

Informe 3: Metodología de Actualización de Precios



Noviembre 28 de 2024

Barranquilla, noviembre 28 de 2024

Señores (as),
Comisión de Regulación de Energía y Gas - CREG
Atn. Dra. Denice Romero López
Asesora - Regulación
Bogotá D.C.

Referencia: Contrato de Consultoría N° 2024-112 ISES SAS

Asunto: Presentación del informe denominado “Metodología de Actualización de Precios”

Estimada Dra. Denice,

Formalizamos ante usted la presentación del informe denominado “Metodología de Actualización de Precios” dentro del marco del desarrollo del contrato número 2024-112 ISES SAS, cuyo objeto es *“Consultoría para identificar las nuevas tecnologías que se estén utilizando para hacer más eficiente el sistema de transmisión, incluyendo activos que aumenten la capacidad de los existentes, así como analizar la variación de precios de las unidades constructivas a considerar en la metodología de transmisión y proponer la forma de actualizarlos”*.

Cordialmente,


C.C 73.144.071

HARRY WILLIAM RAMOS

Gestión Operativa y Administrativa del Contrato
Ingeniería y soluciones Especializadas – ISES S.A.S.
Correo: hramos@ises.com.co
Móvil: 3126024306

CONTENIDO

	Pág.
1. Introducción.....	3
2. Objeto.....	3
3. Alcances y limitaciones	4
4. Antecedentes	4
5. Estado del arte	5
5.1. Preprocesamiento de los datos.....	5
5.2. Selección de variables	7
5.3. Técnicas de modelado de series de tiempo comúnmente utilizadas	8
5.4. Proceso de entrenamiento	11
5.5. Comparación de las métricas de desempeño para el modelado de series de tiempo	12
6. Metodología Propuesta para la Actualización de Costos de Unidad Constructiva	14
6.1. Definición	14
6.2. Diseño funcional de la solución propuesta.....	16
6.3. Preprocesamiento de datos	16
6.4. Análisis factorial	17
6.5. Modelos basados en inteligencia artificial	18
6.6. Selección adecuada de una arquitectura de red neuronal	20
6.7. Proceso de entrenamiento	21
6.8. Métricas utilizadas para la evaluación del desempeño de los modelos	22
7. Implementación	23
7.1. Interfaz de actualización de costos de las unidades constructivas	23
7.2. Resultados	24
8. Conclusiones.....	36
9. Instalación e implementación de la aplicación para la actualización de costos de UC.....	37
10. Referencias	38

1. Introducción

Este documento tiene como objetivo específico desarrollar una metodología para estimar el valor de las Unidades Constructivas UC, considerando de manera integral sus componentes principales: obra civil, montaje, otros costos, costos financieros y DDP, definidos en la regulación vigente. Esta metodología tiene como propósito establecer un marco analítico robusto que facilite la actualización con el mayor ajuste a los precios comerciales de las Unidades Constructivas, asegurando que las estimaciones reflejen adecuadamente los costos reales y se ajusten a las fluctuaciones del mercado. Además, este enfoque pretende capturar la complejidad y variabilidad inherente a cada uno de los componentes, garantizando así una representación precisa de las condiciones dinámicas del sector energético, incluyendo aspectos como las tendencias tecnológicas y el impacto de factores macroeconómicos.

La metodología propuesta permitirá, en primer lugar, identificar las variables significativas que afectan cada uno de los componentes mencionados, lo cual es fundamental para entender los factores que inciden en los costos y para elaborar estrategias que permitan optimizar dichos componentes. Esto incluye no solo el análisis de los costos directos, sino también la consideración de factores externos, como la variabilidad de los precios de materiales, las condiciones macroeconómicas y la influencia de normativas y regulaciones vigentes. En segundo lugar, se pretende estimar valores futuros de los costos, lo cual proporcionará a los actores del sector una herramienta predictiva que facilite la planificación estratégica y la toma de decisiones informadas en torno a inversiones y proyectos de infraestructura eléctrica. Esta estimación de valores futuros se basará en modelos econométricos y escenarios prospectivos que tengan en cuenta las tendencias del mercado y los posibles cambios en las condiciones económicas y regulatorias.

Adicionalmente, la metodología tiene el objetivo de proporcionar resultados interpretables y comprensibles, que puedan ser utilizados por diferentes partes interesadas, tales como entidades reguladoras, desarrolladores y operadores del sistema. Esto permitirá que los resultados de la valoración sean útiles no solo para el análisis técnico, sino también para la formulación de políticas y estrategias que contribuyan al desarrollo eficiente y sostenible del sistema eléctrico nacional. Asimismo, pretende ser lo suficientemente flexible para adaptarse a diferentes contextos y escenarios, permitiendo su actualización continua a medida que surjan nuevos desarrollos tecnológicos y cambios en las condiciones del mercado.

2. Objeto

Desarrollar e implementar una metodología de análisis y cálculo que permita actualizar los costos de las Unidades Constructivas (UC) objeto de estudio, asegurando que estos reflejen las condiciones actuales del mercado y respondan adecuadamente a las necesidades del Sistema de Transmisión Nacional (STN) en Colombia

3. Alcances y limitaciones

El alcance de este estudio es desarrollar una herramienta tecnológica que permita actualizar los precios de las Unidades Constructivas (UC) de transmisión de energía eléctrica de forma sistemática y eficiente. Esta herramienta se basará en técnicas de inteligencia artificial y análisis estadístico multivariado, con el fin de automatizar el proceso y mejorar la precisión de las estimaciones.

La herramienta tendrá las siguientes funcionalidades:

- Identificación automática de variables significativas: Mediante la aplicación de técnicas estadísticas de análisis multivariado, la herramienta identificará las variables que tienen un mayor impacto en el costo de las UC.
- Estimación del costo de las UC: Utilizando técnicas de inteligencia artificial, como redes neuronales y/o *random forest*, la herramienta estimará el costo de las UC con base en las variables significativas identificadas.

Limitaciones:

Para el desarrollo y la aplicación de la herramienta, se requiere:

- Datos históricos de los componentes de las UC: Se necesita una base de datos con información histórica de los precios de los componentes de las UC.
- Variables económicas asociadas: Se requieren datos sobre variables económicas que puedan afectar el costo de las UC, como la inflación, el tipo de cambio, los índices de precios de materiales, entre otros.
- Datos completos y consistentes: La calidad de los datos es fundamental para el correcto funcionamiento de la herramienta. Se requieren datos completos, consistentes y libres de errores.

El desarrollo de esta herramienta tecnológica busca automatizar y optimizar el proceso de actualización de precios de las UC, contribuyendo a la eficiencia y la transparencia en la remuneración de la transmisión de energía eléctrica en Colombia

4. Antecedentes

La valoración de las Unidades Constructivas (UC) en el sector eléctrico se realiza actualmente mediante un procedimiento que emplea el Índice de Precios al Productor (IPP) como única variable de ajuste. Este método consiste en actualizar el valor base de la UC del año anterior según la variación del IPP, para así estimar el valor correspondiente al año siguiente. Sin embargo, los análisis históricos han demostrado que la dependencia exclusiva de una única variable como el IPP no es suficiente para capturar la complejidad de los factores que influyen en el costo real de las UC, lo cual genera desviaciones considerables entre los valores estimados y los valores observados.

Adicionalmente, las metodologías de modelado de costos de las UC no contemplan adecuadamente la inclusión de múltiples variables significativas que impactan directamente el costo de estos elementos. Esta situación evidencia la necesidad de desarrollar un enfoque metodológico más avanzado, que permita modelar los costos de las UC considerando variables como la variación de los commodities asociados con precio de los materiales, otros indicadores macroeconómicos, y el uso de técnicas de inteligencia artificial. Este enfoque integrado facilitaría la obtención de valores actualizados y consistentes, mejorando la precisión y confiabilidad de las estimaciones de costos en el sector eléctrico.

5. Estado del arte

5.1. Preprocesamiento de los datos

El modelado de series de tiempo requiere la fidelidad y representabilidad de los datos disponibles. Las alteraciones en el comportamiento natural de las series temporales pueden inducir ruido o errores en el proceso de construcción del modelado matemático. Muchos autores coinciden y demuestran que esta etapa es importante para el proceso de selección y montaje de un modelo confiable. Los datos de las series de tiempo pueden presentar lagunas, inconsistencias y tendencias con comportamiento exponencial. Se mostrarán las acciones más comúnmente implementadas por los autores para limpiar los datos de series temporales con el fin de mejorar el rendimiento de los modelos obtenidos: Un método para el modelado de series de tiempo econométricas ha sido propuesto en el trabajo propuesto por (S. Liu, Gu, & Bao, 2017).

Se han definido cuatro etapas necesarias para construir cada uno de los modelos matemáticos ARIMA basados en la metodología Box-Jenkins. En primera instancia, para detectar y eliminar errores e inconsistencias para mejorar la calidad de los datos, se agrega el preprocesamiento de datos. Además, todas las lagunas en los datos se identifican y se completan con el promedio de los datos más cercanos. En segunda instancia, la detección y corrección de los datos atípicos se integran como un intervalo de confianza definido por $[\mu - t * \sigma, \mu + t * \sigma]$, donde μ : promedio, t : umbral, σ : desviación estándar. En la última etapa se presentan aquellos casos en los que se muestra tendencia exponencial y heterocedasticidad, lo que hace necesario realizar una transformación logarítmica.

Los autores del trabajo presentado por (Oniscu et al., 2009) realizan una modelización de dos series temporales relacionadas con la precipitación en los ríos Daning y Zhenshui. Para llevar a cabo el acondicionamiento de datos antes de construir los modelos, se proponen y comparan cuatro tipos de técnicas: 1) Moving-Average Model (MA), 2) Principal Components Analysis (PCA), 3) Singular Spectrum Analysis (SSA), 4) Diagonal Averaging (DA). Esta etapa de preprocesamiento se integra a un modelo basado en redes neuronales. El promedio diagonal consiste en encontrar elementos iguales en la matriz resultante y luego generar un nuevo elemento promediando sobre ellos. Los resultados obtenidos mostraron lo siguiente:

- El efecto de la técnica MA en la realización de la previsión no es perceptible.
- La aplicación de PCA para la eliminación de ruido produce una ligera mejora en el rendimiento del modelo.

- SSA produce una mejora considerable en el rendimiento del modelo de previsión.

En el trabajo propuesto por (Zhang & Kline, 2007) surge una alternativa basada en redes neuronales y dos técnicas de preprocesamiento de datos para el modelado trimestral de series de tiempo considerando 48 posibles modelos neuronales con diferentes datos históricos, variables climáticas, variables trigonométricas e indicadores de tiempo como entradas. Se realizan análisis paramétricos y no paramétricos para evaluar y seleccionar el mejor modelo para cada caso. En este trabajo se proponen dos estrategias de transformación: 1) “log”: consiste en aplicar una transformación logarítmica a los datos y 2) “full”: consiste en aplicar la transformación logarítmica junto con una media móvil. Para eliminar la tendencia de la serie de datos, se realiza una aproximación a una tendencia lineal y este componente se resta de los datos. Para eliminar el componente estacional se aplica una media móvil siguiendo un modelo clásico de descomposición aditiva. Los parámetros para descomponer la serie temporal se obtienen a partir de los datos de la propia serie temporal. Otro método para ajustar la estacionalidad de la serie es el propuesto por (Atiya et al., 1999).

Algunos autores incluyen transformaciones más sofisticadas en el acondicionamiento y preprocesamiento de datos, como la propuesta por (Yu et al., 2020) en la que utilizan la Transformada Wavelet Empírica mejorada con el algoritmo Artificial Bee Colonial (ABC). EWT es un algoritmo que integra la formalización del EWT junto con la adaptabilidad de la descomposición modal empírica (EMD). En este trabajo, se utiliza EWT para eliminar el impacto de los valores atípicos, haciendo que los datos sean más adecuados para los modelos de pronóstico ARIMA y Extreme Learning Model (ELM). Para probar su propuesta, los autores utilizan tres conjuntos de datos del sector financiero. Se realiza una comparación del rendimiento (MAPE) de los modelos (LSTM, ARIMA, ANN y el modelo híbrido propuesto). Se muestra una disminución en el MAPE debido al preprocesamiento de los datos. El MAPE más bajo obtenido por la propuesta aquí planteada fue del 0,61% del modelo propuesto.

