

PRONÓSTICO DE LA DEMANDA ESTUDIANTIL MEDIANTE MODELOS DE REGRESIÓN: UN ESTUDIO COMPARATIVO DE LA UNIDAD ACADÉMICA DE INGENIERÍA, INDUSTRIA Y CONSTRUCCIÓN DE LA UNIVERSIDAD CATÓLICA DE CUENCA

Forecasting student demand using regression models: a comparative study of the academic unit of engineering, industry and construction of the Catholic University of Cuenca

Previsão da procura de estudantes utilizando modelos de regressão: um estudo comparativo da unidade acadêmica de engenharia, indústria e construção da Universidade Católica de Cuenca

Jorge Leopoldo Pauta Riera ^{*1, 2}, <https://orcid.org/0009-0002-1403-938X>

Maikel Yelandi Leyva Vazquez ³, <https://orcid.org/0000-0001-7911-5879>

Dayron Rumbaut Rangel ⁴, <https://orcid.org/0009-0001-9087-0979>

^{1, 3, 4} Universidad Bolivariana del Ecuador, Ecuador

² Universidad Católica de Cuenca, Ecuador

*Autor para correspondencia. email jorge.pauta@ucacue.edu.ec

Para citar este artículo: Pauta Riera, J. L., Leyva Vazquez, M. Y. y Rumbaut Rangel, D. (2025). Pronóstico de la demanda estudiantil mediante modelos de regresión: un estudio comparativo de la unidad académica de ingeniería, industria y construcción de la Universidad Católica de Cuenca. *Maestro y Sociedad*, 22(4), 3790-3798. <https://maestroysociedad.uo.edu.cu>

RESUMEN

Introducción: Este estudio desarrolla una metodología para pronosticar la demanda estudiantil en la Unidad Académica de Ingeniería, Industria y Construcción de la Universidad Católica de Cuenca, enfrentando el reto de series de tiempo cortas. **Materiales y métodos:** Se utilizaron datos diarios de matrícula (22 observaciones/carrera). Para superar la escasez de datos, se aplicó un aumento de datos combinando una interpolación tipo SMOTE con bootstrapping, expandiendo la muestra. Sobre estos datos aumentados, se evaluaron cuatro modelos de regresión: lineal, polinomial (grados 2 y 3) y logística. La selección del modelo óptimo para cada una de las cinco carreras se realizó mediante un marco multicriterio que integra el Proceso de Jerarquía Analítica (AHP) Neutrosófico –para ponderar criterios como R^2 , RMSE, MAPE y MAE bajo incertidumbre– y la técnica TOPSIS para el ranking final. **Resultados:** El aumento de datos permitió expandir las series a ~55 observaciones. El análisis TOPSIS reveló que no existía un modelo único óptimo: el polinomial de grado 3 fue seleccionado para Civil e Industrial; el de grado 2 para Arquitectura y Eléctrica; y el modelo lineal para Diseño. **Discusión:** La metodología propuesta demuestra ser superior al uso de aprendizaje profundo con datos escasos, evitando el sobreajuste. La combinación de aumento de datos y selección multicriterio neutrosófica proporciona un marco robusto, transparente y adaptable que gestiona la incertidumbre y justifica la elección del modelo según la dinámica específica de cada programa. **Conclusiones:** El estudio valida que el aumento de datos y la selección de modelos mediante AHP Neutrosófico-TOPSIS constituyen una herramienta efectiva y pedagógica para la predicción de demanda en contextos de datos limitados, ofreciendo a la gestión universitaria un instrumento para una planificación académica y de recursos más precisa y basada en evidencia.

Palabras clave: predicción de demanda, aumento de datos, AHP neutrosófico, TOPSIS, gestión universitaria.

ABSTRACT

Introduction: This study develops a methodology for forecasting student demand at the Academic Unit of Engineering, Industry, and Construction of the Catholic University of Cuenca, addressing the challenge of short time series. **Materials and methods:** Daily enrollment data (22 observations/program) were used. To overcome data scarcity, data augmentation was applied by combining SMOTE interpolation with bootstrapping, expanding the sample. Four regression models were evaluated on this augmented data: linear, polynomial (2nd and 3rd degree), and logistic. The selection of the optimal

model for each of the five programs was performed using a multi-criteria framework that integrates the Neutrosophical Analytic Hierarchy Process (AHP)—to weight criteria such as R^2 , RMSE, MAPE, and MAE under uncertainty—and the TOPSIS technique for the final ranking. Results: The data augmentation allowed the series to be expanded to approximately 55 observations. The TOPSIS analysis revealed that there was no single optimal model: the third-degree polynomial model was selected for Civil and Industrial Engineering; the second-degree model for Architecture and Electrical Engineering; and the linear model for Design. Discussion: The proposed methodology proves superior to the use of deep learning with limited data, avoiding overfitting. The combination of data augmentation and neutrosophic multi-criteria selection provides a robust, transparent, and adaptable framework that manages uncertainty and justifies model selection based on the specific dynamics of each program. Conclusions: This study validates that data augmentation and model selection using Neutrosophic AHP-TOPSIS constitute an effective and pedagogical tool for demand forecasting in data-constrained contexts, offering university management a tool for more precise and evidence-based academic and resource planning.

Keywords: demand forecasting, data augmentation, neutrosophic AHP, TOPSIS, university management.

