

Artificial Intelligence for Project Managers: Predicting Construction Delays

Inteligencia Artificial para líderes de proyectos: anticipando retrasos en construcción

Cristian D. Tobar-Montilla ^{1,2*} , Mariela Muñoz-Añasco ^{2*} , Adriana M. Nieto-Muñoz ^{3*} , Elvia Ruiz-Beltrán ^{4*} 

¹ Centro de Desarrollo Tecnológico Creativ; cristiantobar@unicauca.edu.co

² Universidad del Cauca; mamunoz@unicauca.edu.co

³ Corporación Universitaria Comfacauc-Unicomfacauc; anieto@unicomfacauc.edu.co

⁴ Tecnológico Nacional de México/Instituto Tecnológico de Aguascalientes; erui@aguascalientes.tecnm.mx

* Correspondencia: cristiantobar@unicauca.edu.co

Recibido: 21/11/2025

Aceptado: 17/12/2025

Fecha de Publicación: 18/12/2025

<https://doi.org/10.57173/ritc.v1n19a8>



Derechos de autor: © 2025

Resumen: Los retrasos en proyectos de construcción representan una de las principales causas de sobrecostos y falta de confianza en las empresas del sector, lo que afecta especialmente a las micro, pequeñas y medianas empresas (MiPymes), que enfrentan condiciones de planificación más frágiles y un reducido acceso a herramientas de analítica predictiva. Este estudio propone una metodología basada en el modelo CRISP-DM para anticipar retrasos mediante técnicas de aprendizaje automático supervisadas, como Bosques Aleatorios, Máquinas de Soporte Vectorial, Redes Neuronales, y no supervisadas, como Agrupamiento Jerárquico, utilizando un conjunto de datos sintético que simula las condiciones de proyectos de construcción, y se evaluaron diferentes estrategias para definir umbrales para clasificar el retraso. Los resultados muestran que los enfoques híbridos logran un mejor desempeño que la definición de umbrales fijos, al alcanzar altos niveles de precisión (97%) y revelar patrones latentes, como los “pequeños proyectos con grandes retrasos”, que evidencian la vulnerabilidad de obras de menor escala ante incumplimientos de plazo. La metodología desarrollada muestra que, con pocas variables, como área, duración y precio por metro cuadrado, los líderes de proyectos de las MiPymes pueden implementar modelos predictivos replicables que fortalecen la planificación y la toma de decisiones.

Palabras clave: Retrasos en proyectos de construcción; Aprendizaje automático; Aprendizaje supervisado y no supervisado; CRISP-DM

Abstract: Delays in construction projects are among the main causes of cost overruns and loss of confidence in the sector, especially affecting micro, small, and medium-sized enterprises (MSMEs) that face fragile planning conditions and limited access to predictive analytic tools. This study proposes a methodology based on the CRISP-DM framework to anticipate delays using supervised machine learning techniques, such as Random Forest, Support Vector Machines, and Artificial Neural Networks, along with unsupervised methods like hierarchical clustering. A synthetic dataset simulating construction project conditions was used, and different strategies were evaluated to define thresholds for delay classification. The results show that hybrid approaches outperform fixed threshold definitions, achieving high accuracy (97%) and revealing latent patterns, such as “big delayed little projects,” which highlight the vulnerability of small-scale works to schedule overruns. The developed methodology demonstrates that, with just a few simple variables like built area, project duration, and cost per square meter, MSME project managers can implement replicable predictive models that improve planning and decision-making.

Keywords: Construction project delays; Machine learning; Supervised and unsupervised learning; CRISP-DM

1. Introducción

En Colombia, los proyectos de construcción representan uno de los sectores más dinámicos y estratégicos de la economía del país, pero también uno de los más vulnerables a la incertidumbre en su programación y ejecución. Los retrasos son un fenómeno recurrente que genera sobrecostos y baja productividad en la organización, así como pérdida de confianza por parte de los clientes, públicos o privados, que pueden tener [1], [2], [3]. Estas circunstancias se hacen más evidentes en las MiPymes que componen más del 90% del tejido empresarial del sector y tienen un rol predominante en la generación de empleo y la ejecución de obras de vivienda y públicas [4]. A pesar de su relevancia, la mayoría de estas empresas enfrentan condiciones de planificación más frágiles y escaso acceso a herramientas de capacidad analítica para gestionar datos y predecir riesgos asociados al cumplimiento de los plazos del proyecto [5], como días adicionales de trabajo, extensión de los tiempos de entrega e inicios tardíos de actividades [6].

