

# Protocolo Tesis

## Título de Proyecto

*“Modelo predictivo de costos finales en proyectos de construcción de edificios residenciales en la Ciudad de Guatemala mediante técnicas de Machine Learning”*

## Descripción

El presente trabajo de investigación tiene como objetivo desarrollar un modelo predictivo de costos finales en proyectos de construcción de edificios residenciales en la Ciudad de Guatemala mediante la aplicación de técnicas de Machine Learning. La investigación surge de la necesidad de las empresas constructoras de contar con herramientas más precisas que les permitan anticipar el presupuesto real de un proyecto, considerando los riesgos de sobrecostos derivados de variaciones en materiales, mano de obra, tiempos de ejecución y otros factores.

Para la elaboración del modelo se utilizarán datos históricos provenientes de 20 proyectos desarrollados por la empresa, segmentados en etapas de construcción y con registros de cash-flow. Dichos datos incluirán información de costos planificados y reales, características del proyecto (tipo: condominio o edificio de apartamentos), dimensiones, duración, variaciones por materiales, y desviaciones por etapa.

El modelo predictivo buscará identificar patrones y variables determinantes que influyen en los costos finales, utilizando técnicas de regresión y algoritmos de aprendizaje supervisado. Como resultado, se espera generar una herramienta que permita estimar el costo final de nuevos proyectos con un mayor grado de precisión, apoyando la toma de decisiones estratégicas en la planeación y gestión financiera de la empresa.

La investigación, no obstante, presenta limitaciones derivadas del tamaño de la muestra (20 proyectos), así como la posible falta de estandarización en los registros históricos. Pese a ello, se plantea que este proyecto sirva como un estudio piloto que sienta las bases para la integración de modelos predictivos más robustos en el sector de la construcción en Guatemala.

## Objetivos

### Objetivo Principal

Desarrollar un modelo predictivo basado en técnicas de Machine Learning para estimar con precisión los costos finales de proyectos de construcción de edificios residenciales en la Ciudad de Guatemala a partir de datos históricos.

### Objetivos Específicos

- Identificar y caracterizar los factores principales que influyen en la variación de costos en proyectos de edificios residenciales en la Ciudad de Guatemala.
- Analizar patrones históricos de desviaciones de costos entre presupuestos iniciales y costos finales en el sector residencial local.
- Diseñar y entrenar modelos de Machine Learning para predecir los costos finales de los proyectos.
- Evaluar y comparar diferentes algoritmos de Machine Learning para seleccionar el más efectivo comparando la precisión del modelo mediante métricas estadísticas apropiadas.
- Realizar pruebas de generalización del modelo con datos no utilizados en el entrenamiento. Analizando la sensibilidad del modelo a cambios en las variables de entrada.
- Desarrollar una herramienta o aplicación que permita la implementación práctica del modelo por parte de profesionales del sector.

## Antecedentes

“TRABAJO FIN DE CARRERA Big Data. Técnicas de machine learning para la creación de modelos predictivos para empresas”. Fuente: Repositorio de la Universidad Pontificia Comillas. [\[Enlace\]](#)

“Inteligencia artificial en la gestión de proyectos: caso construcción y obra civil”. Fuente: European Public & Social Innovation Review. [\[Enlace\]](#)

“Índice de Precios de Materiales de Construcción (IPMC) del INE: serie mensual de 89 insumos; hay histórico y reportes 2024–2025”. Fuente: Instituto Nacional de Estadística de Guatemala. [\[Enlace\]](#)

## **Alcance de Proyecto**

### **Fuente de datos**

Recolección de datos de proyectos de construcción de edificios residenciales en la Ciudad de Guatemala. Estos datos incluirán variables como el área de construcción, la ubicación, tiempo de ejecución, presupuestos, costos reales, desviaciones, costos de mano de obra, costos de los subcontratistas, entre otros. Estas variables se determinarán después del análisis sobre los datos. Por otro lado, mediante la construcción de un dataset maestro e integrar la información histórica Incluyendo para hacerla más manejable para el modelo.

### **Modelo de Machine Learning**

Desarrollo y la aplicación de uno o varios modelos de Machine Learning. El objetivo es que estos modelos puedan predecir los costos finales basándose en las variables de entrada. Probablemente se consideren técnicas como la regresión lineal múltiple, árboles de decisión, Random Forest o Gradient Boosting.