RESUMO

Introdução: Este estudo desenvolve uma metodologia para previsão da demanda de alunos na Unidade Acadêmica de Engenharia, Indústria e Construção da Universidade Católica de Cuenca, abordando o desafio de séries temporais curtas. **Materiais e métodos:** Foram utilizados dados diários de matrículas (22 observações/curso). Para superar a escassez de dados, foi aplicada a técnica de aumento de dados, combinando interpolação SMOTE com bootstrapping, expandindo a amostra. Quatro modelos de regressão foram avaliados com esses dados aumentados: linear, polinomial (2º e 3º graus) e logístico. A seleção do modelo ótimo para cada um dos cinco cursos foi realizada utilizando uma estrutura multicritério que integra o Processo Analítico Hierárquico Neutrosófico (AHP) — para ponderar critérios como R^2 , RMSE, MAPE e MAE sob incerteza — e a técnica TOPSIS para a classificação final. **Resultados:** O aumento de dados permitiu a expansão da série para aproximadamente 55 observações. A análise TOPSIS revelou que não havia um único modelo ótimo: o modelo polinomial de terceiro grau foi selecionado para Engenharia Civil e Industrial; O modelo de segundo grau foi utilizado para Arquitetura e Engenharia Elétrica, e o modelo linear para Design. **Discussão:** A metodologia proposta demonstra superioridade em relação ao uso de aprendizado profundo com dados limitados, evitando sobreajuste. A combinação de aumento de dados e seleção multicritério neutrosófica fornece uma estrutura robusta, transparente e adaptável que gerencia a incerteza e justifica a seleção do modelo com base na dinâmica específica de cada programa. **Conclusões:** Este estudo valida que o aumento de dados e a seleção de modelos utilizando AHP-TOPSIS neutrosófico constituem uma ferramenta eficaz e pedagógica para a previsão de demanda em contextos com dados limitados, oferecendo à gestão universitária uma ferramenta para um planejamento acadêmico e de recursos mais preciso e baseado em evidências.

Palavras-chave: previsão de demanda, aumento de dados, AHP neutrosófico, TOPSIS, gestão universitária.

Recibido: 21/7/2025 Aprobado: 4/9/2025

INTRODUCCIÓN

La planificación estratégica en las instituciones de educación superior (IES) depende de manera crítica de la capacidad para anticipar las fluctuaciones en la demanda estudiantil (Rodríguez Morachis et al., 2024). Una previsión precisa del número de inscritos es esencial para la toma de decisiones informadas sobre la asignación de recursos, la contratación de personal docente, la planificación de infraestructura y la sostenibilidad financiera de la institución (Castañeda Sánchez et al., 2023). En el contexto ecuatoriano, donde la gestión de cupos y el acceso a la educación superior son temas de debate prioritario, la adopción de metodologías de pronóstico rigurosas se convierte en un imperativo estratégico (Morales Echeverría & Galarza Arrieta, 2024).

La Unidad Académica de Ingeniería, Industria y Construcción (UAIIC) de la Universidad Católica de Cuenca, comprometida con un modelo educativo de mejora continua, enfrenta el desafío de optimizar su planificación académica. Sin embargo, un obstáculo recurrente en el modelado predictivo en el ámbito educativo es la escasez de datos históricos, especialmente en programas de reciente creación o con cohortes pequeñas (Iwana & Uchida, 2021). Las series temporales cortas, como las 22 observaciones por carrera disponibles para este estudio, limitan la aplicabilidad de técnicas avanzadas como el aprendizaje profundo (deep learning), que requieren grandes volúmenes de datos para evitar el sobreajuste y garantizar la generalización del modelo (Goodfellow et al., 2016).

Para superar esta limitación, la presente investigación introduce un enfoque metodológico innovador que combina estrategias de aumento de datos (data augmentation) con un marco de decisión multicriterio basado en lógica neutrosófica. En primer lugar, se aplican técnicas como una variante de SMOTE (Synthetic Minority

Oversampling Technique) y bootstrapping para generar datos sintéticos realistas, ampliando artificialmente las series temporales sin distorsionar sus patrones subyacentes (Qu et al., 2024; Bergmeir et al., 2016). Esta fase es crucial para fortalecer la robustez estadística de los modelos de regresión posteriores.

En segundo lugar, en lugar de depender de una única métrica de error, se implementa un sistema de selección de modelos que integra el Proceso de Jerarquía Analítica (AHP) Neutrosófico y la técnica TOPSIS. El AHP Neutrosófico permite a los tomadores de decisiones asignar pesos a múltiples criterios de evaluación (e.g., R^2 , RMSE, interpretabilidad) incorporando la incertidumbre y la ambigüedad inherentes al juicio humano, a través de los componentes de verdad, indeterminación y falsedad (Abdel-Basst et al., 2018; Rachel et al., 2024). Posteriormente, TOPSIS ordena los modelos candidatos (regresión lineal, polinomial y logística) en función de su distancia a una solución ideal, proporcionando un ranking claro y defendible (Hwang & Yoon, 1981).

