismo para diversos países constituye un sector priorizado por los beneficios que reporta a la economía de los mismos. La correcta gestión de los destinos turísticos depende en gran medida de la proyección que se realice de la demanda de clientes. Con la introducción de la tecnología 4.0 la industria hotelera ha logrado mayor eficiencia en sus procesos. La implementación de herramientas de inteligencia artificial contribuye a disminuir la incertidumbre asociada a los análisis predictivos. El objetivo general de la investigación es evaluar algoritmos de aprendizaje automatizado para el pronóstico del por ciento de ocupación del hotel Brisas Guardalavaca de Holguín.

## 1. INTRODUCTION

La introducción de las tecnologías 4.0 ha conllevado a realizar transformaciones a los procesos organizacionales. La inteligencia artificial es intrusiva a todos los sectores productivos y de servicios donde el turismo no es una excepción, al enfrentarse también a un futuro más automatizado. La implementación de inteligencia artificial es fundamental para la automatización del marketing y los procesos de servicios en los hoteles (Tussyadiah, 2020). Para mejorar la precisión de la predicción de la demanda turística también se han empleado técnicas de aprendizaje automatizado.

Al efectuar un análisis del tratamiento metodológico dado a los pronósticos (Escalona Morales, 2011; Gunter y Önder, 2015; Huarng y Yu, 2019; Li, Pan, Law y Huang, 2017; Önder y Gunter, 2016; Ramírez, 2013; Salazar Aguilar y Cabrera Ríos, 2007; Tsui, Balli, Gilbey y Gow, 2014; Yang, Pan y Song, 2014; Zhu, Lim, Xie y Wu, 2018) se identificó que los métodos que clasifican dentro de las series de tiempo son los más utilizados en el contexto empresarial, mientras que los métodos de inteligencia artificial aún se están desarrollando. El desarrollo de modelos basados en inteligencia artificial es notable en los estudios de previsión de la demanda turística más recién-

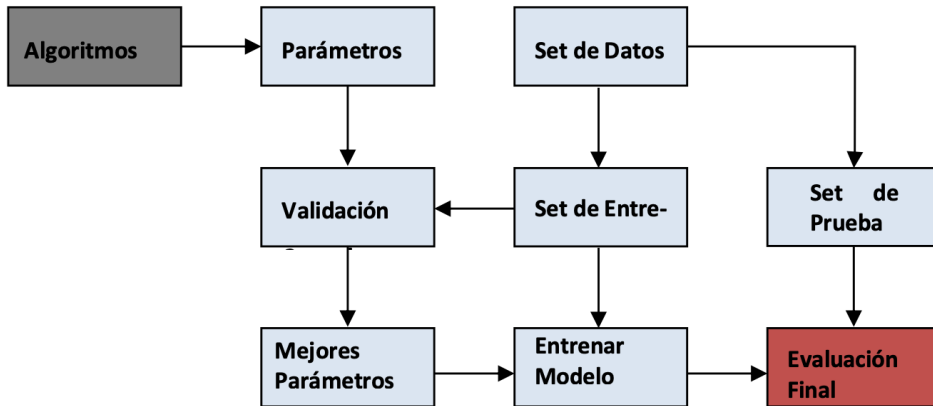
tes. Varios estudios empíricos muestran que los métodos basados en inteligencia artificial superan sus contrapartes econométricas y de series de tiempo que se basan en análisis de big data (Song, Qiu y Park, 2019).

Entre los hoteles de Holguín se encuentra Brisas Guardalavaca perteneciente al grupo de turismo Cubanacán. Se comprueba por revisión documental y entrevistas a los directivos del hotel y trabajadores que el pronóstico de la demanda se realiza utilizando métodos de regresión estadística. Esta situación causa variaciones en las planificaciones al evidenciarse una brecha considerable entre la demanda de clientes pronosticada y la real, lo cual influye además en las desacertadas previsiones de aseguramientos materiales, humanos y financieros necesarios para asumir la demanda. Se identifica como problema científico: Desajustes en los pronósticos de demanda de clientes del hotel Brisas Guardalavaca de Holguín. Para darle solución al problema se definió como objetivo general: evaluar algoritmos de aprendizaje automatizado que ofrezcan buen desempeño para el pronóstico del por ciento de ocupación del hotel.

## 2. MATERIALES Y MÉTODOS

En contraste con el declive de los pronósticos críticos, los modelos basados en inteligencia artificial (IA) han ganado una popularidad cada vez mayor. Este enfoque ha demostrado una capacidad extraordinaria para manejar big data, y se ha demostrado que los modelos basados en IA tienen una precisión superior en la predicción volúmenes de demanda turística (Song *et al.*, 2019). Al respecto, Andía, Arrieta y Sing Long (2019) afirman que existen métodos efectivos para la predicción, como por ejemplo lo son las redes neuronales o los bosques aleatorios.