La implementación de modelos ARIMA en series de tiempo univariadas está limitada por aquellas con características estacionarias, los autores (S. Liu, Gu, & Peng, 2017) proponen una transformación logarítmica para obtener las características propias de las series de tiempo con el fin de construir un modelo de pronóstico de manera óptima. Se implementa un conjunto de pruebas estadísticas para resolver el problema de la identificación imprecisa e inadecuada de las características no estacionarias de la serie temporal. La propuesta está compuesta por las siguientes etapas:

- 1) Mejora de la calidad de los datos. En esta etapa, el reemplazo de los datos faltantes o atípicos se realiza utilizando datos promedio anteriores y/o tardíos.
- 2) Estabilización de la varianza. En esta etapa, se aplica el logaritmo natural a los datos cuando la serie temporal es heterocedástica.
- 3) Aplicación de la prueba unitaria de raíces para la identificación de la periodicidad.
- 4) Algoritmo de autoadaptación y procesamiento de series temporales no estacionarias.

Se utilizaron datos de prueba relacionados con la generación mensual de energía eléctrica en China desde enero de 1990 hasta diciembre de 2015. Los resultados obtenidos mostraron un

mejor desempeño del modelo propuesto con un MAPE de 2.5% respecto al 4.45% del modelo Holt-Winters.

5.2. Selección de variables

Uno de los aspectos más importantes en el modelado de series de tiempo es la selección de las variables a considerar no sólo por su relevancia en la realización de la previsión, sino que también por su incidencia en la reducción de los costos asociados a la disponibilidad de las diferentes variables a ser consideradas. Dependiendo de la interacción entre las variables y las estructuras de los modelos, es posible establecer dos enfoques para la selección de variables: model-free (no depende ni de la estructura ni de la calibración del modelo) y model-based (depende de la estructura del modelo) (Alves da Silva et al., 2008; Fernando et al., 2009). En el model-free, el coeficiente de Pearson se utiliza para determinar la existencia de correlación entre la entrada y la salida del modelo. Por otro lado, la información mutua parcial (PMI) aporta información sobre la existencia de una correlación lineal o no lineal entre los datos de entrada y salida del modelo. El model-based considera que las variables de entrada de los candidatos son significativas si contribuyen a uno o más indicadores de desempeño del modelado. Los indicadores de desempeño comúnmente utilizados son el error cuadrático medio entre los datos observados y pronosticados y el coeficiente de determinación (R^2).

En el trabajo presentado por (Tran et al., 2015) se explican tres métodos para la selección de las variables de entrada más significativas para cuatro series de datos sintéticos y dos series de datos reales. Las técnicas utilizadas para realizar el análisis significativo son: 1) correlación lineal parcial (PLC), 2) información mutua parcial (PMI), y 3) técnica basada en programación genética (GP). La correlación lineal parcial proporciona información sobre cómo se pueden analizar dos variables covariantes de forma lineal en un diagrama de dispersión. La técnica de información mutua evalúa la dependencia entre dos variables a partir de su distribución conjunta y densidad marginal.

Es decir, la información mutua parcial mide la disminución de la incertidumbre de una de las variables como el reconocimiento. Finalmente, está la técnica basada en programación genética que comparte características similares al algoritmo genético, pero éste se basa en la generación y variación de las estructuras del árbol. En este método, una variable de entrada se considera significativa si su número de apariciones en las ecuaciones generadas es mayor que el umbral establecido. Los resultados mostraron que el modelo PLC (debido a su bajo costo computacional) y el modelo GP (debido a su capacidad para detectar una relación no lineal) son recomendados para la selección de las variables.

En el sector salud, el modelado de series de tiempo contribuye al proceso de mejora y optimización. Los autores en (Rasouli et al., 2016) proponen un modelo basado en una red neuronal para determinar la tasa de hospitalización. Para la selección de las variables se ha tomado como referencia los métodos de Box-Jenkins para procesos autorregresivos, en los que se utiliza la tabla de funciones autorregresivas parciales para determinar cuánta información del pasado es necesaria para poder explicar el futuro. La función de autocorrelación parcial ayuda a identificar el orden más alto de un proceso autorregresivo. Este análisis se realiza en cuatro

hospitales obteniendo un MAPE (Error Porcentual Absoluto Medio) mínimo de 1,17% y máximo de 3,95%.

La selección de las variables es un proceso fundamental dentro de la realización de modelos de predicción. En algunas series de tiempo, por ejemplo, en la demanda de energía, es posible encontrar muchas variables que afectan el comportamiento del consumo de energía. Los autores de (Xie & Hong, 2018) toman dos soluciones alternativas que ayudan con la selección de variables con el objetivo de aumentar la eficiencia en la cantidad de variables a utilizar. Los datos de energía y temperatura se utilizan para siete estados de EE. UU. de 2007 a 2011 y de 2001 a 2010, respectivamente. El primer método es holístico y se basa en la selección de la medida de probabilidad de error que sea consistente con la medida de error utilizada para la evaluación final del pronóstico probabilístico. El segundo método se basa en la estrategia heurística que toma un atajo confiando en un error de medición en la selección de variables. Los resultados mostraron que, aunque el método holístico supera ligeramente al método heurístico, no domina. Se obtiene un MAPE mínimo de 2,82% y máximo de 3,19%.

5.3. Técnicas de modelado de series de tiempo comúnmente utilizadas

El modelado de series de tiempo requiere de un conjunto de acciones que permitan depurar, caracterizar, seleccionar y simplificar los datos necesarios para la construcción de los diferentes modelos. Dado que no existe una técnica única que garantice el mejor rendimiento, muchos autores han presentado diferentes trabajos en los que se realiza la comparación entre las diferentes técnicas de modelado (e.g. ARIMA, ES, ANN, SVM, LSTM, RTE, entre otras). Es decir, se han utilizado modelos híbridos con el fin de aprovechar las fortalezas y mitigar las debilidades de las diferentes técnicas. A continuación, se mostrará un resumen de los trabajos relacionados con estos temas:

Las redes neuronales también se han explorado mediante enfoques univariados, como el presentado por (Oğcu et al., 2012) en el que se utilizan redes neuronales y una extensión de máquinas de vectores de soporte para la regresión no lineal. Para el entrenamiento de la red neuronal propuesta y el ajuste de regresión se utiliza una colección de datos de consumo que abarca un período de cuarenta y un (41) años. Los datos de entrada consisten en variables derivadas de los datos de consumo, estas son: tiempo, mes e índice de estacionalidad, datos de consumo con uno o dos años de anticipación. Los aspectos más importantes a considerar dentro de lo presentado por (Oğcu et al., 2012). Los resultados (MAPE) mostraron que el rendimiento del modelo SVR (3,3%) superó al mostrado por las redes neuronales (3,9%).

El proceso de valoración de unidades constructivas y el enfoque de precisión entre lo que se estima y su valor real ha llevado a realizar estudios sobre el comportamiento de los diferentes componentes que conforman estas unidades. El estado del arte segrega el análisis de los modelos de acuerdo con los alcances y limitaciones intrínsecas al planteamiento y complejidad de las variables a correlacionar

El análisis de las series temporales de valores de unidades constructivas puede resultar complejo cuando se recurre a métodos convencionales de estimación que basan su principio en establecer

una relación a priori entre la variable independiente y las variables explicativas. Este desafío no es único del sector de valoración de activos eléctricos; por ejemplo, en el campo de la estimación de demanda de energía eléctrica, estudios como el propuesto por (Olaya, 2007a) demostraron las limitaciones de utilizar métodos simplificados. En dicho estudio, se evolucionó de considerar únicamente la hora del día como variable explicativa para la demanda, a incorporar variables macroeconómicas como el PIB, logrando mejoras significativas en la precisión de las estimaciones. Este caso resulta particularmente relevante para nuestro contexto actual, donde la valoración de unidades constructivas enfrenta una situación análoga: la dependencia de una única variable resulta insuficiente para capturar la complejidad total del fenómeno.

Dentro de los aportes más relevantes que podemos extraer de estas experiencias se destacan:

- Así como se estableció una fuerte relación entre la demanda eléctrica y variables macroeconómicas como el PIB, es razonable esperar que el valor de las unidades constructivas esté igualmente influenciado por múltiples variables económicas además del IPP.
- De la misma manera que resultó suficiente utilizar información reciente para estimaciones de demanda, se puede inferir que un histórico reciente de valores de unidades constructivas, junto con variables económicas actuales, proporcionaría una base sólida para las estimaciones.
- La experiencia en el sector eléctrico ha demostrado las limitaciones de los modelos univariados, lo que refuerza la necesidad de evolucionar más allá del uso exclusivo del IPP para la valoración de unidades constructivas.
- Los modelos multivariados han presentado consistentemente mejor rendimiento en diversos campos del sector eléctrico, sugiriendo un camino similar para la valoración de activos.

Establecer la idoneidad de un tipo de modelo no es una tarea sencilla, de hecho, aún en la actualidad resulta complicado indicar tácitamente cuál de todos es el mejor. En el estado del arte es posible encontrar aproximaciones que buscan comparar y analizar el rendimiento de algunos modelos comúnmente utilizados como redes neuronales y técnicas de machine. Sin embargo, es importante resaltar que la base de rendimiento es únicamente establecida para unas condiciones de validación en particular, lo que hace necesario un análisis cuidadoso para seleccionar la metodología más apropiada para el caso específico de valoración de unidades constructivas.

Los modelos de estimación basados en una única variable se caracterizan por presentar un rendimiento limitado cuando se busca predecir el valor de unidades constructivas a mediano y largo plazo, mientras que los métodos multivariados se convierten en una mejor opción. Como referencia, en el campo de estimación de demanda eléctrica, el trabajo presentado por (Taylor et al., 2006) realizó un análisis comparativo de seis tipos de modelos de predicción, incluyendo métodos univariados sencillos y métodos más sofisticados como el análisis de componentes principales (PCA). Los resultados de este estudio son particularmente relevantes para nuestro contexto, donde se destacan:

- La superioridad de los métodos que incorporan múltiples variables sobre los métodos simples univariados.
- Los métodos de regresión no paramétricos demostraron ser una opción viable al balancear simplicidad y precisión debido al número reducido de parámetros involucrados.
- El método simple de referencia fue significativamente superado por métodos más sofisticados.
- Se logró una mejora sustancial en la precisión cuando se incorporaron técnicas de corrección de errores, sugiriendo la importancia de incluir mecanismos de ajuste en los modelos de valoración.
- Se utilizaron datos de prueba relacionados con la generación mensual de energía eléctrica en China desde enero de 1990 hasta diciembre de 2015. Los resultados obtenidos mostraron un mejor desempeño del modelo propuesto con un MAPE de 2.5% respecto al 4.45% del modelo Holt-Winters.