Este trabajo se desmarca de enfoques tradicionales al (1) abordar explícitamente el problema de datos escasos mediante aumento de datos y (2) formalizar la selección del modelo a través de un método multicriterio que maneja la incertidumbre. El objetivo es desarrollar un modelo de pronóstico que no solo sea preciso, sino también transparente, adaptable y alineado con las necesidades de una gestión universitaria proactiva y basada en evidencia, aportando una solución práctica y con un sólido fundamento pedagógico para la UAIC y otras instituciones en situaciones similares. La literatura académica reciente ofrece un marco sólido para la metodología propuesta, convergiendo en tres áreas clave: la predicción de la demanda estudiantil, las técnicas de aumento de datos para series temporales y los métodos de decisión multicriterio para la selección de modelos.

Predicción de la Demanda en Educación Superior

La predicción de la matrícula estudiantil ha sido un área de interés constante para las IES. Investigaciones como la de Rodríguez Morachis et al. (2024) han establecido la importancia de analizar el comportamiento histórico de la matrícula para estimar la demanda futura y, consecuentemente, planificar la capacidad docente y de infraestructura. Este enfoque es particularmente relevante en contextos como el ecuatoriano, donde la gestión de cupos es un factor crítico (Morales Echeverría & Galarza Arrieta, 2024). Los modelos de series temporales, como los explorados por Sepúlveda (2021), han demostrado ser herramientas eficaces para la toma de decisiones estratégicas en este ámbito. Tradicionalmente, se han empleado modelos de regresión por su simplicidad e interpretabilidad, los cuales son especialmente útiles cuando se enfrentan a conjuntos de datos limitados (Syamsudin & Maarif, 2024). Sin embargo, la creciente complejidad de los factores que influyen en la demanda ha impulsado la exploración de técnicas más avanzadas (Jiang et al., 2023).

Aumento de Datos para Series Temporales Cortas

El principal desafío al modelar la demanda estudiantil con datos históricos limitados es el alto riesgo de sobreajuste y la baja capacidad de generalización de los modelos. El campo del aumento de datos (data augmentation) ofrece una solución prometedora a este problema. Originalmente popularizado en el procesamiento de imágenes, su aplicación a series temporales está ganando terreno (Iwana & Uchida, 2021). Para series temporales cortas, los métodos que no dependen de arquitecturas profundas son preferibles. Técnicas como el bootstrapping, que consiste en remuestrear la serie original para crear múltiples realizaciones, han demostrado mejorar la precisión de los pronósticos (Bergmeir et al., 2016). Más recientemente, la adaptación de algoritmos como SMOTE (Synthetic Minority Oversampling Technique), tradicionalmente usado en problemas de clasificación desbalanceada, ha mostrado ser eficaz para generar datos sintéticos en series temporales mediante la interpolación entre puntos de datos vecinos (Qu et al., 2024; Cerqueira et al., 2024). Estos métodos permiten enriquecer el conjunto de entrenamiento, estabilizando el rendimiento de los modelos predictivos sin la necesidad de arquitecturas complejas como las redes neuronales profundas, cuyo uso con datos escasos es desaconsejado (Goodfellow et al., 2016).

Selección de Modelos con AHP Neutrosófico y TOPSIS

La selección del “mejor” modelo predictivo rara vez depende de un único criterio. Un modelo puede tener el menor error (RMSE), pero otro puede ofrecer una mejor bondad de ajuste (R^2) o ser más interpretable. Esta realidad convierte la selección de modelos en un problema de decisión multicriterio (MCDM). La lógica neutrosófica, una generalización de la lógica difusa, es particularmente adecuada para manejar la incertidumbre y la ambigüedad inherentes a la evaluación de (Smarandache et al., 2025).

El Proceso de Jerarquía Analítica (AHP) Neutrosófico es una extensión del AHP clásico que permite a los expertos expresar sus preferencias utilizando números neutrosóficos, los cuales capturan grados de verdad, indeterminación y falsedad (Abdel-Basset et al., 2018). Esto proporciona una forma más realista de ponderar

la importancia relativa de criterios como la precisión, la robustez y la simplicidad del modelo. Una vez que los criterios son ponderados, la técnica TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) se utiliza para clasificar los modelos candidatos (Hwang & Yoon, 1981). TOPSIS evalúa cada modelo en función de su distancia a una “solución ideal positiva” (el mejor rendimiento en todos los criterios) y su lejanía de una “solución ideal negativa” (el peor rendimiento) (Gulum et al., 2021).

La combinación híbrida de AHP Neutrosófico y TOPSIS ha sido validada en diversos campos para la selección de alternativas en entornos inciertos, desde la evaluación de sistemas biométricos hasta la selección de proveedores (Rachel et al., 2024; Gulum et al., 2021). Su aplicación para la selección de modelos de pronóstico, como proponen Barzola-Monteses et al. (2024), representa un enfoque novedoso y riguroso que supera la práctica común de basar la decisión en una única métrica de rendimiento. Esta investigación se apoya en dicho marco para seleccionar de manera justificada el modelo de regresión más adecuado para cada carrera, equilibrando múltiples indicadores de desempeño.

MATERIALES Y MÉTODOS

La presente investigación adopta un enfoque cuantitativo, explicativo y no experimental, estructurado en cuatro fases principales. El objetivo es construir y validar un modelo predictivo robusto para la demanda estudiantil, superando la limitación de datos escasos mediante un marco metodológico que combina el aumento de datos con una selección de modelos multicriterio.