En la práctica, los líderes de proyecto de construcción han recurrido a la experiencia de expertos y consultores para anticipar retrasos, el índice de importancia relativa (RII) y análisis estadísticos descriptivos en sus históricos. Estas aproximaciones han sido útiles para priorizar causas de retrasos y generar marcos de decisión basados en la experiencia acumulada de profesionales del sector [7]. Sin embargo, su dependencia a la subjetividad reduce su eficacia en entornos más complejos para el proyecto.

Este escenario es relevante para la aplicación de técnicas de Inteligencia Artificial (IA) como el aprendizaje automático (*Machine Learning*, ML) que detectan patrones en los datos históricos y predicen las condiciones ante las cuales se puede producir un retraso en el proyecto. Estudios recientes han demostrado que modelos de aprendizaje automático supervisado (que se caracterizan por estar impulsados por la tarea, es decir, requieren ejemplos previos para entrenar el modelo) como Bosques Aleatorios (BA), Máquinas de Soporte Vectorial (MSV) y Redes Neuronales (RNA) pueden capturar relaciones complejas entre variables técnicas, financieras y de gestión [8], [9]. Asimismo, su implementación integrada con modelos no supervisados de aprendizaje (descubren relaciones, agrupaciones o estructuras que no son evidentes en el conjunto de datos) como las técnicas de agrupamiento como *K-means* y agrupamiento jerárquico aumenta la precisión en la predicción de retrasos [10].

Sin embargo, la adopción de estas herramientas por parte de las MiPymes sigue siendo limitada debido a, primero, que los modelos de inteligencia artificial son percibidos como “cajas negras” sobre las cuales no es posible explicar con claridad cómo se llega a una determinada predicción, y segundo, la falta de metodologías adaptadas a la escala de las MiPymes. En respuesta a esta brecha, este trabajo presenta una metodología replicable basada en el método CRISP-DM (*Cross Industry Standard Process for Data Mining*) [11], para anticipar retrasos de obra utilizando ML. De esta manera, las MiPymes tienen un marco de trabajo de referencia que pueden aplicar de forma práctica incluso en bases de datos de pocas variables y avanzar hacia una gestión predictiva de retrasos.

2. Enfoque CRISP-DM y sus resultados

El método se estructuró según el modelo CRISP-DM. Este enfoque permite desarrollar soluciones de minería de datos sistemáticas, estándar y replicables para la industria. En esta sección se describen las fases de proceso junto con los resultados obtenidos de cada una de ellas, con el propósito de mostrar de manera práctica cómo el enfoque puede ser aplicado por líderes de proyectos en MiPymes constructoras. El esquema general de las fases del CRISP-DM y el flujo de procesamiento de datos se presentan en la Figura 1.