La siguiente tabla muestra un resumen de las técnicas usualmente utilizadas en el modelado de series de tiempo:

Tabla 1. Resumen de las técnicas usualmente utilizadas en el modelado de series de tiempo

Autor	Técnica de modelado					
	LSTM	MLP	ARIMA	SVM	ANFIS	FL
(F. Liu et al., 2020)	✓	✓	✓	✓		
(Aldhyani et al., 2020)	✓			✓	✓	
(Bauer et al., 2020)		✓	✓	✓		
(Yang & Yang, 2020)	✓					
(Zheng et al., 2020)	✓	✓		✓		
(Jiang et al., 2019)	✓		✓			
(Godfrey & Gashler, 2017)	✓		✓	✓		
(Jaramillo et al., 2017)				✓		
(Moreira de Andrade & Nunes da Silva, 2009)		✓	✓		✓	
(Huang et al., 2003)		✓	✓			
(Azadeh et al., 2008)		✓	✓			
(Senjyu et al., 2002)		✓				
(Oğcu et al., 2012)				✓		
(Huang et al., 2003)		✓	✓			
(Beccali et al., 2004)		✓				
(Fan & Chen, 2006)				✓		
(Pai & Hong, 2005)		✓	✓	✓		
(Yalcinoz & Eminoglu, 2005)		✓				
(Dalvand et al., 2011)			✓			
(Olaya, 2007b)			✓			

Autor	Técnica de modelado					
	LSTM	MLP	ARIMA	SVM	ANFIS	FL
(Song et al., 2005)						✓
(Pacheco et al., 2004)		✓			✓	
(Darbellay & Slama, 2000)		✓	✓			
(OLANREWAJU, 2011)		✓				
(Gabriel I. S. Ruas, 2008)		✓		✓		
(Barbounis et al., 2006; Park et al., 1991)		✓				
(G et al., 2006)						✓
(Nagasaka & Mamun, 2004)		✓				
(Miguel et al., 2007)		✓				
(Mizutani et al., 2005)		✓				
(R.-H. Liang & Cheng, 2000)		✓				

5.4. Proceso de entrenamiento

El modelo de red neuronal propuesto por (Bataneh & Marler, 2017) se centró en reducir la cantidad de datos de entrada necesarios en el proceso de entrenamiento. Se aplica una red neuronal de base radial a problemas de regresión; La nueva estructura consiste en un proceso de entrenamiento de varias etapas en el que se combinan los mínimos cuadrados ortogonales (OLS) con un gradiente de optimización. La validación del modelo se realiza sobre datos y pronóstico de estrés y fuerza en la rodilla para la prevención de lesiones en diferentes escenarios (tarea) en los que los resultados y tiempos estimados se obtienen con otros modelos como el de mínimos cuadrados ortogonales (OLS). y avance. La red de retropropagación (FFN) se compara con sus valores reales. La estructura propuesta denominada Opt-RBN mostró, en algunos escenarios, resultados favorables en los datos de entrenamiento, aunque el tiempo del proceso de entrenamiento es un poco mayor, los autores lo consideraron un buen tiempo porque es inferior a un minuto.

El modelo Back Propagation Neural Network (BPNN) se utiliza en el proceso de entrenamiento para determinar el número de neuronas necesarias en las dos capas ocultas de una red neuronal para pronosticar la magnitud en la escala de Richter de los terremotos en una región del Mar de Filipinas (Lin et al., 2018). Con el uso de datos de registro de los terremotos en la región de 1990 a 2014, se construyen varias series de modelos BPNN para pronosticar la magnitud del terremoto y compararla con los datos reales. A partir de los datos, se concluye que un número de 10 neuronas por capa oculta es el número ideal para el modelo de pronóstico. Dado que, los errores de pronóstico del modelo BPNN con 10 neuronas en cada capa oculta son muy similares a los modelos que usan más de 10 neuronas por capa, lo que provoca un mayor tiempo de cálculo para resultados similares. Los resultados se comparan con los valores reales de las magnitudes de los terremotos que ya han ocurrido y los autores concluyeron que el modelo BPNN pronosticó un resultado confiable de la magnitud del terremoto en la escala de Richter.

Para resolver el problema de categorización de pruebas, se utiliza el modelo de red neuronal para la categorización. Uno de los modelos propuestos es la red neuronal de retropropagación mejorada (IBPNN) de (Li, C. et al., 2014) con un proceso computacional paralelo, acelera el entrenamiento de la red neuronal para la categorización de texto. El algoritmo BPNN utiliza un Sun Cluster con 34 nodos (procesadores). El IBPNN paralelo está integrado con la técnica SVD (descomposición de valores singulares), donde la entrada de la red neuronal se representa como un vector de características de baja dimensión. La validación se realiza utilizando diferentes bases de datos donde se modifica el número de procesadores de 1 a 32 lo que produjo una mejora en los tiempos de ejecución sin disminuir la precisión del texto categorizado. Los resultados mostraron que el IBPNN paralelo junto con la técnica SVD logra un proceso de entrenamiento más rápido, adaptativo y confiable en la categorización de textos.

La Tabla 2 muestra un resumen de técnicas comúnmente utilizadas en el proceso de entrenamiento de los modelos de series de tiempo.

Tabla 2. Resumen de técnicas comúnmente utilizadas para el entrenamiento de modelos de series de tiempo

Autor	Técnicas de entrenamiento				
	RBF	OLS	BPNN	FFN	SGD
(Gracia et al., 2006)	✓				
(Bataineh & Marler, 2017)		✓		✓	
(Wang et al., 2017)					✓
(Lin et al., 2018)			✓		
(Li et al., 2014)			✓		
(Gu et al., 2018)	✓				
(Y. Liang et al., 2019)				✓	
(Hacibeyoglu & Ibrahim, 2018)				✓	
(Hannah Jessie Rani & Aruldoss Albert Victoire, 2018)	✓				
(Chouikhi & Alimi, 2018)	✓	✓		✓	

5.5. Comparación de las métricas de desempeño para el modelado de series de tiempo

En la Tabla 3 se muestra un resumen de las técnicas comúnmente utilizadas para evaluar el desempeño de los modelos de series de tiempo.

Tabla 3. Resumen de técnicas comúnmente utilizadas en la comparación de desempeño de modelos de series de tiempo

Autor	Métricas de comparación						
	MSE	RMSE	MPE	MAPE	MAE	AMRE	CDF
(Awolusi et al., 2019)		✓	✓				
(Errachdi & Benrejeb, 2018)	✓						
(Wong et al., 2018)		✓		✓	✓		
(Ramyar & Kianfar, 2019)	✓						
(Moshkbar-Bakhshayesh, 2020)						✓	✓
(Garg et al., 2016)	✓						
(Zambri et al., 2015)		✓					
(Tabaraki & Khodabakhshi, 2020)		✓				✓	
(Kumar et al., 2015)				✓			
(Ebrahimi et al., 2017)		✓					

Comentarios finales

Análisis Factorial:

- El estado del arte destaca la importancia del análisis factorial como herramienta para identificar variables significativas en modelos económicos.
- La técnica ha demostrado ser efectiva para reducir la dimensionalidad de los datos mientras mantiene la información relevante.

Redes Neuronales:

- Las arquitecturas simples de una capa han mostrado buen rendimiento en diversos estudios del sector eléctrico.
- Los modelos monocapa permiten un balance adecuado entre capacidad predictiva e interpretabilidad.
- Se han reportado MAPE entre 1.5% y 4% en aplicaciones similares con arquitecturas simples.

Variables Significativas:

- Las variables macroeconómicas han demostrado ser predictores importantes en modelos de valoración del sector eléctrico.
- La literatura sugiere que la combinación de variables históricas y económicas proporciona mejores resultados que el uso de variables individuales.

Aspectos Metodológicos:

- El análisis estadístico previo resulta fundamental para la selección apropiada de variables.
- La normalización y preprocesamiento de datos son cruciales para el rendimiento del modelo.
- La validación rigurosa de los modelos es una práctica constante en la literatura.

Consideraciones Generales:

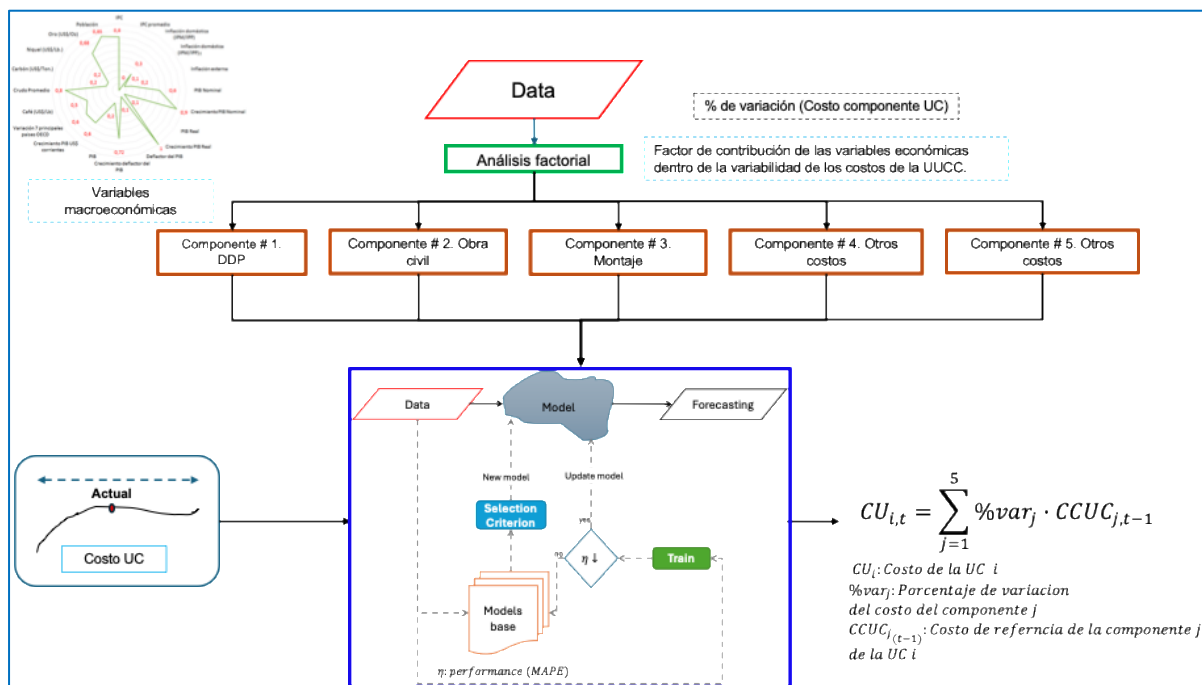
- No existe un consenso sobre el modelo ideal, pero hay una tendencia hacia estructuras simples y robustas.
- La combinación de análisis factorial y redes neuronales aparece frecuentemente en la literatura reciente.
- La interpretabilidad del modelo se mantiene como un factor importante en aplicaciones prácticas.
- Los estudios enfatizan la importancia de balancear complejidad y precisión.

6. Metodología Propuesta para la Actualización de Costos de Unidad Constructiva**6.1. Definición**

La variación de los costos de las unidades constructivas es un fenómeno que no es sencillo de modelar debido a los diferentes inductores que afectan la variabilidad de cada uno de los componentes que la definen: DDP, obra civil, montaje y demás costos financieros. Los componentes antes mencionados se ven afectados por algunas características macroeconómicas del mercado como: precio de minerales, precio de materiales como el cobre, acero, aluminio, oro, tasas de cambio, indicadores de crecimiento económico, inflación, etc.

Establecer relaciones entre cada una de las variables no es una labor sencilla de describir haciendo uso de modelos estadísticos convencionales. Ante esto se propone la integración de modelos basados en inteligencia artificial que permitan establecer relaciones entre las series de tiempo que rigen el comportamiento de los costos de cada unidad constructiva objeto de estudio y las diferentes variables macroeconómicas dispuestas en la base de datos disponible. Para la actualización de los costos de las unidades constructivas se tomó como referencia la propuesta de análisis y modelado de series de tiempo planteada por los autores de los trabajos (Jimenez et al., 2019; Jimenez Mares et al., 2023). En este tipo de planteamientos se busca seleccionar, depurar, filtrar, agrupar y simplificar el conjunto de variables exógenas (e.g. macroeconómicas) que tengan un nivel de significancia dentro del proceso a modelar.

Imagen 1. Metodología propuesta para la actualización de los costos de las UCs.