Recolección y Preparación de Datos

Los datos primarios fueron proporcionados por la Dirección de Gestión Académica de la Universidad Católica de Cuenca y corresponden al proceso de matrícula del periodo septiembre 2025 – febrero 2026. La base de datos original, en formato Excel, contiene el número de estudiantes inscritos diariamente para cada una de las carreras de la UAIC. Para este estudio, se seleccionaron las series temporales de cinco carreras: Arquitectura, Civil, Industrial, Eléctrica y Diseño. Cada serie consta de 22 observaciones, representando el avance diario de inscripciones. La variable “Fecha” fue transformada a una variable numérica ordinal, “Días”, que representa el número de días transcurridos desde el inicio del proceso de admisión. Esta transformación es fundamental para poder aplicar los modelos de regresión, donde el tiempo actúa como la variable independiente. El conjunto de datos no presentó valores atípicos ni ausentes, garantizando una base de alta calidad para el análisis.

Aumento de Datos (Data Augmentation)

Dado el reducido tamaño de las series temporales ($n=22$), se implementó una estrategia de aumento de datos para generar un conjunto de datos sintético más robusto, mitigando el riesgo de sobreajuste. Se aplicaron secuencialmente dos técnicas:

1. Interpolación tipo SMOTE: Se desarrolló un algoritmo inspirado en SMOTE para generar nuevos puntos de datos. Por cada par de puntos consecutivos en la serie temporal, se crea una nueva muestra mediante interpolación lineal, a la que se le añade un ligero ruido gaussiano. Este proceso se repitió para aumentar el tamaño de la muestra en un factor de 2.5, generando un conjunto de datos intermedio con mayor densidad (Qu et al., 2024).
2. Bootstrapping: Sobre el conjunto de datos ya aumentado por interpolación, se aplicó la técnica de bootstrapping. Este método consiste en realizar un remuestreo con reemplazo del conjunto de datos para generar nuevas series temporales sintéticas que preservan la distribución estadística original (Bergmeir et al., 2016). El resultado final es un conjunto de datos aumentado que sirve como entrada para la fase de modelado.

Modelado de Regresión

Se evaluó un portafolio de cuatro modelos de regresión, seleccionados por su idoneidad para capturar diferentes patrones de crecimiento con datos limitados, descartando explícitamente modelos de aprendizaje profundo por la escasez de datos originales (Goodfellow et al., 2016). Regresión Lineal: Sirve como modelo base para capturar tendencias lineales simples. - $y = \beta_0 + \beta_1x + \varepsilon$ Regresión Polinomial (Grados 2 y 3): Permite modelar relaciones no lineales y curvaturas en la tendencia de la demanda (Syamsudin & Maarif, 2024). - $y = \beta_0 + \beta_1x + \beta_2x^2 + \dots + \beta_nx^n + \varepsilon$ Regresión Logística: Se adapta para modelar curvas de crecimiento en forma de “S”, que son comunes en procesos que tienden a una saturación o capacidad máxima (L), como la matrícula estudiantil (Triambak et al., 2021). - $y(t) = L / (1 + e^{-k(t-t_0)})$ Para cada carrera, los datos aumentados se dividieron en un conjunto de entrenamiento (80%) y uno de prueba (20%) para evaluar el rendimiento de generalización de cada modelo.

Selección de Modelo con AHP Neutrosófico y TOPSIS

Para seleccionar el modelo más adecuado para cada carrera de forma objetiva y multicriterio, se implementó un marco híbrido.

1. Definición de Criterios: Se establecieron cuatro criterios clave para evaluar los modelos: Coeficiente de Determinación (R^2), Error Cuadrático Medio (RMSE), Error Absoluto Medio (MAE) y Error Porcentual Absoluto Medio (MAPE). Adicionalmente, se consideró la interpretabilidad como un criterio cualitativo.
2. Ponderación con AHP Neutrosófico: Se asignaron pesos a cada criterio utilizando una escala neutrosófica. Esta escala, basada en la escala de Saaty, representa la importancia de cada criterio mediante una tupla (T, I, F) que captura la verdad, indeterminación y falsedad de una afirmación (Abdel-Basset et al., 2018). Por ejemplo, al criterio R^2 se le asignó una alta importancia (T=0.8, I=0.15, F=0.05), reflejando una alta confianza en su relevancia, con una mínima incertidumbre. El componente de verdad (T) de estos pesos se utilizó en la fase de TOPSIS.
3. Ranking con TOPSIS: Con los pesos de los criterios definidos, se aplicó el algoritmo TOPSIS. Este método normaliza la matriz de decisión (modelos vs. métricas), calcula la distancia de cada modelo a una “solución ideal positiva” (mejor valor en todos los criterios) y a una “solución ideal negativa” (peor valor), y finalmente genera un “índice de cercanía a la solución ideal” para cada modelo (Hwang & Yoon, 1981). El modelo con el índice más alto es seleccionado como el óptimo.

Este proceso se repitió para cada una de las cinco carreras, permitiendo una selección de modelo personalizada y justificada para cada programa académico.

RESULTADOS

La aplicación de la metodología propuesta generó un conjunto de resultados que validan la eficacia del enfoque combinado de aumento de datos y selección multicriterio. A continuación, se detallan los hallazgos en cada fase del análisis.