### **Evaluación de Modelo**

Evaluar el rendimiento del modelo para determinar qué tan preciso es en sus predicciones. Esto podría incluir métricas como el error cuadrático medio (RMSE) o el error absoluto medio (MAE). También se detectarán cuáles variables son las que más influyen dentro de los costos del proyecto.

## **Límites de Proyecto**

### **Disponibilidad y calidad de los datos**

La recolección de los datos de proyectos históricos puede estar incompleta o inconsistentes, esto podría afectar la precisión del modelo. Otro aspecto para tomar en cuenta es el tamaño de la muestra, el modelo tendrá capacidad limitada de generalización por lo que se necesitarían más elementos para realizar un modelo robusto y estable.

## **Factores externos**

El modelo podría no considerar factores que son difíciles de cuantificar, como los cambios en las regulaciones gubernamentales, las fluctuaciones imprevistas en los precios de los materiales a nivel global o los eventos socioeconómicos inesperados (como una pandemia o una crisis económica) que impactan directamente en los costos.

## **Geográfico**

Los resultados aplican solo a la empresa y al contexto de la Ciudad de Guatemala, por lo que, los resultados no necesariamente aplicables a otros departamentos de Guatemala o países con diferentes condiciones económicas y regulatorias.

## **Contenido Esperado**

### **Capítulo 1. Introducción**

- 1.1 Planteamiento del problema
- 1.2 Formulación del problema
- 1.3 Justificación del estudio
- 1.4 Objetivo general
- 1.5 Objetivos específicos
- 1.6 Alcances y limitaciones

### **Capítulo 2. Marco Teórico**

- 2.1 Conceptos de costos en la construcción
  - 2.1.1 Costos directos
  - 2.1.2 Costos indirectos
  - 2.1.3 Sobrecostos y desviaciones
- 2.2 Gestión de proyectos de construcción
  - 2.2.1 Etapas del ciclo de vida de un proyecto
  - 2.2.2 Condominios vs. edificios de apartamentos
- 2.3 Cash-flow y análisis financiero en la construcción
  - 2.3.1 Presupuestos iniciales
  - 2.3.2 Control de ejecución
- 2.4 Introducción al Machine Learning
  - 2.4.1 Tipos de aprendizaje supervisado
  - 2.4.2 Modelos aplicados en predicciones de costos (Regresión, Random Forest, XGBoost, Redes Neuronales)
- 2.5 Estado del arte

- 2.5.1 Modelos predictivos en construcción a nivel internacional
- 2.5.2 Estudios previos en Latinoamérica
- 2.5.3 Brechas de investigación en Guatemala

### **Capítulo 3. Metodología**

- 3.1 Tipo de investigación (aplicada, cuantitativa, predictiva)
- 3.2 Diseño metodológico
  - 3.2.1 Recolección de datos
  - 3.2.2 Variables seleccionadas (costos iniciales, etapas de construcción, variaciones, tiempos, materiales)
  - 3.2.3 Procesamiento de datos (limpieza, normalización, codificación)
- 3.3 Desarrollo del modelo predictivo
  - 3.3.1 Selección de algoritmos (ej. regresión múltiple, Random Forest, XGBoost)
  - 3.3.2 Entrenamiento y validación del modelo
  - 3.3.3 Métricas de evaluación (RMSE, MAE,  $R^2$ )
- 3.4 Herramientas tecnológicas de desarrollo
- 3.5 Validación de resultados
  - 3.5.1 Comparación contra costos finales reales
  - 3.5.2 Discusión sobre desviaciones

### **Capítulo 4. Resultados y Análisis**

- 4.1 Análisis descriptivo de los datos
  - 4.1.1 Distribución de proyectos por tipo (condominio / edificio)
  - 4.1.2 Distribución de desviaciones de costos
- 4.2 Resultados de los modelos predictivos
  - 4.2.1 Métricas obtenidas
  - 4.2.2 Comparación entre algoritmos
- 4.3 Discusión de resultados
  - 4.3.1 Factores más influyentes en las desviaciones

### **Capítulo 5. Conclusiones y Recomendaciones**

- 5.1 Conclusiones
- 5.2 Recomendaciones prácticas para empresas constructoras
- 5.3 Recomendaciones para futuras investigaciones