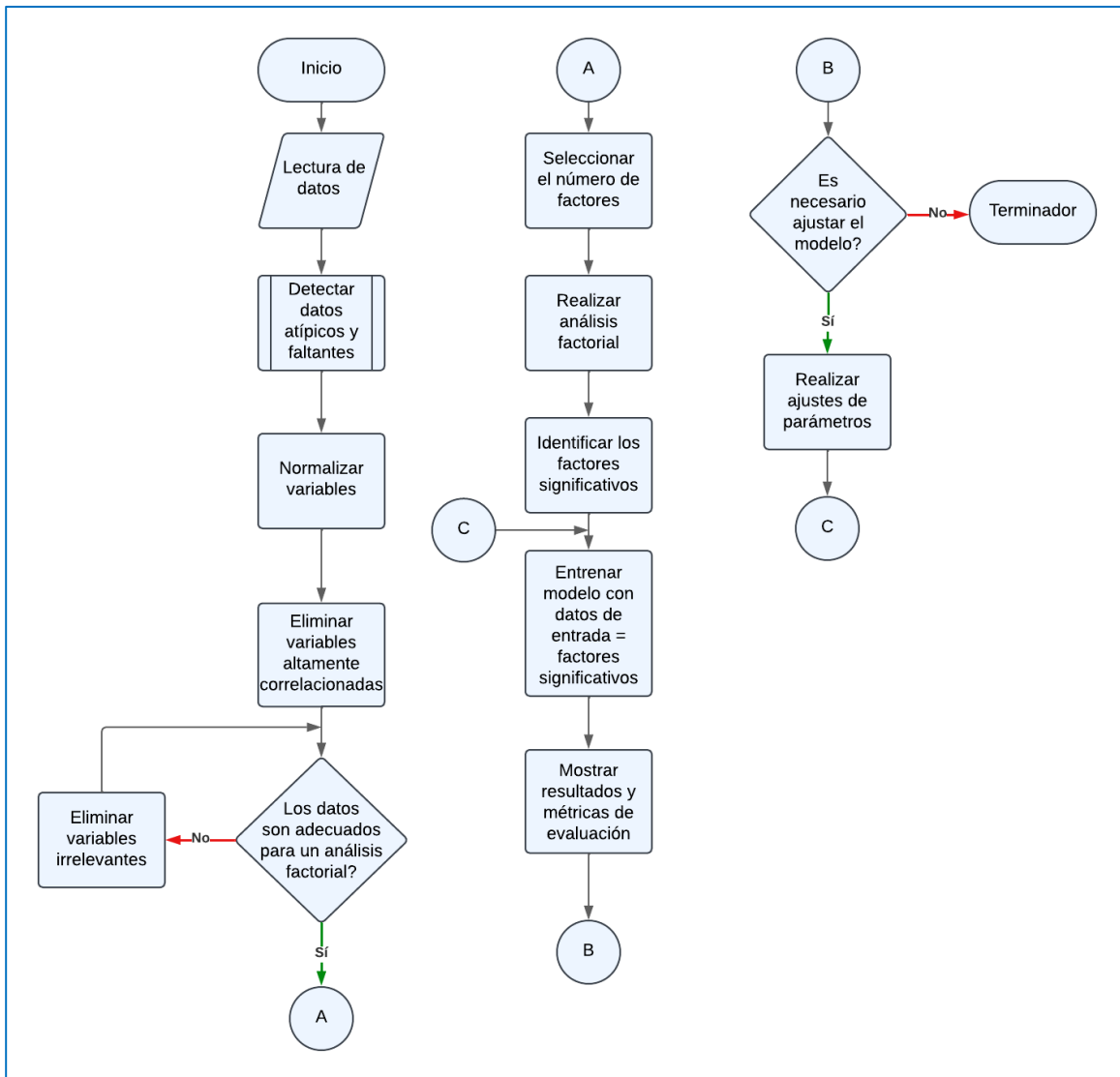
En la imagen 1 se muestra cada una de las etapas y/o procesos que definen la metodología de actualización de costos aquí propuesta. En primera instancia, se propone la implementación de un análisis estadístico multivariado que involucra análisis factorial para evaluar por cada componente y unidad constructiva las variables que resulten significativas dentro del proceso de modelado. El análisis factorial es una técnica estadística de reducción de dimensiones que describe cada variable en términos de una combinación lineal de un pequeño número de factores comunes no observables y un factor único para cada variable. De esta manera, el objetivo es encontrar los factores comunes que recojan el máximo de información de las variables originales (Díaz Monroy, 2007). Una vez determinado el número y valores de los factores para cada uno de los componentes que describen cada unidad constructiva, se procede al entrenamiento de los diferentes modelos basados en inteligencia artificial disponibles. Para la propuesta aquí planteada, se incluyen redes neuronales artificiales y bosques aleatorios, por sus siglas en inglés, random forest. De manera iterativa se lleva a cabo el proceso de entrenamiento de cada uno de los modelos disponibles con el objetivo de evaluar y seleccionar aquel con el mejor desempeño (MAPE) para cada caso. Como resultado de esta metodología se obtendrá un modelo basado en redes neuronales o bosques aleatorios por cada componente y unidad constructiva.

La propuesta aquí planteada proporcionará a la CREG la posibilidad de una actualización continua sin la necesidad de consultar los diferentes proveedores del mercado. Lo anterior, gracias a que se proporcionará una interfaz que tomando como referencia los valores futuros de las diferentes variables exógenas, se actualizarán los costos de cada unidad constructiva seleccionada.

6.2. Diseño funcional de la solución propuesta

La metodología propuesta se fundamenta en cada una de las etapas funcionales mostradas en la imagen 2:

Imagen 2. Metodología propuesta para la actualización de los costos de las UCs



6.3. Preprocesamiento de datos

El preprocesamiento de datos es fundamental para garantizar estimaciones confiables en la valoración de unidades constructivas. El primer paso consiste en la estructuración adecuada de los datos, identificando y separando los componentes principales: obra civil, montaje, otros costos y costos financieros. El componente DDP se excluye del análisis por ser determinístico, representando un 24.5% del valor base que se calcula directamente.

El tratamiento de valores faltantes sigue un protocolo estricto: no se permiten valores ausentes en los componentes principales de la unidad constructiva para mantener la integridad de los datos. Para las variables económicas y explicativas, se utiliza la media como método de imputación, documentando cualquier reemplazo realizado para mantener la trazabilidad del proceso.

Para el manejo de valores atípicos se implementa el método IQR (Rango Intercuartílico), estableciendo límites superior e inferior ($Q3 + 1.5IQR$ y $Q1 - 1.5IQR$ respectivamente). En los componentes de la unidad constructiva, los *outliers* solo se detectan y registran, mientras que en las variables económicas se aplica *winsorización*¹ para corregir valores extremos sin perder la esencia de las variaciones del mercado.

La normalización de variables mediante estandarización (media cero y desviación estándar uno) se aplica a todas las variables numéricas. Este paso es crucial dado que manejamos diferentes escalas, desde millones en valores de componentes hasta porcentajes en variables económicas. La estandarización permite la comparabilidad directa entre variables y optimiza el rendimiento del modelo posterior.

Finalmente, se verifica la calidad de los datos procesados mediante la validación de consistencia y relaciones lógicas entre componentes. Los datos resultantes se estructuran en matrices adecuadas para el análisis factorial, verificando previamente los requisitos de correlación y adecuación muestral necesarios para este análisis.

6.4. Análisis factorial

El análisis factorial constituye una etapa crítica para la identificación de variables significativas en la valoración de unidades constructivas. La metodología inicia con la verificación de factibilidad mediante el test de esfericidad de *Bartlett*, que evalúa si la matriz de correlaciones difiere significativamente de una matriz identidad, y el test *KMO* (Kaiser-Meyer-Olkin), que mide la adecuación muestral. Estos tests determinan si el conjunto de datos es apropiado para el análisis factorial.

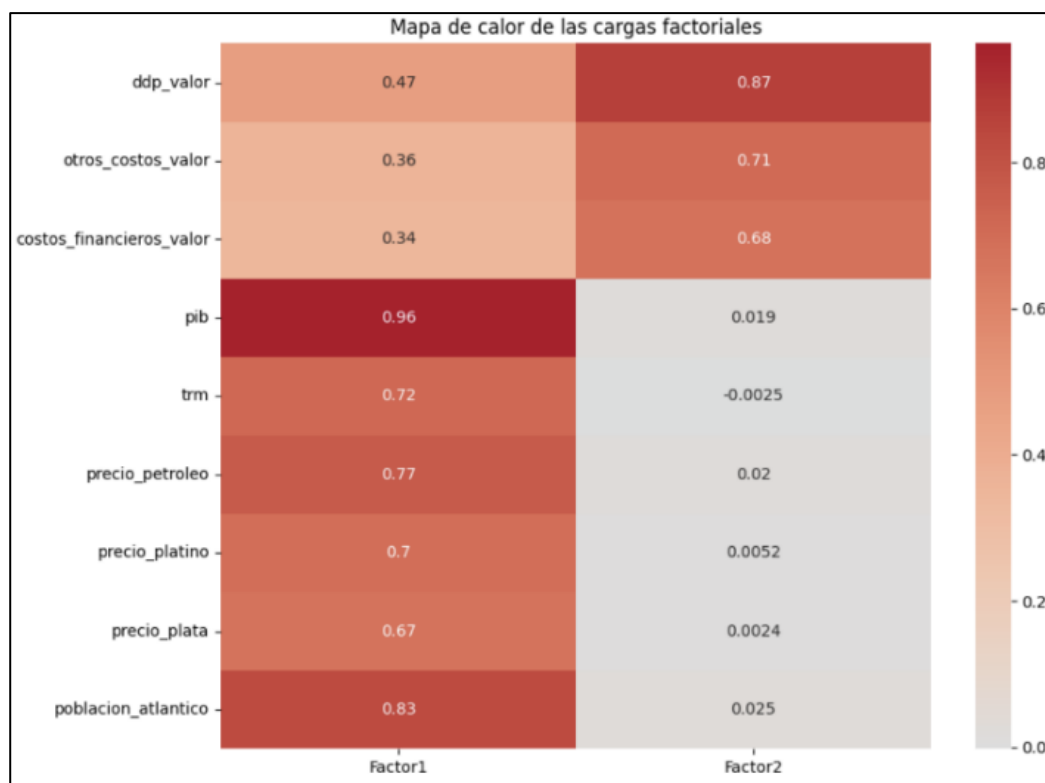
Una vez confirmada la factibilidad, se procede a la identificación de factores principales mediante el criterio de *Kaiser* (eigenvalores mayores a 1) y el análisis del *scree plot*. Las variables se agrupan según sus cargas factoriales, considerando significativas aquellas con valores absolutos superiores a 0.5. La rotación *Varimax* se aplica para maximizar la interpretabilidad de los factores resultantes.

La estructura factorial resultante permite reducir la dimensionalidad del problema mientras se mantiene la información relevante para el modelo. Las correlaciones entre variables y factores se visualizan mediante una matriz de calor (ver Figura 3), facilitando la identificación de patrones

¹ La winsorización es una técnica estadística que se utiliza para limitar los valores extremos en conjuntos de datos, reduciendo efectivamente la influencia de los valores atípicos en los análisis estadísticos.

y relaciones significativas. Este proceso asegura que solo las variables más influyentes sean consideradas en el modelo posterior (Jimenez et al., 2019).

Imagen 3. Mapa de calor de cargas factoriales



En la Imagen 3 se muestra un ejemplo de un mapa de calor resultado de un análisis factorial para un caso de estudio. Como resultado se muestran dos factores donde se evidencia que el factor 1 se encuentra caracterizado por las pib, trm, precio_petróleo, precio_platino, precio_plata y población_atlántico, mientras que el factor 2 se encuentra definido por las variables ddp_valor, otros_costos-valor y costos_financieros_valor.

6.5. Modelos basados en inteligencia artificial

Desde una visión más amplia, se puede definir como un sistema que propone soluciones a casos complejos donde los algoritmos convencionales no pueden resolver (Jabbar et al., 2010). Este concepto puede estar sujeto a ambigüedades, por lo que es necesario aclarar que un sistema de inteligencia computacional debe cumplir los siguientes criterios (Eberhart & Shi, 2007).

- Sólo se ocupa de los datos numéricos (nivel bajo).
- Tiene un componente de reconocimiento de patrones.
- Sentido Numérico –Actividades del ser humano como resolución de problemas, toma de decisiones y aprendizaje–.