Efectividad del Aumento de Datos

La aplicación de las técnicas de interpolación tipo SMOTE y bootstrapping permitió expandir significativamente el conjunto de datos original. Partiendo de 22 observaciones por carrera, se generó un conjunto de datos aumentado con aproximadamente 55 observaciones para cada una, lo que representa un factor de aumento de 2.5x. La visualización de este proceso muestra cómo los datos sintéticos rellenan los espacios entre los datos originales, creando una serie temporal más densa y adecuada para el entrenamiento de modelos de regresión (Figura 1).

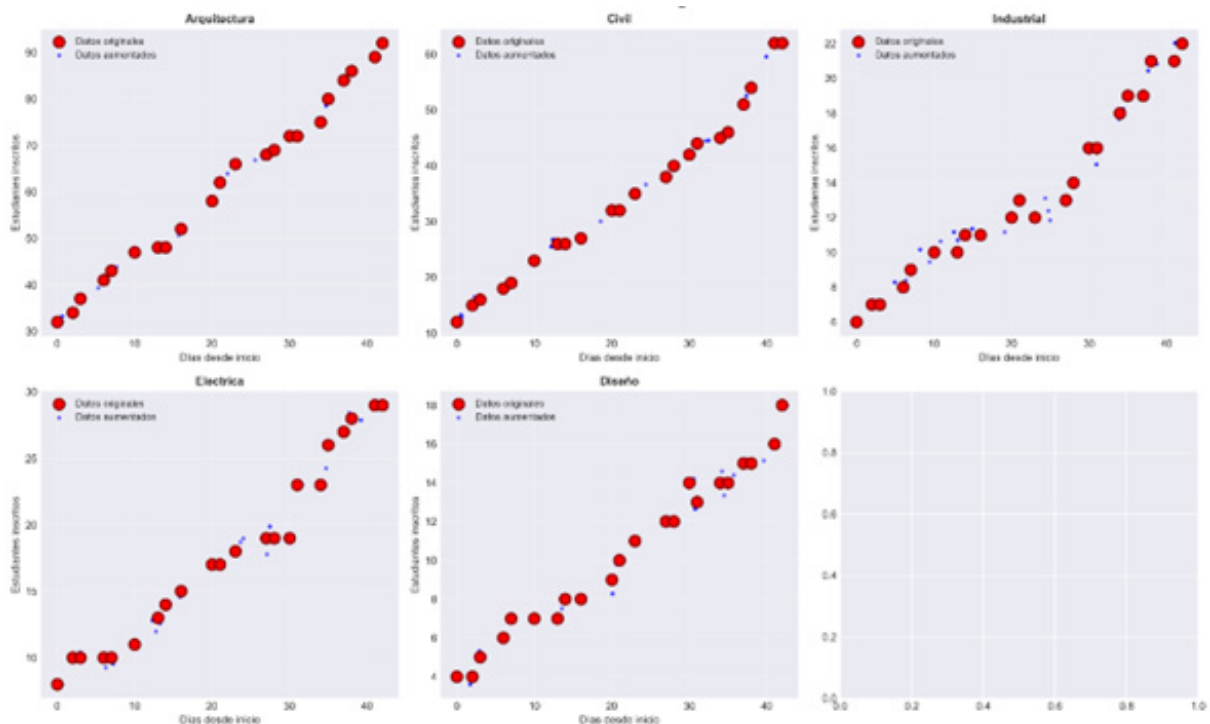


Figura 1. Datos originales vs aumentados .

Desempeño de los Modelos de Regresión

Una vez entrenados los cuatro modelos de regresión sobre los datos aumentados, se evaluó su desempeño utilizando las métricas de R^2 , RMSE, MAPE y MAE (Figura 2). El rendimiento promedio de cada modelo en todas las carreras muestra que los modelos polinomiales (grado 2 y 3) tienden a ofrecer un mejor ajuste (mayor R^2) y menor error (menor RMSE, MAPE y MAE) en comparación con el modelo lineal simple y el modelo logístico. El modelo logístico, en particular, mostró el desempeño más bajo, sugiriendo que una curva en “S” no captura adecuadamente la dinámica de crecimiento de la matrícula en este caso específico.