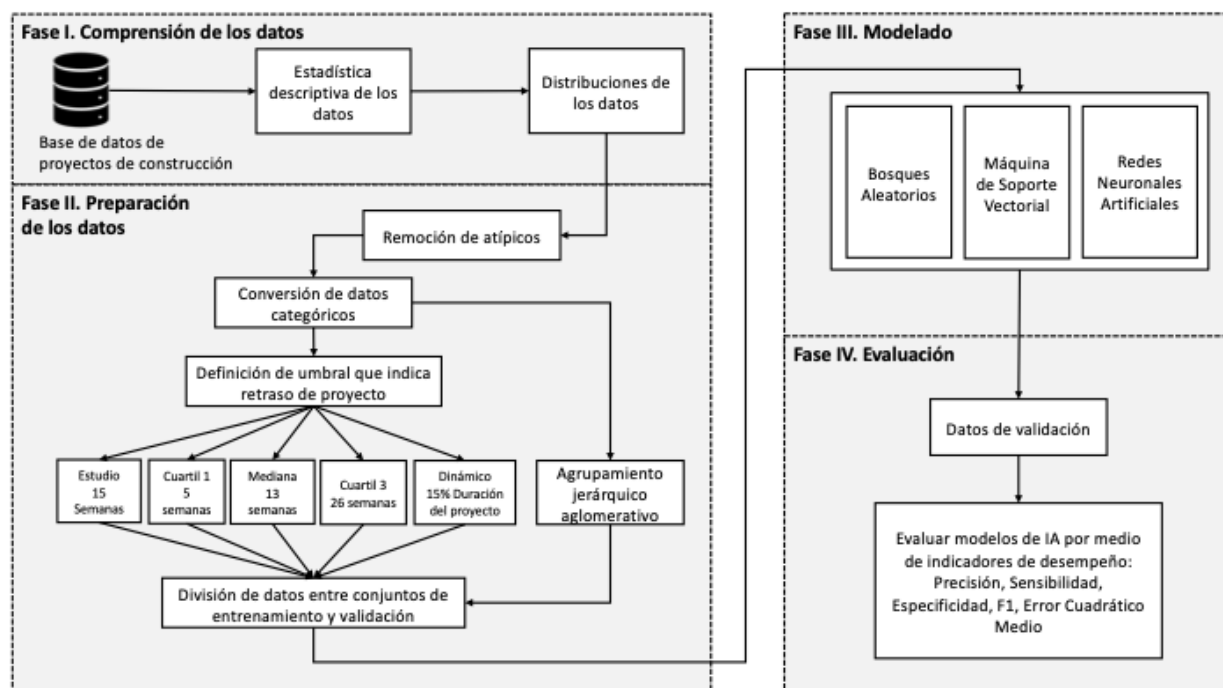


Figura 1. Esquema metodológico basado en el modelo CRISP-DM para la predicción de retrasos en proyectos de construcción.

Tabla 1. Descripción estadística de las variables del conjunto de datos utilizado en el estudio.

Nombre de la Variable	Descripción	Media	Desv. Est.	Min	Max	N
Área construida	Área de construcción, sumando los pisos en caso de un edificio de varios niveles.	8573.88	5835.11	50.76	32385.79	3753
Precio por metro cuadrado	Costo (en dólares) aproximado de construcción por metro cuadrado.	737.78	251.40	182.09	1903.91	3753
Duración del proyecto	Duración total del proyecto de construcción en semanas.	54.64	19.99	13.00	122.00	3753
Semanas de retraso	Retraso total del proyecto de construcción en semanas.	17.80	16.50	0	113.00	3753
Tipología	Unifamiliar	0.12				(450) 3753
	Colectivo	0.37				(1388) 3753
	Comercial	0.38				(1426) 3753
	Otros	0.13				(487) 3753

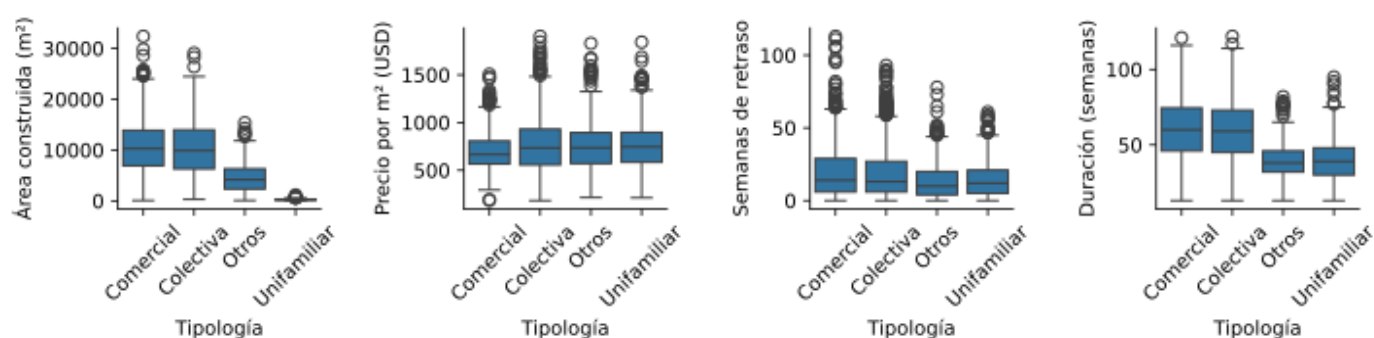


Figura 2. Distribución del área construida, precio por metro cuadrado, duración del proyecto y semanas de retraso en proyectos de construcción según su tipología.

2.1. Compresión de los datos

El estudio utilizó la base de datos desarrollada por [12], que contiene información simulada sobre características físicas, económicas y temporales del proyecto, empleando distribuciones normales controladas. De esta manera, se recrean distintos escenarios de proyectos constructivos.