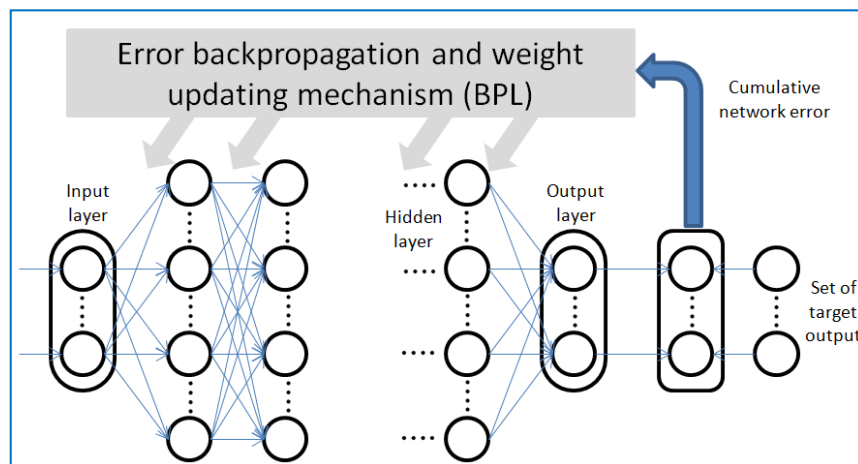
Este tipo de sistemas combina elementos de aprendizaje, adaptación, evolución y lógica difusa para diseñar programas con cierto nivel de inteligencia. Se basa en la experiencia y proporciona las decisiones correctas a pesar de las limitaciones. (Konar, 2005).

La inteligencia computacional aplicada al modelado de series temporales ha permitido desarrollar diversas aplicaciones orientadas a la previsión de series temporales, análisis de riesgos, perfilación de clientes, redes inteligentes, etc.

Redes Neuronales (RN)

Este tipo de técnica de inteligencia artificial se basa en el funcionamiento de las redes neuronales biológicas del cerebro humano. Por tanto, están compuestos por unidades llamadas neuronas y también son capaces de aprender de la experiencia, generalizar a partir de ejemplos anteriores y abstraer las principales características de una serie de datos. Así, las RN, cambian su comportamiento dependiendo del entorno, proporcionan respuestas correctas a entradas con pequeñas variaciones, entre otras ventajas (Bilbao, 2008; Flórez & Fernández, 2008).

Imagen 4. Topología de las redes neuronales (Fakhreddine O. Karray & Silva, 2004)



La Imagen 4 muestra la estructura básica de una red neuronal entrenada con algoritmo de retropropagación. La retropropagación es un método para entrenar la red neuronal utilizando tres o más capas. La topología comúnmente utilizada es la de tres capas: una inferior y una superior oculta. La capa inferior o capa de entrada son las únicas unidades de procesamiento que reciben entradas del exterior, por lo tanto, los puntos sirven como distribuidores y no realizan una operación de cálculo. La capa (en el caso de una topología de más de tres capas) entre la capa inferior y la superior se denomina oculta, en ella las unidades procesadoras están interconectadas con capas adyacentes y realizan el cálculo de las operaciones. Finalmente, la capa de salida es la que representa la respuesta de la red. La práctica de esta topología es monitoreada, es decir, los patrones de entrada coinciden y se deben ingresar las salidas objetivo (Bilbao, 2008; Flórez & Fernández, 2008).

El modelo neuronal seleccionado consiste en una arquitectura simple de una sola capa oculta, donde el número de neuronas corresponde directamente a la cantidad de variables significativas

identificadas en el análisis factorial. La capa de entrada recibe las variables normalizadas seleccionadas, mientras que la capa de salida proporciona directamente el valor estimado de la unidad constructiva.

Para garantizar la estabilidad y rendimiento del modelo, se implementan elementos clave de regularización: normalización de datos de entrada, regularización L1 y L2 para evitar sobreajuste, y early stopping para determinar el punto óptimo de entrenamiento. El optimizador Adam con una tasa de aprendizaje conservadora permite una convergencia estable.

6.6. Selección adecuada de una arquitectura de red neuronal

Después de seleccionar la estructura de la red neuronal a utilizar el reto está en determinar la cantidad de parámetros óptima a utilizar. Un buen modelo debe brindar un error razonable en la fase de entrenamiento, validación y de prueba sin recaer en sobreajuste. Como fase inicial de diseño se debe llevar a cabo un análisis estadístico exploratorio que permita identificar las variables significativas que mejoren el desempeño del modelo. La selección de la red con un mejor desempeño no es una tarea fácil ya que no hay una teoría formal que permita simplificar esta tarea.

En el estado del arte se recomienda la aplicación de teorías relacionadas con el diseño de experimentos y validación cruzada con el fin de recorrer la zona de experimentación que incluya la mayor variación de los posibles parámetros para la selección de la red neurona con un mejor desempeño. Otro aspecto por considerar es la cantidad de parámetros disponibles y su afectación en la complejidad del modelo, por tal motivo se recomienda considerar el criterio de Akaike con el fin de penalizar la adición de parámetros extras. En resumen, la selección de la configuración y estructura adecuada de la red neuronal es crucial para el proceso de pronóstico. El acondicionamiento de los datos, por ejemplo, normalización, re-escalamiento o transformación también brindan una mejoría en el desempeño de los modelos basados en redes neuronales.

- **Random Forest**

El bosque aleatorio es un algoritmo de aprendizaje automático flexible y fácil de usar que produce, incluso sin ajuste de hiperparámetros, un gran resultado la mayor parte del tiempo. También es uno de los algoritmos más utilizados, por su simplicidad y diversidad (puede usarse tanto para tareas de clasificación como de regresión)(Schonlau & Zou, 2020).

Un modelo basado en árbol implica dividir recursivamente el conjunto de datos dado en dos grupos según un determinado criterio hasta que se cumpla una condición de parada predeterminada. En la parte inferior de los árboles de decisión se encuentran los llamados nodos de hojas u hojas.

La Imagen 5 muestra una partición recursiva de un espacio de entrada bidimensional con límites alineados con los ejes, es decir, cada vez que el espacio de entrada se divide en una dirección paralela a uno de los ejes. Aquí la primera división se da en $x_2 \geq a_2$. Luego, los dos subespacios se dividieron nuevamente: la rama izquierda se dividió en $x_1 \geq a_4$. La rama derecha fue primero

dividida en $x_1 \geq a_1$, y una de sus subramas fue dividida en $x_2 > a_3$. La Figura 6 es una representación gráfica del subespacio particionado.

Imagen 5. Partición binaria recursiva de subespacios bidimensionales

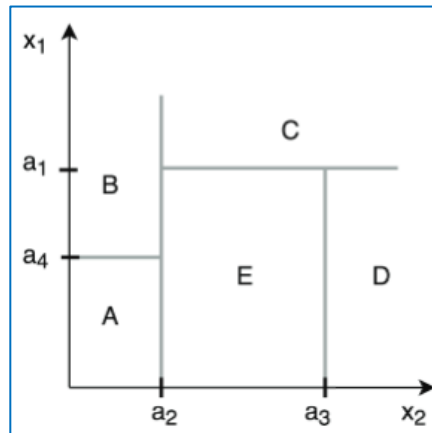
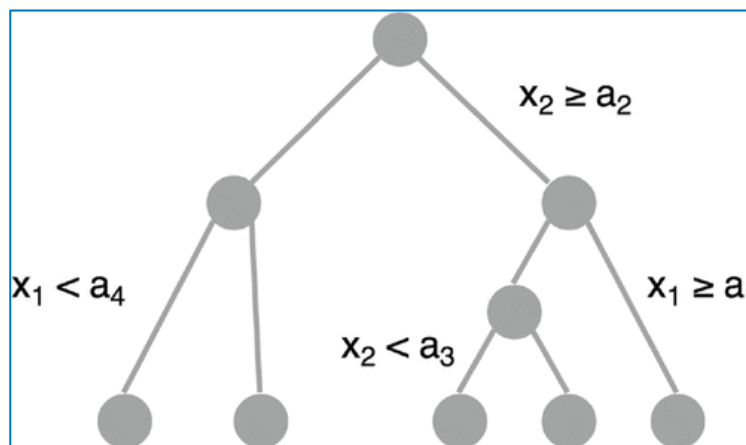


Imagen 6. Representación gráfica del árbol de decisión.



Dependiendo de cómo se establezcan los criterios de partición y detención, se pueden diseñar árboles de decisión tanto para tareas de clasificación (resultado categórico, por ejemplo, regresión logística) como para tareas de regresión (resultado continuo).

6.7. Proceso de entrenamiento

En la metodología propuesta en la Imagen 1, se plantea una estrategia de entrenamiento y selección de los modelos de series de tiempo disponibles (RNA y RF). En esta propuesta el entrenamiento de los modelos es integrado a un proceso iterativo que busca evaluar todas las configuraciones establecidas para seleccionar aquella con el mejor desempeño (MAPE) para cada clase de los modelos disponibles (Calvetti & Somersalo, 2007).

6.8. Métricas utilizadas para la evaluación del desempeño de los modelos

En esta sección hemos explicado las técnicas de modelado de series de tiempo comúnmente utilizadas. La fase de construcción y selección de los modelos de predicción requiere de un adecuado proceso de entrenamiento y evaluación de las posibles configuraciones para cada técnica. Para la construcción de un modelo particular de una serie real o simulada es necesario dividir los datos de la serie de tiempo en dos partes: conjunto de entrenamiento y conjunto de validación. Los datos ubicados dentro del conjunto de entrenamiento serán la base para la construcción del modelo deseado.

A su vez el conjunto de validación puede subdividirse en dos: conjunto de validación utilizado para evaluar el desempeño del modelo obtenido durante el entrenamiento y conjunto de prueba que es utilizado para evaluar el desempeño del modelo seleccionado a finalizar el proceso de entrenamiento. Dependiendo del conjunto de variables a considerar y algunas características como la magnitud, se hace necesario aplicar técnicas de preprocesamiento y transformación a escalas comunes a todas, por ejemplo, normalización, transformación logarítmica, diferenciación. Una vez obtenido el resultado de parte del modelo debe aplicarse la transformación inversa a la realizada con las variables de entrada para poder medir el desempeño en las mismas unidades de la serie de tiempo original.

A continuación, se describirán las métricas de desempeño comúnmente utilizadas. $e_i = y_i - \hat{y}_i$, donde,

y_i : valor actual.

$\hat{y}_i$: valor previsto.

e_i : error calcula para los n datos de la serie de tiempo.

- **Mean Absolute Percentage Error (MAPE)**

Esta metrica se define en la ecuacion (2.8-4):

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right| \times 100 \quad (6.8-1)$$

Algunas de sus propiedades son:

- Es independiente de la escala, pero si es afectada por las transformaciones realizadas.
- No muestra la dirección del error.
- Los valores extremos no son penalizados.

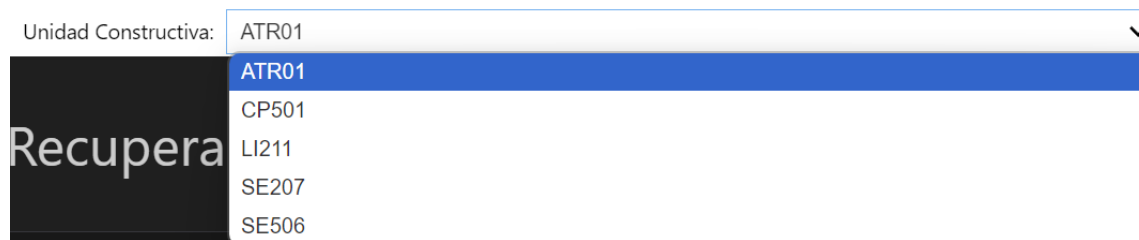
La validación del modelo se realiza mediante división temporal de datos, manteniendo un conjunto de validación cronológicamente posterior al conjunto de entrenamiento, lo que simula las condiciones reales de aplicación. El rendimiento se evalúa principalmente mediante el error porcentual absoluto medio (MAPE), buscando mantener errores dentro de rangos aceptables para aplicaciones prácticas.