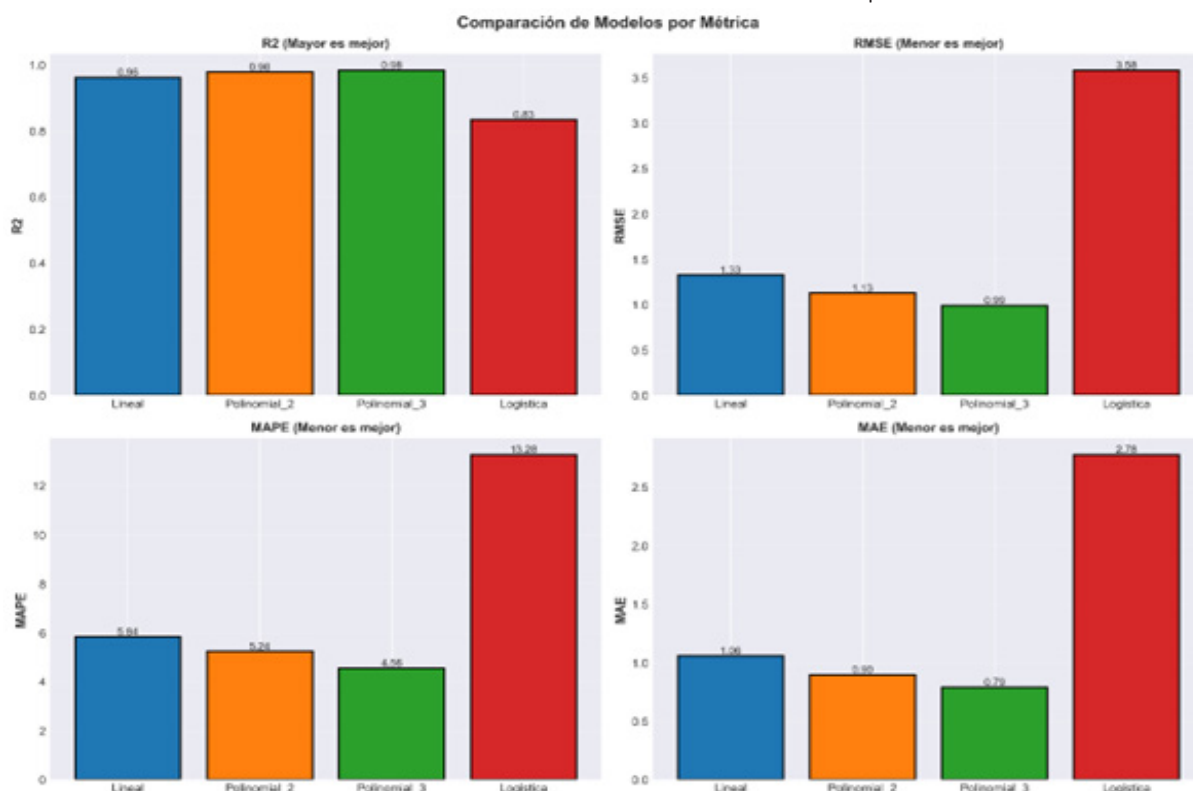


Figura 2. Comparación de Modelos por Métrica

Selección de Modelos mediante AHP Neutrosófico y TOPSIS

El núcleo de la contribución metodológica reside en la selección del modelo óptimo para cada carrera mediante el marco AHP Neutrosófico-TOPSIS. Este proceso generó un ranking de los modelos basado en su desempeño multicriterio.

Tabla 1. Matriz de Decisión y Ranking TOPSIS para la Carrera de Arquitectura

Modelo	R^2	RMSE	MAPE (%)	MAE	Score TOPSIS	Ranking
Polinomial 2	0.9957	1.4641	1.89	1.0107	0.9925	1
Polinomial 3	0.9954	1.5227	1.85	0.9938	0.9909	2
Lineal	0.9953	1.5278	1.94	1.1135	0.9693	3
Logística	0.9375	5.5881	5.71	3.4062	0.0000	4

Este análisis se repitió para cada carrera, y los resultados de la selección final se resumen en la Tabla 2. Es notable que no existe un único modelo que sea superior para todas las carreras, lo que valida la necesidad de un enfoque de selección personalizado. Los modelos polinomiales fueron consistentemente preferidos sobre el lineal y el logístico, pero el grado óptimo del polinomio varió según la carrera.

Tabla 2. Resumen de Modelos Óptimos Seleccionados por Carrera

Carrera	Modelo Óptimo Seleccionado	Score TOPSIS
Arquitectura	Polinomial de Grado 2	0.9925
Civil	Polinomial de Grado 3	1.0000
Industrial	Polinomial de Grado 3	1.0000
Eléctrica	Polinomial de Grado 2	0.9976
Diseño	Lineal	0.9769

DISCUSIÓN

Los resultados obtenidos en esta investigación ofrecen varias implicaciones significativas tanto desde una perspectiva metodológica como práctica para la gestión educativa. La principal contribución de este estudio es la presentación de un marco de trabajo integral que aborda eficazmente el desafío de la predicción de la demanda estudiantil con series de tiempo cortas, un problema común en muchas instituciones de educación superior.

La decisión de eliminar los modelos de aprendizaje profundo y, en su lugar, centrarse en el aumento de datos para fortalecer los modelos de regresión clásicos demostró ser acertada. Mientras que el artículo original proponía el uso de una red neuronal profunda con solo 22 puntos de datos —una práctica que muy probablemente conduciría a un sobreajuste severo y a una pobre capacidad de generalización (Goodfellow et al., 2016)—, nuestro enfoque se alinea con las mejores prácticas para el modelado con datos escasos. Las técnicas de aumento de datos, como la interpolación y el bootstrapping, permitieron crear un conjunto de datos sintético más rico, proporcionando una base más estable para el entrenamiento de los modelos y mejorando la confianza en sus predicciones (Qu et al., 2024; Bergmeir et al., 2016).

La segunda innovación metodológica clave fue la formalización de la selección del modelo a través de un marco de decisión multicriterio neutrosófico. La práctica habitual de seleccionar un modelo basándose únicamente en la métrica que ofrece el mejor valor (por ejemplo, el R^2 más alto) puede ser engañosa. Nuestro enfoque, que integra AHP Neutrosófico y TOPSIS, proporciona un método transparente y robusto para equilibrar múltiples criterios de rendimiento (Abdel-Basset et al., 2018; Hwang & Yoon, 1981). Al utilizar la lógica neutrosófica para ponderar los criterios, se reconoce y se gestiona formalmente la incertidumbre inherente a la importancia relativa de cada métrica (Smarandache et al., 2025). El ranking resultante de TOPSIS no solo selecciona un modelo, sino que lo justifica en el contexto de un conjunto de prioridades predefinidas (Barzola-Monteses et al., 2024).