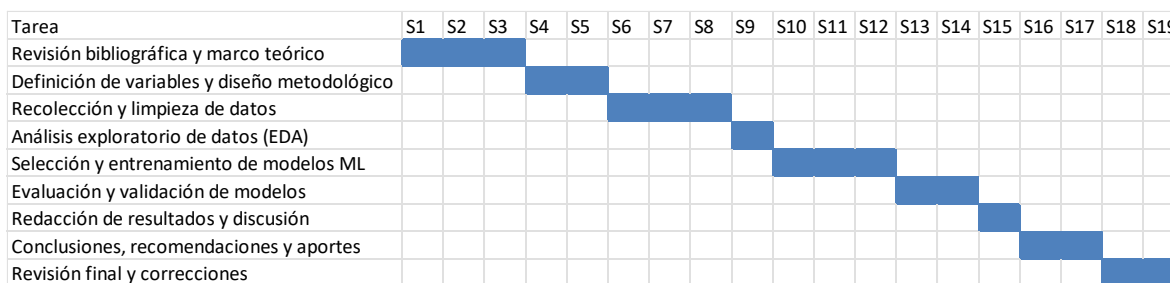
## Asesor sugerido

Ing. Juan Carlos López

## Planificación General

Inicio	Fin	Semanas	Días	Tarea
2025-08-25	2025-09-15	3	21	Revisión bibliográfica y marco teórico
2025-09-15	2025-09-29	2	14	Definición de variables y diseño metodológico
2025-09-29	2025-10-20	3	21	Recolección y limpieza de datos
2025-10-20	2025-10-27	1	7	Análisis exploratorio de datos (EDA)
2025-10-27	2025-11-17	3	21	Selección y entrenamiento de modelos ML
2025-11-17	2025-12-01	2	14	Evaluación y validación de modelos
2025-12-01	2025-12-08	1	7	Redacción de resultados y discusión
2025-12-08	2025-12-22	2	14	Conclusiones, recomendaciones y aportes
2025-12-22	2026-01-05	2	14	Revisión final y correcciones

## Diagrama de Gantt



## **Bibliografía**

- Ashworth, A., & Perera, S. (2019). *Cost studies of buildings* (6th ed.). Routledge. <https://doi.org/10.4324/9780429276508>
- Hendrickson, C. (2008). *Project management for construction: Fundamental concepts for owners, engineers, architects, and builders*. Prentice Hall.
- Cheng, J. C., Law, K. H., & Pan, W. (2020). Predicting construction project cost performance: A machine learning approach. *Automation in Construction*, 118, 103278. <https://doi.org/10.1016/j.autcon.2020.103278>
- Sonmez, R., & Ontepeli, B. (2009). Predesign cost estimation of urban railway projects with parametric modeling. *Journal of Civil Engineering and Management*, 15(4), 405–409. <https://doi.org/10.3846/1392-3730.2009.15.405-409>
- Kim, H., & Park, M. (2020). Machine learning approach to predicting construction cost overruns. *Journal of Construction Engineering and Management*, 146(2), 04019097. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0001730](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001730)
- Flyvbjerg, B. (2017). *The Oxford handbook of megaproject management*. Oxford University Press.
- Baccarini, D. (1996). The concept of project complexity—a review. *International Journal of Project Management*, 14(4), 201–204. [https://doi.org/10.1016/0263-7863\(95\)00093-3](https://doi.org/10.1016/0263-7863(95)00093-3)
- Cámara Guatemalteca de la Construcción. (2022). *Informe anual de la construcción en Guatemala*. Guatemala: CGC.
- Instituto Nacional de Estadística de Guatemala. (2021). *Índice de precios de materiales de construcción*. Guatemala: INE.

Guatemala, 21 de agosto de 2025

Ingeniero  
**Carlos Alfredo Azurdia**  
**Coordinador de Privados y Trabajos de Tesis**  
**Escuela de Ingeniería en Ciencias y Sistemas**  
**Facultad de Ingeniería - USAC**

Respetable Ingeniero Azurdia:

Por este medio hago de su conocimiento que acepto asesorar el trabajo de investigación del estudiante **LEONARDO RONEY MARTÍNEZ MALDONADO** con carné **201780044 y NO PASAPORTE E0169262** titulado **“MODELO PREDICTIVO DE COSTOS FINALES EN PROYECTOS DE CONSTRUCCIÓN DE EDIFICIOS RESIDENCIALES EN LA CIUDAD DE GUATEMALA MEDIANTE TÉCNICAS DE MACHINE LEARNING”**, así mismo me comprometo a ejercer mis funciones ajustándome a las fechas que en conjunto con el estudiante hemos definido en el plan de trabajo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,



**Ing. Juan Carlos López Huertas**  
Colegiado No. 17063