El trabajo sigue la lógica que se detalla en la figura 1, como algoritmos se utilizan los de regresión lineal, regresión polinomial, máquinas de vectores de soporte (MVS), red neuronal artificial perceptrón multicapa ( $RNA_{PM}$ ) y bosques aleatorios. De cada algoritmo se obtiene el mejor modelo en base a la selección de parámetros utilizando las técnicas de búsqueda en cuadrícula y validación cruzada. Como métricas en la evaluación final se emplean el error cuadrático medio (ECM) y el coeficiente de determinación ( $R^2$ ).



**Figura1.** Selección de parámetros y evaluación de los modelos con las técnicas de búsqueda en cuadrícula y validación cruzada.

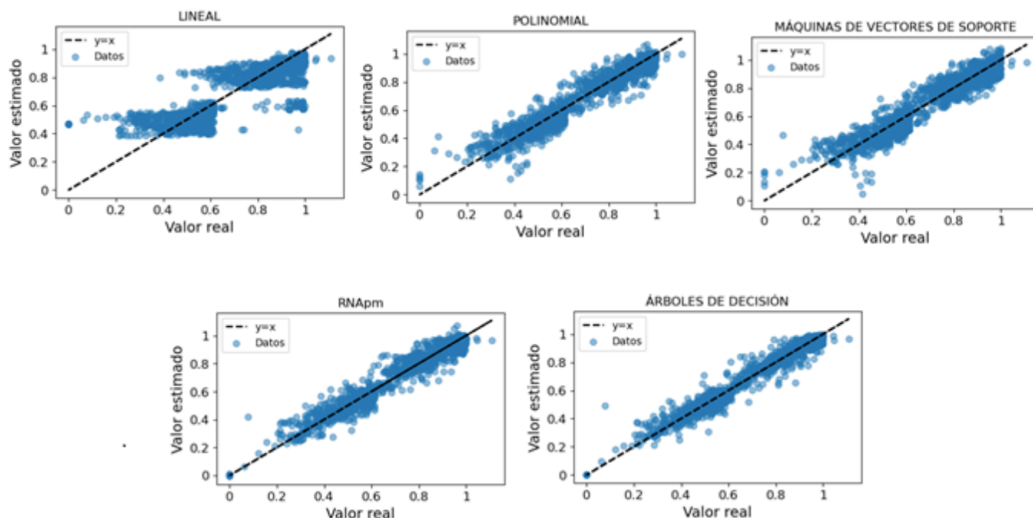
**Fuente:** Adaptado de Müller y Guido (2016).

Para obtener los modelos por regresión lineal se utilizan diferentes métodos regularización y valores del parámetro  $\alpha$ . En el caso de la regresión polinomial se consideran diferentes grados del polinomio. Los modelos generados con MVS utilizan diferentes  $kernel$  y valores de los parámetros  $C$  y  $\gamma$ . Para los modelos de  $RNA_{PM}$  se analizan las diferentes alternativas de arquitecturas en función del número de capas ocultas, cantidad de neuronas dentro de las capas ocultas, las funciones de activación de las capas ocultas y el método para optimizar la función de rendimiento en el entrenamiento. Con la variación de  $n\_estimators$  y  $max\_depth$  se construyen los modelos de bosques aleatorios.

### 3. RESULTADOS Y DISCUSIÓN

Para el pronóstico de la demanda del Hotel se partió de los datos obtenidos del departamento comercial quienes facilitaron para la investigación el por ciento de ocupación histórico (booking) de los últimos siete años. Los 258 modelos obtenidos están diseñados según los datos de cuatro variables de entrada o independientes (sección, día, mes y año) y una variable de salida o dependiente (por ciento de ocupación).

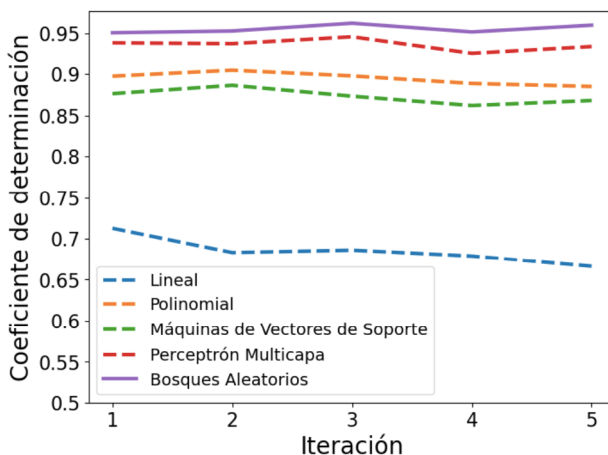
La figura 2 muestra la comparación de las predicciones del por ciento de ocupación según los modelos de regresión lineal, polinomial, MVS,  $RNA_{PM}$  y bosques aleatorios que se obtuvieron con mejor métrica y su comportamiento real con los datos del entrenamiento.