La base de datos, incluye 3573 registros y tiene las siguientes variables: i) Área construida como la superficie total construida del proyecto en m², ii) Precio por metro cuadrado en dólares (\$/m²), iii) Duración del proyecto en semanas, iv) Semanas de retraso al cronograma inicial, en semanas, y v) tipología o tipo de edificación construida como unifamiliar, colectivo, comercial y otros.

Esta etapa incluyó la revisión estadística descriptiva (Ver Tabla 1) donde se aprecia que, en promedio, los proyectos tienen un área construida de 8573.88 m² con una desviación estándar de 5835.11 m², con un precio medio de 737.78 USD/m². La duración media de las obras fue de 54.6 semanas, con un retraso promedio de 17.8 semanas. Los proyectos comerciales (38%) y colectivos (37%) concentran la mayor proporción de observaciones, seguidos por los proyectos unifamiliares (12%) y otros tipos (13%).

Por otra parte, con el propósito de visualizar el comportamiento de las variables, se identificaron sus distribuciones por tipología (Ver

Figura 2). Se observa una alta variabilidad en el área construida, especialmente en proyectos comerciales y colectivos, así como valores más elevados y dispersos, coherentes con edificaciones de mayor escala y complejidad. En contraste, los proyectos unifamiliares muestran áreas significativamente menores y una menor dispersión, lo que refleja su homogeneidad en tamaño y alcance. En cuanto al precio por metro cuadrado, la mayoría de las tipologías mantienen valores concentrados entre 500 y 1000 USD/m². La distribución de las semanas de retraso presenta medianas relativamente similares entre las tipologías, con una ligera tendencia a retrasos mayores en los proyectos comerciales y colectivos. Finalmente, la duración total de los proyectos muestra patrones consistentes con el área construida: las obras comerciales y colectivas tienden a extenderse por más de 60 semanas, mientras que las unifamiliares se mantienen en 40 semanas.

2.2. Preparación de los datos

En esta etapa se ejecutaron la limpieza y la transformación de variables. Inicialmente, se detectaron valores atípicos y se codificaron las categorías de tipología constructiva. Se definieron distintos umbrales para generar las etiquetas de la variable de salida. De esta manera, se asignó el valor 1 a los proyectos con retraso y 0 a los sin retraso en la tarea de clasificación. Para ello, se implementaron 5 estrategias que simulan distintos criterios que un experto podría sugerir al líder de proyecto como umbral de alarma para retrasos no admisibles: valor propuesto por [12] (15 semanas), mediana (13 semanas), primer y tercer

cuartil (5 y 26 semanas respectivamente) de las semanas de retraso, un umbral dinámico proporcional (15%) a la duración del proyecto.