7. Implementación

7.1. Interfaz de actualización de costos de las unidades constructivas

En la aplicación se ha de seleccionar en primer lugar la unidad constructiva a evaluar, para esto, se dispone de un selector cuyas opciones corresponden a los nombres (identificaciones) de cada Unidad Constructiva. Por ejemplo, en caso de disponer una base de datos con 5 unidades constructivas para el entrenamiento, estas serán las disponibles en el selector (ver Imagen 7).

Imagen 7. Ejemplo listado de UUCC disponibles.



Unidad Constructiva: ATR01

ATR01

CP501

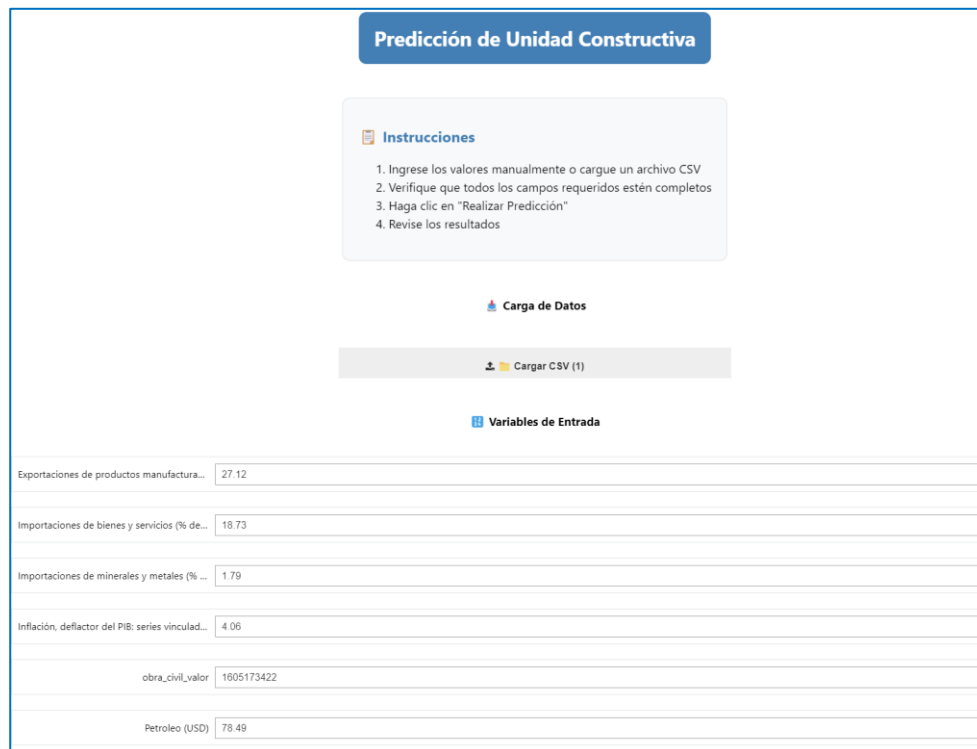
LI211

SE207

SE506

Una vez es seleccionada la unidad constructiva, se procede con el ingreso de los datos requeridos por el modelo, para esto, puede llenar las casillas por cada variable macroeconómica/elemento como también se puede cargar un archivo csv con las variables necesarias.

Imagen 8. Sección de ingreso de datos



Predicción de Unidad Constructiva

Instrucciones

1. Ingrese los valores manualmente o cargue un archivo CSV
2. Verifique que todos los campos requeridos estén completos
3. Haga clic en "Realizar Predicción"
4. Revise los resultados

Carga de Datos

Cargar CSV (1)

Variables de Entrada

Exportaciones de productos manufactura...	27.12
Importaciones de bienes y servicios (% de...	18.73
Importaciones de minerales y metales (% ...	1.79
Inflación, deflactor del PIB: series vinculad...	4.06
obra_civil_valor	1605173422
Petroleo (USD)	78.49

En caso de que el usuario desee llevar a cabo el ejercicio de actualización de los costos de manera individual para cada unidad constructiva, el usuario debe diligenciar cada uno de los datos de las variables que aparecen en el formulario mostrado en la Imagen 8.

El formato del archivo csv se puede observar al final de la sección (ver Imagen 9)

Imagen 9. Formato de CSV esperado

Formato de CSV Esperado												
	Exportaciones de productos manufacturados (% de las exportaciones de mercancías)	Importaciones de bienes y servicios (% del PIB)	Importaciones de minerales y metales (% de importaciones de mercancías)	Inflación, deflactor del PIB: series vinculadas (% anual)	obra_civil_valor	Petroleo (USD)	piib	cobre x (USD)	Acero (USD)	Inversión extranjera directa, entrada neta de capital (% del PIB)	oro x (USD)	valor_actual
0	0	0	0	0	0	0	0	0	0	0	0	0

Una vez se han ingresado los datos, ya sea por cada casilla o a través de un archivo .csv, se debe de presionar el botón “Realizar predicción” para poder ejecutar el modelo de la UC seleccionada y así poder visualizar el costo actualizado y tasa de variación con respecto al costo actual.

Imagen 10. Resultados de un modelo de actualización de costos de la UC.

Realizar Predicción

Resultados de la Predicción

Costo UC actual: \$12,919,313,583.00

Costo UC actualizado: \$14,690,324,218.75

Incremento: 13.71%

	Exportaciones de productos manufacturados (% de las exportaciones de mercaderías)	Importaciones de bienes y servicios (% del PIB)	Importaciones de minerales y metales (% de importaciones de mercaderías)	Inflación, deflactor del PIB: series vinculadas (% anual)	obra_civil_valor	Petroleo (USD)	piib	cobre_x (USD)	Acero (USD)	Inversión extranjera directa, entrada neta de capital (% del PIB)	oro_x (USD)	valor_actual
0	27.12	18.73	1.79	4.06	1605.173422	78.49	1.14	234.22	570.0	3.46	972.4	12919.313583

7.2. Resultados

Descripción de los datos

Para validar el desempeño de los modelos obtenidos se tomó como referencia la base datos de costos internacionales de Perú publicadas en OSINERGMIN 2023 y nacionales publicados en la circular 090 de 2021 y la circular CREG 076 de 2024 para las UC convencionales y no asimiladas. Para el caso de las UC de infraestructura nueva se tomaron como referencia los datos provistos por la circular CREG 076 de 2024 y la circular CREG 092 de 2024. Como ventana de análisis se consideró un periodo comprendido entre el año 2011 hasta el 2022. Los datos de las variables macroeconómicas utilizadas dentro del proceso de análisis multivariado se tomaron de las fuentes y vigencias mostradas en la Tabla 4.

Tabla 4. Fuentes de las variables exógenas utilizadas por la metodología propuesta

Variable	Fuente
Año	Base de datos UC
DDP	

Variable	Fuente
Costo obra civil	
Costo montaje	
Otros costos	
Costos financieros	
Costo total UC	
IPC	https://www.sintraprevi.org/pdf/indicadores/ipc.pdf
Petróleo (USD)	https://es.investing.com/commodities/brent-oil-historical-data
TRM	https://suameca.banrep.gov.co/estadisticas-economicas/#!/informacionSerie/1/Tasa%20de%20cambio%20del%20peso%20colombiano/Tasa%20de%20cambio%20Representativa%20del%20Mercado%20%28TCRM%29
Precio del oro (USD)	https://www.cochilco.cl/web/historico-precio-de-los-metales/
Precio del cobre (USD)	
Acero (USD)	
Aluminio (USD)	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías)	https://datos.bancomundial.org/pais/colombia
Importaciones de minerales y metales (% de importaciones de mercaderías)	
Crecimiento de la población (%) anual	
PIB	
Ahorro bruto (% del PIB)	
Inflación, deflactor del PIB: series vinculadas (% anual)	
Inversión extranjera directa, entrada neta de capital (% del PIB)	
Importaciones de bienes y servicios (% del PIB)	

Los datos se normalizan para evitar diferencias en las magnitudes que puedan afectar la coherencia de la distribución de los datos. El proceso de normalización se realiza a través de la Ecuación (1) para cada variable a considerar para el pronóstico de series de tiempo. Los datos normalizados se establecen en 0 a 1.

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

donde,

x : dato del conjunto de datos.

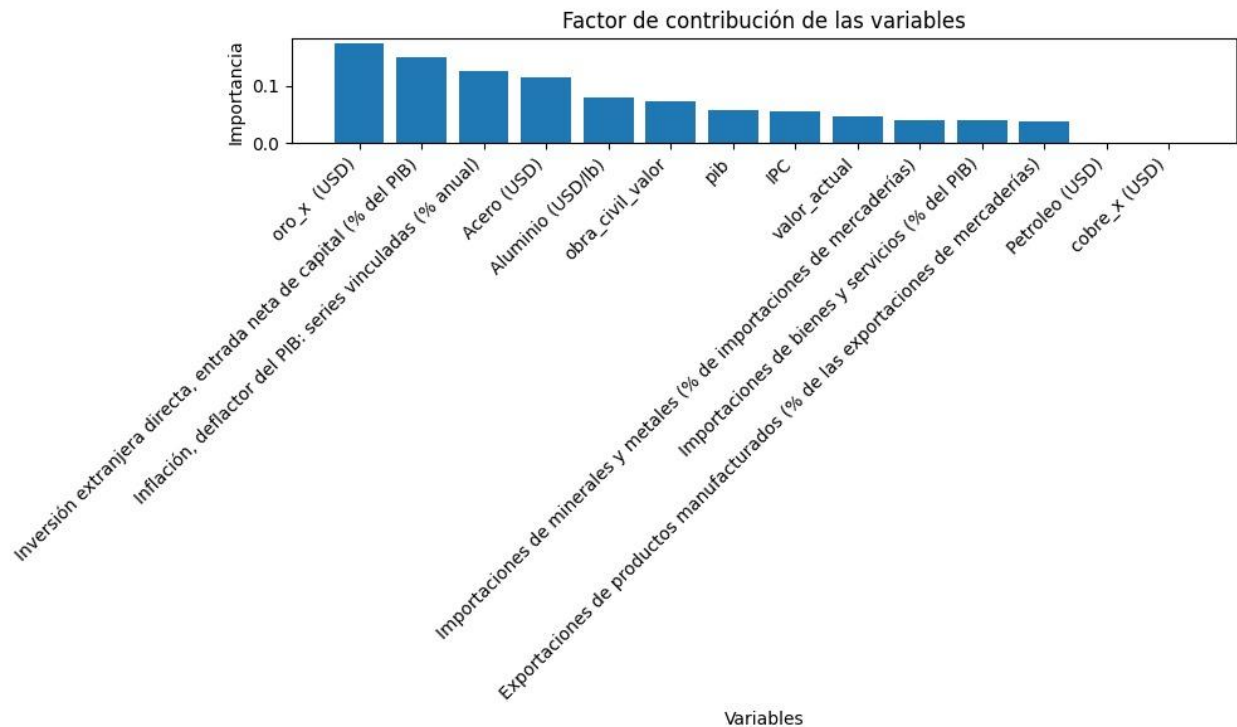
μ : media del conjunto de datos

σ : desviación del conjunto de datos.

Como resultado del proceso de análisis factorial aplicado al conjunto de variables disponibles, se obtuvo como factores requeridos los mostrados en la Figura 11.

Imagen 11. Mapa de calor del análisis factorial

Imagen 12. Factor de contribución de las variables económicas dentro de la variabilidad de los costos de las unidades constructivas



En la Imagen 12 se muestra el factor de contribución de cada una de las variables consideradas dentro del análisis utilizado para la ponderación de cada una de las variables macroeconómicas. Los resultados de esta etapa brindan información complementaria al análisis factorial aquí propuesto.

Las correlaciones entre variables y factores resultantes del proceso de análisis exploratorio para las variables macroeconómicas y los costos de las UC se visualizan mediante una matriz de calor

(ver Imagen 11). La Imagen 11 muestran las cargas factoriales, donde se evidencia la necesidad de cuatro (4) factores. Teniendo en cuenta las cargas factoriales se agrupan de la siguiente manera:

- Factor mano de obra: obra civil y precio del acero.
- Factores materiales: PIB, precio del cobre y precio del petróleo.
- Factor costos de importación: inflación, inversión extranjera e importaciones de bienes y servicios.
- Factor costos de nacionalización y exportaciones: exportaciones de productos manufacturados e importaciones de minerales y metales.