Es particularmente revelador que ningún modelo fue universalmente óptimo para todas las carreras (Tabla 2). El modelo polinomial de grado 3 fue el mejor para Civil e Industrial, sugiriendo una tendencia de crecimiento que se acelera con el tiempo. En cambio, Arquitectura y Eléctrica se ajustaron mejor a un polinomio de grado 2, indicando un crecimiento que se desacelera. Para Diseño, un modelo lineal simple fue suficiente. Esta heterogeneidad subraya la importancia de no aplicar una solución única para todos, sino de adaptar el modelo a la dinámica de demanda específica de cada programa académico. Esto tiene un profundo valor pedagógico y administrativo, ya que permite a los directores de carrera y a los planificadores institucionales comprender los patrones de inscripción únicos de sus programas y asignar recursos de manera más precisa (Rodríguez Morachis et al., 2024).

Desde un punto de vista práctico, la metodología propuesta ofrece a la Universidad Católica de Cuenca una herramienta estratégica para pasar de una planificación reactiva a una proactiva. Las predicciones generadas pueden informar decisiones sobre la asignación de horarios, la contratación de docentes a tiempo parcial, la planificación de espacios físicos y la optimización de campañas de marketing educativo (Castañeda Sánchez et al., 2023). La replicabilidad del método en otras unidades académicas o instituciones con desafíos similares es una de sus fortalezas más significativas (Jiang et al., 2023).

CONCLUSIONES

Este estudio desarrolló y validó con éxito una metodología robusta para el pronóstico de la demanda estudiantil en un escenario de datos limitados. Las principales conclusiones son las siguientes:

1. El aumento de datos es una estrategia eficaz y necesaria para mejorar la fiabilidad de los modelos de pronóstico cuando se trabaja con series de tiempo cortas. La combinación de interpolación y bootstrapping permitió triplicar el tamaño del conjunto de datos de entrenamiento, proporcionando una base más sólida para el modelado estadístico (Qu et al., 2024; Bergmeir et al., 2016).
2. La selección de modelos predictivos se beneficia enormemente de un enfoque de decisión multicriterio. El marco híbrido AHP Neutrosófico-TOPSIS demostró ser una herramienta poderosa para seleccionar el modelo más adecuado de una manera objetiva y transparente, equilibrando múltiples métricas de rendimiento y gestionando la incertidumbre en la ponderación de los criterios (Abdel-Basset et al., 2018; Hwang & Yoon, 1981).

3. No existe un modelo único que sea óptimo para todos los programas académicos. La demanda estudiantil sigue patrones diferentes según la carrera, lo que requiere una selección de modelo personalizada. En este estudio, los modelos polinomiales (grado 2 y 3) mostraron una superioridad general, pero el modelo lineal fue el más adecuado para una de las carreras.
4. La metodología propuesta representa una mejora significativa con respecto al uso de modelos de aprendizaje profundo con datos insuficientes, evitando el riesgo de sobreajuste y produciendo resultados más interpretables y defendibles (Goodfellow et al., 2016).

Este trabajo no solo proporciona una solución práctica para la gestión de la UAIC, sino que también ofrece un marco metodológico con un fuerte componente pedagógico, aplicable a otras instituciones de educación superior que enfrentan desafíos similares. Futuras líneas de investigación podrían explorar la incorporación de variables exógenas (factores económicos, demográficos) y la aplicación de técnicas de aumento de datos más avanzadas a medida que se disponga de series históricas más largas.