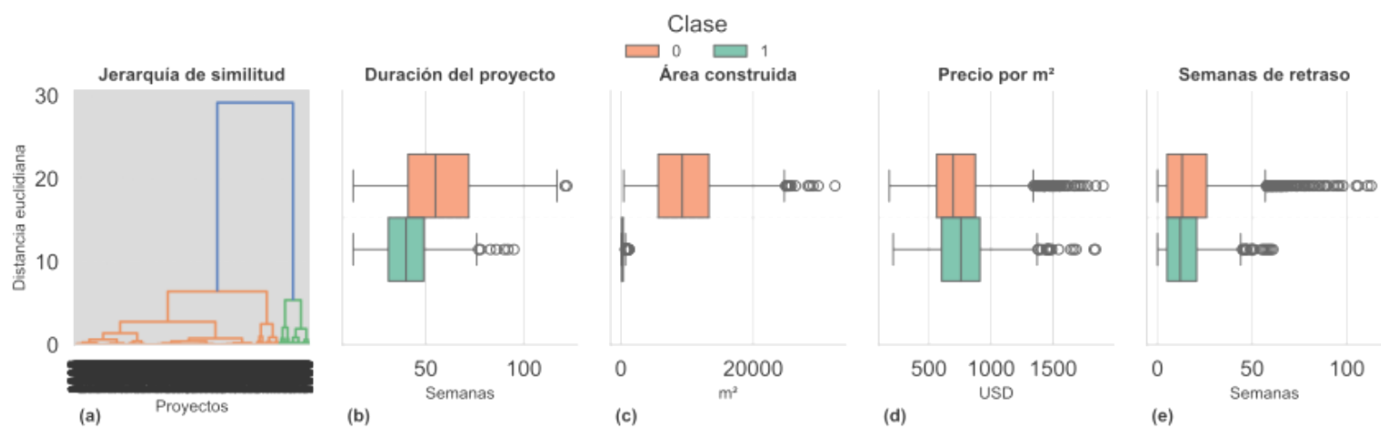


Figura 3. Dendrograma jerárquico y distribución de variables por clase. (a) Agrupamiento jerárquico de los proyectos mediante el método de Ward, mostrando la similitud entre los proyectos y las dos clases mayoritarias. (b-e) Distribución de las variables de duración del proyecto, área construida, precio por metro cuadrado y semanas de retraso según la clase identificada. La clase 1 (verde) corresponde al grupo de “pequeños proyectos con grandes retrasos”.

Finalmente, como alternativa a la definición manual de umbrales, se utilizó el agrupamiento jerárquico para la etiquetación de la variable de salida, lo que permite agrupar proyectos con comportamientos similares.

2.2.1. Agrupamiento jerárquico

El agrupamiento jerárquico es una técnica de aprendizaje no supervisado que permite agrupar proyectos según su similitud. A diferencia de las estrategias que utilizaron umbrales fijos para clasificar los proyectos como retrasados o no retrasados, esta técnica permitió generar grupos a partir de la estructura interna de los datos, sin imponer reglas externas.

El número de grupos se eligió a partir del dendrograma de la agrupación (Ver **Error! No se encuentra el origen de la referencia.** a). Cada línea del gráfico representa la fusión de dos grupos de proyectos que presentan similitudes en sus características, y la altura de la unión indica el grado de disimilitud entre ellos. De esta forma, el dendrograma revela la estructura jerárquica interna de los datos y permite determinar un punto de corte para definir un número deseado de grupos. Para este estudio se definen dos clases.

La Figura 3.b-e presenta los diagramas de caja de las variables de la base de datos, que evidencian el comportamiento de los proyectos clasificados en las dos clases identificadas por el algoritmo de agrupamiento (etiquetadas como “0” y “1”).

Los proyectos del grupo 0 presentan mayores áreas y duraciones, lo que sugiere que corresponden a obras de mayor escala y complejidad. En contraste, los del grupo 1 son más pequeños, de menor costo por metro cuadrado, pero registran niveles más altos de retraso. Estos proyectos, denominados “Pequeños proyectos con grandes retrasos”, podrían indicar que, aunque más pequeños, enfrentan desafíos distintos que afectan su cumplimiento, posiblemente influenciados por restricciones logísticas, falta de planeación o debilidad en los mecanismos de control. Este resultado demuestra la utilidad del enfoque no supervisado para detectar patrones que pueden no ser evidentes al emplear criterios arbitrarios, como la mediana o los cuartiles.

Finalmente, se dividió el conjunto de datos en dos subconjuntos: uno para entrenamiento y otro para prueba (o validación), asegurando que los modelos fueran evaluados sobre datos que no habían sido vistos durante la fase de entrenamiento. Esta

división se realizó de forma aleatoria, reservando el 30% de los datos para prueba (1126 proyectos) y el 70% restante para el entrenamiento del modelo (2626 proyectos).