Estimación de los parámetros de la red neuronal.

Por otro lado, en lo que respecta al proceso de entrenamiento de los modelos basados en redes neuronales dado que no existe una teoría formal que permita garantizar la selección de parámetros y configuración óptimos de las redes neuronales que brinden un máximo nivel de generalización para fines de pronóstico, se propone una metodología de diseño experimental. La metodología planteada en (Jiménez et al., 2021) evidenció como factor significativo dentro de la configuración de las redes neuronales el número de neuronas en la capa oculta.

Por lo tanto, para llevar a cabo la selección de la red neuronal adecuada, se aplicó el algoritmo de entrenamiento propuesto en (Jiménez et al., 2021); en donde se utilizó un *script* desarrollado en Python utilizando la librería scikit-learn y tensor flow que busca modificar parámetros de la red neuronal recorriendo una amplia zona de experimentación para evaluar el MAPE y seleccionar aquella configuración con mejor desempeño. En las Tablas 5 y 6 se muestran los parámetros obtenidos para las redes neuronales con mejor desempeño.

Tabla 5. Configuración de la red neuronal multicapa seleccionada.

Parámetros	Valores
Número de capas ocultas	2
Numero de neuronas en la capa oculta	[32, 16]
Funcion de activación capas ocultas	Unidad lineal rectificada (ReLU)
Funcion de activación capa de salida	Lineal
Kernel regularizador	l2(0.01)

Tabla 6. Configuración de la red neuronal monocapa seleccionada

Parámetros	Valores
Número de capas ocultas	2
Funcion de activación	Lineal
Kernel regularizador	l2(0.01, 0.01)
Kernel inicializador	HeNormal

Estimación de los parámetros de los modelos basados en random forest.

Para entrenar modelos RTE se propone un proceso de optimización basado en la teoría bayesiana. La teoría de optimización bayesiana se utiliza para llevar a cabo la optimización global de funciones multimodales, como modelos basados en sistemas de inteligencia computacional. Un hipermodelo bayesiano representa el comportamiento de los datos y la información cualitativa a través de la distribución de probabilidades (Jimenez Mares et al., 2023). A continuación, se muestran los parámetros obtenidos para este tipo de modelos como resultado del proceso antes mencionado. En las Tablas 7 y 8 se muestran los parámetros de diseño de los modelos RTE con mejor desempeño en el entrenamiento para los datos de costos UC analizados.

Tabla 7. Configuración del modelo basado en *random forest* seleccionado

Parámetros	Valores
Número de estimadores	300
Max_depth	20

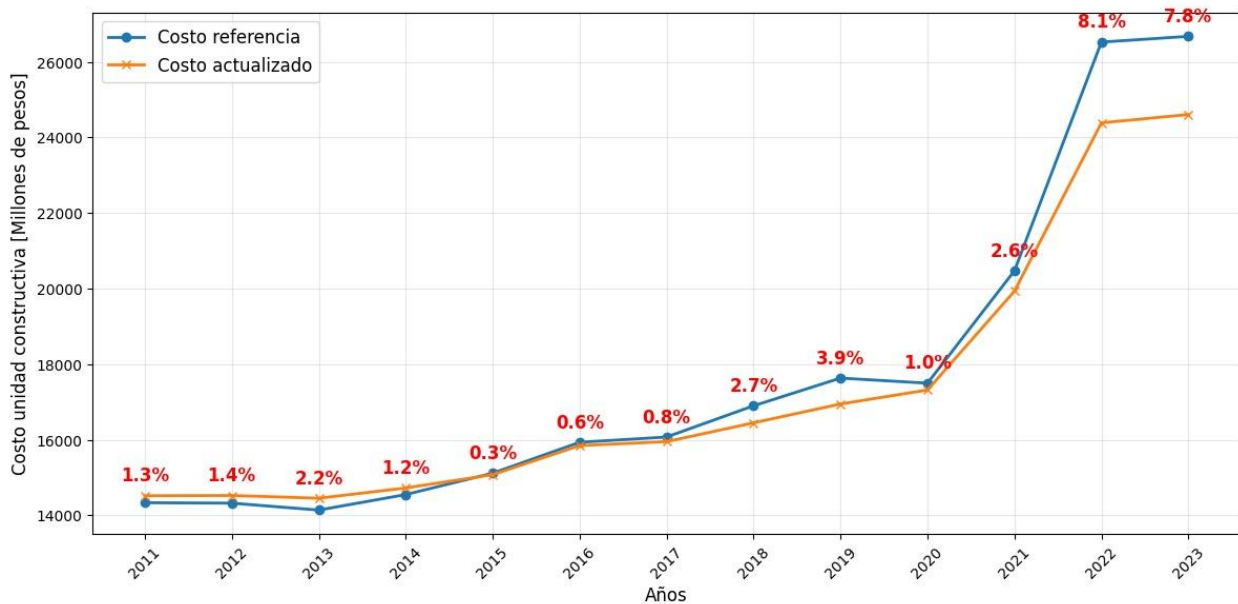
Tabla 8. Configuración del modelo XGBRegressor (árbol de decisión) seleccionado

Parámetros	Valores
Número de estimadores	500
Max_depth	4
Learning rate	0.01
Tree method	Hist
reg_alpha	0.01
reg_lambda	0.1
gamma	0.05

Resultados de los modelos de actualización de los costos de las unidades constructivas

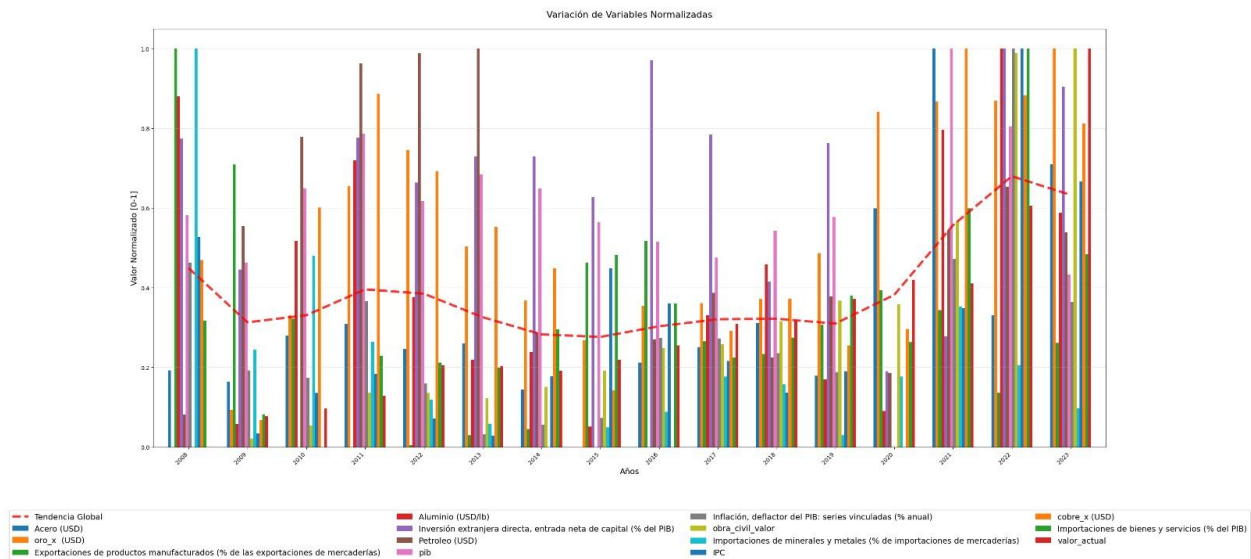
Como resultado de la implementación de la metodología mostrada en la Imagen 1 para la totalidad de las UC objeto de estudio, se logró alcanzar una MAPE igual a 4.18%. En la Imagen 12 se muestra el error obtenido en una UC por cada año analizado. En ella se evidencia un máximo error de 3.4% y un mínimo error de 0.1%, lo cual permite concluir que la metodología propuesta brinda la posibilidad de ensamblar modelos para la estimación de los costos de las UCs con niveles de confianza adecuados.

Imagen 13. Resultados del proceso de entrenamiento de los modelos obtenidos para la actualización de costos de la UC



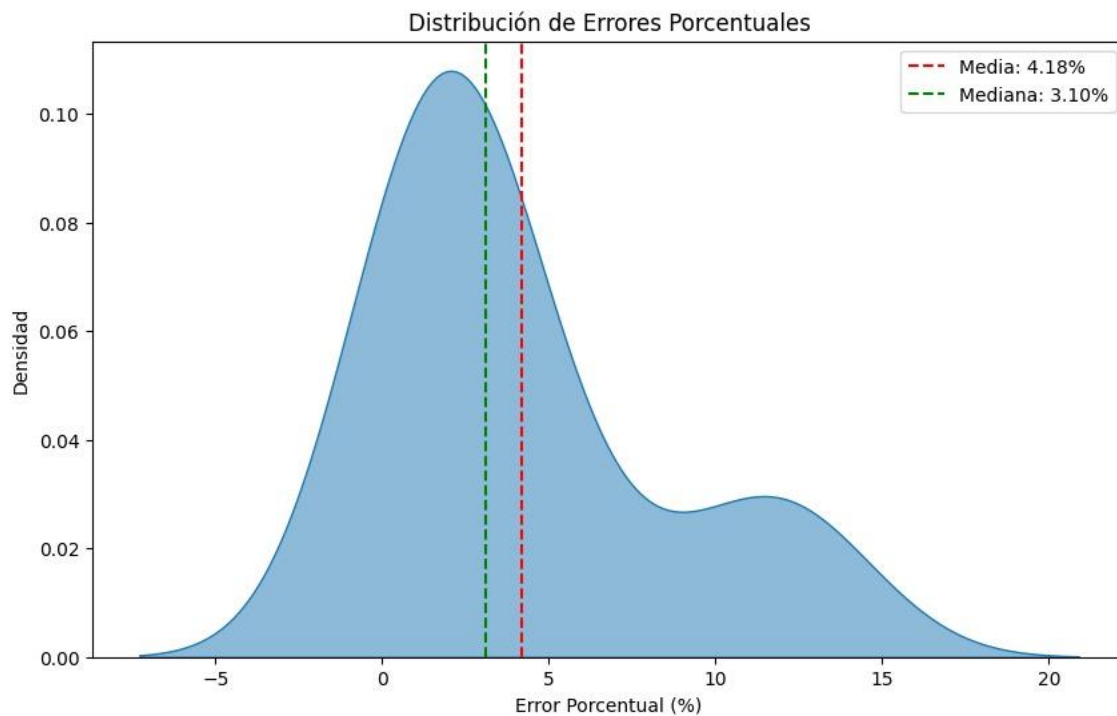
Adicionalmente, los resultados mostrados en la Imagen 13 evidencian efectos como la pandemia y otras situaciones exógenas como conflictos armados a nivel mundial, que posiblemente afectan el desempeño del modelo, el cual sufrió un detrimento en los niveles de error a partir del año 2020; sin embargo, se observa la flexibilidad y capacidad para seguir la tendencia creciente de los costos de las UC ocasionado por el aumento de los costos de la logística de transporte. Estos resultados muestran que a pesar de verse afectado por factores externos que no son inherentes a los datos el modelo se mantuvo estable en su comportamiento general.

Imagen 14. Comportamiento de las variables macroeconómicas seleccionadas para los años 2011 – 2022.



En la Imagen 14 se muestran las variables macroeconómicas seleccionadas a través de la metodología propuesta. En esta gráfica es posible evidenciar un incremento en los valores de estas variables explicativas en los últimos tres años, lo cual permite sustentar el aumento de los costos de las UC analizadas.

Imagen 115. Función de densidad de probabilidad de los errores obtenidos durante el proceso de entrenamiento de los modelos para la actualización de costos de la UC.