**Figura 2.** Comparación de las predicciones del por ciento de ocupación.

**Fuente:** elaboración propia.

Los mejores modelos seleccionados por cada algoritmo fueron evaluados con la técnica de búsqueda en cuadrícula y validación cruzada. En la figura 3 se muestran los coeficientes de determinación que se obtuvieron, se aprecia que el de bosques aleatorios es el modelo que mejor generaliza este tipo de pronóstico.



**Figura 3.** Coeficientes de determinación obtenidos por búsqueda en cuadrícula y validación cruzada.

**Fuente:** elaboración propia.

La tabla 1 muestra la evaluación final de los modelos con los datos del set de prueba. El de mejor resultado se corrobora es el de bosques aleatorios seguido de la RNA<sub>PM</sub>.

**Tabla 1.** Métricas de los algoritmos en el set de prueba.

| Algoritmos         | R <sup>2</sup> | ECM   |
|--------------------|----------------|-------|
| Bosques Aleatorios | 0.953          | 0.032 |
| RNA <sub>PM</sub>  | 0.935          | 0.040 |
| Polinomial         | 0.908          | 0.050 |
| MVS                | 0.882          | 0.060 |
| Lineal             | 0.687          | 0.087 |

**Fuente:** elaboración propia.

#### 4. CONCLUSIONES

1. Las investigaciones consultadas confirman la factibilidad de utilizar algoritmos de aprendizaje automatizado como técnicas de pronóstico.
2. Se aplicó el procedimiento de Müller y Guido (2016) para la evaluación de los mejores modelos con las técnicas de búsqueda en cuadrícula y validación cruzada.
3. El modelo que ofrece mejor aproximación en las predicciones del por ciento de ocupación del hotel es el de bosques aleatorios con un coeficiente de determinación de un 0.953 y un error cuadrático medio de 0.032, le sigue la RNA<sub>PM</sub> y el menos recomendado es el modelo de regresión lineal.

#### REFERENCIAS BIBLIOGRÁFICAS

- Andía, M. E., Arrieta, C., & Sing Long, C. A.** (2019). Una guía conceptual para usar y entender Big Data en la investigación clínica. *Revista Médica Clínica Las Condes*, 30(1), 83-94. doi: <https://doi.org/10.1016/j.rmclc.2018.11.003>
- Escalona Morales, P.** (2011). *Dinámica de los flujos turísticos de los Mercados emisores hacia el destino Cuba*. Universidad de Holguín, Facultad de Ingeniería Industrial y Turismo .... Retrieved from <https://repositorio.uho.edu.cu/xmlui/handle/uho/4882>
- Gunter, U., & Önder, I.** (2015). Forecasting international city tourism demand for Paris: Accuracy of uni-and multivariate models employing monthly data. *Tourism Management*, 46, 123-135. doi: <https://doi.org/10.1016/j.tourman.2014.06.017>

- Huang, K.-H., & Yu, T. H.-K.** (2019). Application of Google Trends to Forecast Tourism Demand. *Journal of Internet Technology*, 20(4), 1273-1280.
- Li, X., Pan, B., Law, R., & Huang, X.** (2017). Forecasting tourism demand with composite search index. *Tourism Management*, 59, 57-66. doi: <https://doi.org/10.1016/j.tourman.2016.07.005>
- Önder, I., & Gunter, U.** (2016). Forecasting tourism demand with Google Trends for a major European city destination. *Tourism Analysis*, 21(2-3), 203-220. doi: <https://doi.org/10.3727/108354216X14559233984773>
- Ramírez, A. M. A.** (2013). *Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución*. Universidad Tecnológica de Pereira. Facultad de Ingenierías Eléctrica .... Retrieved from <https://core.ac.uk/download/pdf/71397537.pdf>
- Salazar Aguilar, M. A., & Cabrera Ríos, M.** (2007). Pronóstico de demanda por medio de redes neuronales artificiales. *Ingenierías*, 10(35), 6-12.
- Song, H., Qiu, R. T., & Park, J.** (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338-362. doi: <https://doi.org/10.1016/j.annals.2018.12.001>
- Tsui, W. H. K., Balli, H. O., Gilbey, A., & Gow, H.** (2014). Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42, 62-76. doi: <https://doi.org/10.1016/j.tourman.2013.10.008>
- Tussyadiah, I.** (2020). A review of research into automation in tourism: Launching the Annals of Tourism Research Curated Collection on Artificial Intelligence and Robotics in Tourism. *Annals of Tourism Research*, 81, 102883. doi: <https://doi.org/10.1016/j.annals.2020.102883>
- Yang, Y., Pan, B., & Song, H.** (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4), 433-447. doi: <https://doi.org/10.1177/0047287513500391>
- Zhu, L., Lim, C., Xie, W., & Wu, Y.** (2018). Modelling tourist flow association for tourism demand forecasting. *Current Issues in Tourism*, 21(8), 902-916. doi: <https://doi.org/10.1080/13683500.2016.1218827>