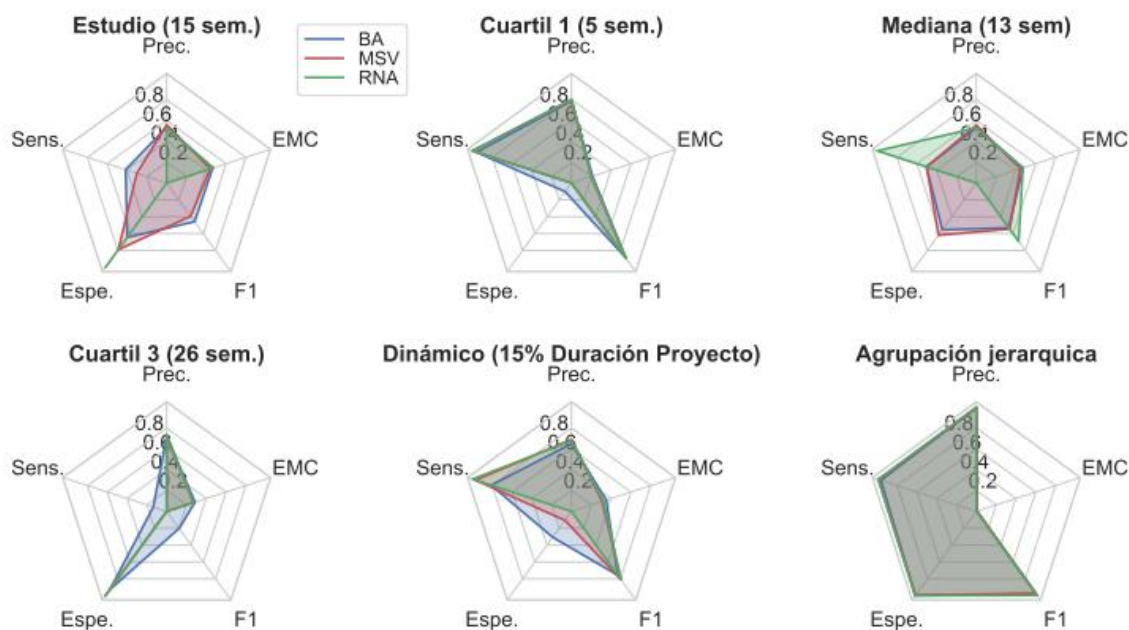


Figura 4. Desempeño comparativo de tres modelos de clasificación (BA, MSV y RNA) bajo seis enfoques para identificar retrasos en proyectos de construcción. Cada diagrama muestra las métricas de precisión (Prec.), sensibilidad (Sens.), especificidad (Espe.), F1 Score y el error medio cuadrático (EMC) para cada enfoque.

2.3. Modelado

Una vez que se definieron las etiquetas de clasificación para los proyectos (retrasados o no retrasados), y se dividieron los conjuntos de datos en entrenamiento y prueba, se procedió a entrenar diferentes modelos de inteligencia artificial con el objetivo de predecir esta condición sobre el conjunto de datos.

Los modelos implementados fueron: Bosques Aleatorios (BA), Máquinas de Soporte Vectorial (MSV) y Redes Neuronales Artificiales (RNA). La elección de estos algoritmos se debe a su capacidad para adaptarse a distintos tipos de relaciones entre variables y a su amplio uso en tareas de clasificación binaria.

2.4 Evaluación

Finalmente, los modelos de BA, MSV y RNA se evaluaron mediante indicadores clásicos de desempeño de las tareas de clasificación: precisión, sensibilidad, especificidad, f1-score y error medio cuadrático (EMC). Esta fase permitió determinar el modelo con mejor equilibrio entre desempeño y capacidad de generalización bajo cinco esquemas distintos para clasificar los proyectos retrasados y un sexto para clasificar “pequeños proyectos con grandes retrasos”. La Figura 4 presenta los gráficos de radar, en los que la escala radial va de 0 (centro) a 1 (periferia), lo que indica el valor alcanzado por cada métrica. Cuanto más cercana a la periferia esté la línea de un modelo, mejor es su desempeño en esa métrica.

De esta manera, los resultados evidenciaron un comportamiento diferenciado según el umbral empleado para definir los proyectos retrasados. Con un umbral fijo de 15 semanas, la precisión de los modelos fue moderada (53-56 %) y la sensibilidad fue baja, especialmente en la red neuronal. Al utilizar el primer cuartil (5 semanas), la sensibilidad aumentó hasta alcanzar el máximo valor (1) en SVM y RNA, aunque con una pérdida de especificidad, lo que generó una alta proporción de falsos positivos. En contraste, el tercer

cuartil (26 semanas) mostró un efecto opuesto: sensibilidad muy baja y especificidad superior al 90%, lo que revela un comportamiento conservador en la detección de retrasos.

El umbral basado en la mediana (13 semanas) ofreció un equilibrio moderado entre sensibilidad y precisión (entre 0.5 y 0.6), mientras que el enfoque adaptativo, definido como el 15% de la duración del proyecto, logró combinar alta sensibilidad (hasta 1.0 en RNA) con precisión superior al 0.67 y un error cuadrático medio inferior a 0.35. Finalmente, el agrupamiento jerárquico resultó ser el método más sobresaliente: los tres modelos alcanzaron métricas superiores al 97% en precisión, sensibilidad, especificidad y F1-score, con un error mínimo (0.0009-0.0075), lo que evidencia la eficacia de complementar un modelo de clasificación con un enfoque no supervisado para generar etiquetas relevantes y detectar patrones ocultos de retraso.