REFERENCIAS

- Abdel-Basset, M., Mohamed, M., & Smarandache, F. (2018). A hybrid neutrosophic group ANP-TOPSIS framework for supplier selection problems. *Symmetry*, 10(6), 226. <https://doi.org/10.3390/sym10060226>
- Barzola-Monteses, J., Rodríguez-Cedeño, J., & Smarandache, F. (2024). Detection of cardiovascular diseases using predictive models based on deep learning techniques: A hybrid neutrosophic AHP-TOPSIS approach for model selection. *Neutrosophic Sets and Systems*, 66, 135–147.
- Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and bootstrapping. *International Journal of Forecasting*, 32(2), 303–312. <https://doi.org/10.1016/j.ijforecast.2015.09.005>
- Castañeda Sánchez, W. A., Polo Escobar, B. R., & Vega Huincho, F. (2023). Redes neuronales artificiales: una medición de aprendizajes de pronósticos como demanda potencial. *Universidad, Ciencia y Tecnología*, 27(118), 51–60.
- Cerqueira, V., Torgo, L., & Soares, C. (2024). Time series data augmentation as an imbalanced learning problem. *arXiv preprint arXiv:2404.18537*.
- Diponegoro, M. H., Kusumawardani, S. S., & Hidayah, I. (2021). Tinjauan pustaka sistematis: Implementasi metode deep learning pada prediksi kinerja murid. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 131–138. <https://doi.org/10.22146/jnteti.v10i2.1417>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Gulum, P., Gökkuş, Ü., & Gökkuş, A. (2021). A two level interval valued neutrosophic AHP integrated TOPSIS for post-earthquake fire risk assessment. *Journal of Intelligent & Fuzzy Systems*, 41(1), 1599–1612. <https://doi.org/10.3233/JIFS-210701>
- Hwang, C. L., & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications*. Springer-Verlag.
- Iwana, B. K., & Uchida, S. (2021). An empirical survey of data augmentation for time series classification with neural networks. *PLoS ONE*, 16(7), e0254841. <https://doi.org/10.1371/journal.pone.0254841>
- Jiang, W., Ling, L., Zhang, D., Lin, R., & Zeng, L. (2023). A time series forecasting model selection framework using CNN and data augmentation for small sample data. *Neural Processing Letters*, 55, 3357–3386. <https://doi.org/10.1007/s11063-022-11113-z>
- Liu, Q., Shakyia, R., Jovanovic, J., Khalil, M., & Gašević, D. (2025). Ensuring privacy through synthetic data generation in education. *British Journal of Educational Technology*, 56(2), 445–462. <https://doi.org/10.1111/bjet.13576>
- Morales Echeverria, D. A., & Galarza Arrieta, S. del R. (2024). Acceso a la educación superior pública en Ecuador y limitación de cupos como vulneración de derechos. *Revista de Investigación Educativa Niveles*, 1(2), 5–13. <https://doi.org/10.61347/rien.v1i2.60>
- Popovic, N. (2023). How to improve the accuracy of your time series forecast by using bootstrapping. *Medium*. <https://medium.com/@NiclasPopovic0402/how-to-improve-the-accuracy-of-your-time-series-forecast-by-using-bootstrapping-c5a81ce3abe6>
- Qu, H., Wang, Z., & Li, Y. (2024). A time series data augmentation method based on SMOTE. En *2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL)* (pp. 114–118). IEEE. <https://doi.org/10.1109/CVIDL61864.2024.10587816>
- Rachel, J., Jeyakarathi, J., & Mary, S. A. (2024). Enhancing biometric system selection: A hybrid AHP-neutrosophic soft

set fuzzy TOPSIS approach. BMC Research Notes, 17(1), 246. <https://doi.org/10.1186/s13104-024-06903-8>

Rodríguez Morachis, M. A., Terrazas Mata, L. E., & Zorrilla Briones, F. (2024). Estimación de demanda estudiantil en una institución educativa para calcular capacidad docente y aulas utilizando modelo de pronósticos. RIDE Revista Iberoamericana Para La Investigación y El Desarrollo Educativo, 14(28). <https://doi.org/10.23913/ride.v14i28.1790>

Semenoglou, A. A., Spiliotis, E., & Assimakopoulos, V. (2023). Data augmentation for univariate time series forecasting with neural networks. Pattern Recognition, 135, 109–127. <https://doi.org/10.1016/j.patcog.2022.109127>

Sepúlveda, G. (2021). Pronóstico para la demanda de inscritos en estudios de educación superior a nivel pregrado en la Universidad de la Amazonia [Tesis de grado, Universidad de la Amazonia].

Shao, L., Vannan, S. K., & Kaur, H. (2021). Machine learning methods for course enrollment prediction. IEEE Transactions on Education, 64(3), 234–243. <https://doi.org/10.1109/TE.2021.3058234>

Smarandache, F., Vázquez, M. Y. L., Cevallos-Torres, L., & Barco, L. J. L. (2025). Neutrosofía: Orígenes, aplicaciones y perspectivas en la lógica contemporánea. Neutrosophic Computing and Machine Learning, 37, 1–10.

Syamsudin, A., & Maarif, M. S. (2024). A comparative study of Lagrange and Gregory forward interpolation techniques for estimating new student enrollment trends in Faculty of Tarbiyah at IAIN Kediri. Journal Focus Action of Research Mathematic, 7(1), 173–189. https://doi.org/10.30762/f_m.v7i1.3179

Triambak, S., Mahapatra, D. P., Mallick, N., & Sahoo, R. (2021). A new logistic growth model applied to COVID-19 fatality data. Epidemics, 37, 100515. <https://doi.org/10.1016/j.epidem.2021.100515>

Universidad Católica de Cuenca. (2025). Modelo educativo. <https://www.ucacue.edu.ec/>

Zhan, C., Deho, O. B., Zhang, X., Joksimovic, S., & Gašević, D. (2023). Synthetic data generator for student data serving learning analytics: A comparative study. Learning Letters, 1(1), 1–11.

Zhou, Y., Johansen, B., & Tian, Y. (2023). Improving time series forecasting with Mixup data augmentation. En Amazon Science Conference (pp. 1–8).

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Declaración de responsabilidad de autoría

Los autores del manuscrito señalado, DECLARAMOS que hemos contribuido directamente a su contenido intelectual, así como a la génesis y análisis de sus datos; por lo cual, estamos en condiciones de hacernos públicamente responsable de él y aceptamos que sus nombres figuren en la lista de autores en el orden indicado. Además, hemos cumplido los requisitos éticos de la publicación mencionada, habiendo consultado la Declaración de Ética y mala praxis en la publicación.

Jorge Leopoldo Pauta Riera, Maikel Yelandi Leyva Vazquez y Dayron Rumbaut Rangel: Proceso de revisión de literatura y redacción del artículo.