Los resultados obtenidos evidencian que el desempeño de los modelos depende de la forma en que se define el umbral de retraso, que puede definirse arbitrariamente por expertos o a partir de las características técnicas, económicas y temporales del proyecto. En la siguiente sección se discuten estas diferencias y sus implicaciones para la gestión predictiva del retraso del proyecto en las MiPymes constructoras.

3. Discusión

Los resultados de este estudio indican que es posible anticipar retrasos en proyectos de construcción mediante modelos de aprendizaje automático, incluso con bases de datos reducidas y de carácter sintético. El hallazgo más significativo provino del agrupamiento jerárquico, que permitió identificar una clase de proyectos de menor escala, pero altamente propensos al incumplimiento de plazos, denominados “pequeños proyectos con grandes retrasos”, lo que demuestra que los retrasos significativos no son exclusivos de grandes obras, sino también de proyectos pequeños y menos complejos.

El desempeño de los modelos supervisados varió según el criterio de clasificación del retraso. Con umbrales fijos, la sensibilidad fue mayoritariamente baja; esto significa que los modelos fallan al identificar proyectos que efectivamente se retrasan, lo cual, para las MiPymes constructoras, es crítico porque implica que las empresas no se enteran a tiempo de posibles retrasos reales, lo que impide tomar medidas preventivas. En cambio, con umbrales más flexibles (como el 15% de la duración del proyecto), los modelos alcanzaron un equilibrio entre sensibilidad y precisión, evitando los sesgos asociados a valores fijos arbitrarios. El enfoque no supervisado resultó el más sólido, con métricas superiores al 97% en precisión, sensibilidad y especificidad, al permitir que las propias características del proyecto definieran las clases. Estos hallazgos coinciden con los resultados de [3] y [10], que destacan la complementariedad entre métodos híbridos (supervisados y no supervisados) para mejorar la interpretabilidad y la precisión de los modelos.

Desde una perspectiva práctica, anticipar retrasos mediante estos modelos ofrece beneficios directos para los líderes de proyectos de las MiPymes constructoras, principalmente en la reducción de sobrecostos y en la mejora de la planificación. Tal como plantea [2], prever desviaciones permite asignar recursos de manera preventiva y reducir gastos derivados de la extensión de plazos. Además, la implementación de modelos predictivos replicables dentro del marco CRISP-DM fortalece la capacidad de los líderes de proyecto para identificar riesgos tempranos y priorizar intervenciones [13].

Finalmente, aunque el uso de una base sintética limita la generalización, el estudio cumple con el propósito metodológico: evidenciar que las MiPymes constructoras pueden incorporar analítica predictiva con pocas variables, utilizando herramientas abiertas como Python y su librería *scikit-learn*, para generar valor a partir de sus propios datos. Afirmación con la que coinciden [14] y [15].

Como líneas de trabajo futuras, se plantea la validación empírica de los modelos con bases de datos reales provenientes de empresas locales, lo que permitirá calibrar los algoritmos a las condiciones específicas del contexto colombiano. Asimismo, se propone explorar tareas de regresión, como la predicción de semanas exactas de retraso o de

incrementos de costo, para ampliar la capacidad explicativa de los modelos. Finalmente, integrar estas herramientas en sistemas de monitoreo continuo y en tableros de control empresariales podría fortalecer la adopción práctica de la inteligencia artificial en la planeación de proyectos de construcción.

4. Conclusiones

El estudio demuestra que es posible anticipar retrasos en proyectos de construcción mediante modelos de aprendizaje automático, incluso con bases de datos reducidas y de carácter sintético. Los resultados confirman que los enfoques híbridos, que combinan métodos supervisados y no supervisados, ofrecen un mejor desempeño que los criterios fijos basados en umbrales de semanas, logrando una mayor precisión y revelando patrones ocultos en los datos. Entre ellos, destaca la identificación de los “pequeños proyectos con grandes retrasos”, lo que evidencia que las obras de menor escala, comunes en las MiPymes constructoras, también son altamente vulnerables al incumplimiento de plazos.

La aplicación del estándar CRISP-DM permitió estructurar un proceso metodológico replicable que guía desde la comprensión de los datos hasta la evaluación de los modelos predictivos. Este marco demuestra que, con variables simples como la área construida, la duración y el precio por metro cuadrado, las MiPymes pueden generar herramientas de apoyo analítico que fortalecen la planificación y la toma de decisiones.

No obstante, al basarse en un conjunto de datos sintético, los resultados deben interpretarse como un ejercicio metodológico exploratorio. Aun así, el trabajo aporta una hoja de ruta inicial para que las MiPymes del sector de la construcción adopten la analítica predictiva como soporte para mitigar sobrecostos y anticipar desviaciones en las programaciones de obra.

Fondos: Esta investigación fue financiada por el Centro de Desarrollo Tecnológico Creativ, mediante la subvención número COD BPIN 2020000100538 del Ministerio de Ciencia, Tecnología e Innovación de Colombia.

Conflictos de interés: Los autores declaran no tener ningún conflicto de interés.

Referencias

- [1] A. Asadi, M. Alsubaey, and C. Makatsoris, “A machine learning approach for predicting delays in construction logistics,” *International Journal of Advanced Logistics*, vol. 4, no. 2, pp. 115–130, May 2015, doi: 10.1080/2287108x.2015.1059920.
- [2] A. Gondia, S. M. Asce, ; Ahmad Siam, ; Wael El-Dakhakhni, F. Asce, and A. H. Nassar, “Machine Learning Algorithms for Construction Projects Delay Risk Prediction,” 2019, doi: 10.1061/(ASCE).
- [3] S. Ouansrimeang and K. Wisaeang, “Analyzing the critical delay factors for construction projects in the public sector using relative importance index and machine learning techniques,” *Journal of Infrastructure, Policy and Development*, vol. 8, no. 8, 2024, doi: 10.24294/jipd.v8i8.6208.
- [4] BBVA Research, “Detrás del andamio: la estructura del sector de la construcción en Colombia,” Feb. 2024.
- [5] J. F. Gil Argüelles, “Analítica De Datos Para La Estandarización De La Productividad En Proyectos De Vivienda Vis En Colombia: Estudio De Los Históricos De Ejecución De Las Actividades,” Universidad de los Andes, Bogotá, 2022.
- [6] A. M. Correa González, “Factores que generan retrasos en la Entrega de Proyectos de Obra Civil en Choachí, Cundinamarca,” Corporación Universitaria Minuto de Dios, 2024.
- [7] J. P. Fitzsimmons, R. Lu, Y. Hong, and I. Brilakis, “CONSTRUCTION SCHEDULE RISK ANALYSIS - A HYBRID MACHINE LEARNING APPROACH,” *Journal of Information Technology in Construction*, vol. 27, pp. 70–93, 2022, doi: 10.36680/j.itcon.2022.004.

-
- [8] K. R. Erzaij, A. M. Burhan, W. A. Hatem, and R. H. Ali, "Prediction of the Delay in the Portfolio Construction Using Naïve Bayesian Classification Algorithms," *Civil and Environmental Engineering*, vol. 17, no. 2, pp. 673–680, Dec. 2021, doi: 10.2478/cee-2021-0066.
 - [9] M. O. Sanni-Anibire, R. M. Zin, and S. O. Olatunji, "Machine learning - Based framework for construction delay mitigation," Jul. 01, 2021, *International Council for Research and Innovation in Building and Construction*. doi: 10.36680/j.itcon.2021.017.
 - [10] Z. Zeng and Y. Gao, "Cost Control Management of Construction Projects Based on Fuzzy Logic and Auction Theory," *IEEE Access*, vol. 12, pp. 130292–130304, 2024, doi: 10.1109/ACCESS.2024.3438291.
 - [11] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.
 - [12] V. Hernández Guadalupe, "Construction Management Supervised IA.," 2022, *Kaggle*. doi: 10.34740/KAGGLE/DSV/4094678.
 - [13] S. Khursheed, S. Sharma, V. K. Paul, M. Al-Farouni, and C. Yoldosheva, "Forecasting Delays and Budget overruns resulting from Material-Related Factors," in *E3S Web of Conferences*, EDP Sciences, Aug. 2024. doi: 10.1051/e3sconf/202456302045.
 - [14] J. Batselier and M. Vanhoucke, "Construction and evaluation framework for a real-life project database," *International Journal of Project Management*, vol. 33, no. 3, pp. 697–710, Apr. 2015, doi: 10.1016/j.ijproman.2014.09.004.
 - [15] Z. H. Ali, A. M. Burhan, M. Kassim, and Z. Al-Khafaji, "Developing an Integrative Data Intelligence Model for Construction Cost Estimation," *Complexity*, vol. 2022, 2022, doi: 10.1155/2022/4285328.