En la Imagen 15 se muestra la función de distribución de probabilidad de los errores obtenidos durante el proceso de entrenamiento para los modelos de actualización de los costos de las unidades constructivas. En la gráfica se evidencia que el error se ajusta a una distribución bimodal con una media de 4.18% y una mediana de 3.10%, lo cual, evidencia la pertinencia en el desempeño de los modelos.

Los resultados mostrados en las Imágenes 13,14 y 15 son una muestra aplicada a una UC, la cual sirve de referencia para el análisis general del comportamiento de los modelos obtenidos. Este ejercicio se realizó para cada una de las UC definidas en el alcance del presente proyecto.

A continuación, se muestran algunos casos de estudio en donde se aplica la metodología propuesta y se comparan los resultados obtenidos en la actualización de los costos de las UC con los obtenidos por las metodologías vigentes tomadas como referencia.

- Caso # 1: UC ATR01

Imagen 116. Resultados de actualización de los costos de la UC ATR01 para el año 2011.

Actualización para UC 2011: ATR01	Unidad Constructiva (UC)
valor actual de 2010: 1571100000.0	Costo UC metodología actual año anterior
Actualización: 14520469726.5625	Costo UC Actualizado
Incremento: 824.2231383465407%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 14335342964.0	Costo UC metodología actual
Error porcentual: 1.2914010011996528%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	Variables seleccionadas para el modelo
Acero (USD): 720.0	
oro_x (USD): 1571.1	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 16.47	
Aluminio (USD/lb): 108.65	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.37	
Petroleo (USD): 108.35	
pib: 6.950000000000002	
Inflación, deflactor del PIB: series vinculadas (% anual): 6.39	
obra_civil_valor: 1386915761.0	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.81	
IPC: 3.73	
cobre_x (USD): 399.66	
Importaciones de bienes y servicios (% del PIB): 20.190000000000005	
valor_actual: 13172.076014	

- Caso # 2: UC CP209

Imagen 117. Resultados de actualización de los costos de la UC CP209 para el año 2012.

Actualización para UC 2012: CP209	Unidad Constructiva (UC)
valor actual de 2011: 16540000.0	Costo UC metodología actual año anterior
Actualización: 1487730102.5390625	Costo UC Actualizado
Incremento: 8894.74064412976%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 1466989038.9999998	Costo UC metodología actual
Error porcentual: 1.4138526592673977%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 655.0	
oro_x (USD): 1668.0	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 16.54	
Aluminio (USD/lb): 91.55	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.06	
Petroleo (USD): 110.1	
pib: 3.91	
Inflación, deflactor del PIB: series vinculadas (% anual): 3.6200000000000006	
obra_civil_valor: 141928255.6	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.66	
IPC: 2.44	
cobre_x (USD): 360.59	
Importaciones de bienes y servicios (% del PIB): 20.02	
valor_actual: 1468.144829	

- Caso # 3: UC CP501

Imagen 118. Resultados de actualización de los costos de la UC CP501 para el año 2013.

Actualización para UC 2013: CP501	Unidad Constructiva (UC)
valor actual de 2012: 83699000.0	Costo UC metodología actual año anterior
Actualización: 1430685913.0859375	Costo UC Actualizado
Incremento: 1609.3225881861642%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 1399631962.0	Costo UC metodología actual
Error porcentual: 2.218722630595192%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 670.0	
oro_x (USD): 1409.7	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 16.92	
Aluminio (USD/lb): 83.699	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.24	
Petroleo (USD): 111.0	
pib: 5.13	
Inflación, deflactor del PIB: series vinculadas (% anual): 1.91	
obra_civil_valor: 135411593.0	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.6	
IPC: 1.9399999999999997	
cobre_x (USD): 332.12	
Importaciones de bienes y servicios (% del PIB): 19.9	
valor_actual: 1417.83642	

- Caso # 4: UC LI211

Imagen 119. Resultados de actualización de los costos de la UC LI211 para el año 2014.

Actualización para UC 2014: LI211	Unidad Constructiva (UC)
valor actual de 2013: 4240000.0	Costo UC metodología actual año anterior
Actualización: 7336333007.8125	Costo UC Actualizado
Incremento: 172926.7218823703%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 7249505882.000001	Costo UC metodología actual
Error porcentual: 1.1976971565480694%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 550.0	
oro_x (USD): 1265.6	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 17.14	
Aluminio (USD/lb): 84.69	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.24	
Petroleo (USD): 58.91	
pib: 4.5	
Inflación, deflactor del PIB: series vinculadas (% anual): 2.24	
obra_civil_valor: 701375195.1	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.54	
IPC: 3.66	
cobre_x (USD): 311.26	
Importaciones de bienes y servicios (% del PIB): 20.85	
valor_actual: 7044.747953	
Variables seleccionadas para el modelo	

- Caso # 5: UC LI231

Imagen 20. Resultados de actualización de los costos de la UC LI231 para el año 2015.

Actualización para UC 2015: LI231	Unidad Constructiva (UC)
valor actual de 2014: 37960000.0	Costo UC metodología actual año anterior
Actualización: 3665132568.359375	Porcentaje de variación respecto al año inmediatamente anterior
Incremento: 9555.249126341872%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 3674933707.0000005	Costo UC metodología actual
Error porcentual: 0.26670246110715695%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 401.0	
oro_x (USD): 1159.0	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 23.43	
Aluminio (USD/lb): 75.331	
Inversión extranjera directa, entrada neta de capital (% del PIB): 3.96	
Petroleo (USD): 37.96	
pib: 2.9599999999999995	
Inflación, deflactor del PIB: series vinculadas (% anual): 2.45	
obra_civil_valor: 355542486.1999999	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.59	
IPC: 6.77	
cobre_x (USD): 249.23	
Importaciones de bienes y servicios (% del PIB): 22.71	
valor_actual: 3537.338875	
Variables seleccionadas para el modelo	

- Caso # 6: UC LI511

Imagen 21. Resultados de actualización de los costos de la UC LI511 para el año 2016.

Actualización para UC 2016: LI511	Unidad Constructiva (UC)
valor actual de 2015: 2090000.0000000007	Costo UC metodología actual año anterior
Actualización: 65194769531.25	Porcentaje de variación respecto al año inmediatamente anterior
Incremento: 3119266.963217702%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 65555899552.99999	Costo UC metodología actual
Error porcentual: 0.5508734137009736%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 620.0	
oro_x (USD): 1250.7	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 24.25	
Aluminio (USD/lb): 72.796	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.9	
Petroleo (USD): 57.75	
pib: 2.0900000000000007	
Inflación, deflactor del PIB: series vinculadas (% anual): 5.15	
obra_civil_valor: 6342402170.0	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.63	
IPC: 5.75	
cobre_x (USD): 220.56	
Importaciones de bienes y servicios (% del PIB): 21.49	
valor_actual: 62194.555679	

- Caso # 7: UC REA01

Imagen 22. Resultados de actualización de los costos de la UC REA01 para el año 2017.

Actualización para UC 2017: REA01	Unidad Constructiva (UC)
valor actual de 2016: 5130000.0	Costo UC metodología actual año anterior
Actualización: 3157983398.4375	Costo UC Actualizado
Incremento: 61459.13057383041%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 3182176398.0	Costo UC metodología actual
Error porcentual: 0.760265822400836%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 660.0	
oro_x (USD): 1257.8	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 20.46	
Aluminio (USD/lb): 89.3	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.39	
Petroleo (USD): 66.23	
pib: 1.3600000000000003	
Inflación, deflactor del PIB: series vinculadas (% anual): 5.13	
obra_civil_valor: 307869202.2	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.72	
IPC: 4.09	
cobre_x (USD): 279.68	
Importaciones de bienes y servicios (% del PIB): 20.14	
valor_actual: 3154.57981	

- Caso # 8: UC SE201

Imagen 23. Resultados de actualización de los costos de la UC SE201 para el año 2018.

Actualización para UC 2018: SE201	Unidad Constructiva (UC)
valor actual de 2017: 11097426990000.02	Costo UC metodología actual año anterior
Actualización: 1116436401.3671875	Costo UC Actualizado
Incremento: -99.99899396823933%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 1147044589.0	Costo UC metodología actual
Error porcentual: 2.668439215558036%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 723.0	
oro_x (USD): 1268.6	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 19.98	
Aluminio (USD/lb): 95.68	
Inversión extranjera directa, entrada neta de capital (% del PIB): 3.38	
Petroleo (USD): 54.44	
pib: 2.560000000000001	
Inflación, deflactor del PIB: series vinculadas (% anual): 4.63	
obra_civil_valor: 110974269.90000002	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.7	
IPC: 3.18	
cobre_x (USD): 295.88	
Importaciones de bienes y servicios (% del PIB): 20.64	
valor_actual: 1090.968013	

- Caso # 9: UC SE511

Imagen 24. Resultados de actualización de los costos de la UC SE511 para el año 2019.

Actualización para UC 2019: SE511	Unidad Constructiva (UC)
valor actual de 2018: 1570000.0	Costo UC metodología actual año anterior
Actualización: 149219863.89160156	Costo UC Actualizado
Incremento: 9404.449929401373%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 155254484.6	Costo UC metodología actual
Error porcentual: 3.88692199387742%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	
Acero (USD): 586.0	
oro_x (USD): 1392.3	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 21.08	
Aluminio (USD/lb): 81.244	
Inversión extranjera directa, entrada neta de capital (% del PIB): 4.33	
Petroleo (USD): 65.52	
pib: 3.1899999999999999	
Inflación, deflactor del PIB: series vinculadas (% anual): 4.0	
obra_civil_valor: 15020560.870000001	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.57	
IPC: 3.8	
cobre_x (USD): 272.14	
Importaciones de bienes y servicios (% del PIB): 21.69	
valor_actual: 148.8010625	

- Caso # 10: UC SE514

Imagen 25. Resultados de actualización de los costos de la UC SE514 para el año 2020.

Actualización para UC 2020: SE514	Unidad Constructiva (UC)
valor actual de 2019: 1609999.9999999998	Costo UC metodología actual año anterior
Actualización: 2483023925.78125	Costo UC Actualizado
Incremento: 154125.08855784163%	Porcentaje de variación respecto al año inmediatamente anterior
Valor real: 2508841597.0	Costo UC metodología actual
Error porcentual: 1.0290674090234324%	Error calculado entre el costo UC de referencia y el actualizado por el modelo
Los datos utilizados son	Variables seleccionadas para el modelo
Acero (USD): 1020.0	
oro_x (USD): 1770.1	
Exportaciones de productos manufacturados (% de las exportaciones de mercaderías): 22.4	
Aluminio (USD/lb): 77.293	
Inversión extranjera directa, entrada neta de capital (% del PIB): 2.76	
Petroleo (USD): 51.53	
pib: -7.19	
Inflación, deflactor del PIB: series vinculadas (% anual): 1.48	
obra_civil_valor: 242725406.8	
Importaciones de minerales y metales (% de importaciones de mercaderías): 1.72	
IPC: 1.6099999999999999	
cobre_x (USD): 280.35	
Importaciones de bienes y servicios (% del PIB): 20.53	
valor_actual: 2528.187841	

8. Conclusiones

- En este proyecto se ha propuesto una metodología para el desarrollo de un modelo de actualización de los costos de las UC acorde con las características de las variables macroeconómicas más significativas. Para lograr lo anterior, se recurrió a técnicas estadísticas de análisis multivariado tales como el análisis factorial con el fin de caracterizar la serie de costos de UC y su correlación con variables exógenas como los indicadores económicos.
- Como resultado del análisis multivariado propuesto, se lograron seleccionar variables relacionadas con el costo de los materiales (precio del acero, precio del cobre y precio del aluminio), PIB, IPC, exportaciones de productos manufacturados (% de las exportaciones de mercaderías), importaciones de minerales y metales (% de importaciones de mercaderías). Estas variables demostraron brindar información a los modelos sobre los efectos que tiene el comportamiento de los costos de los materiales de construcción de la UC, inflación (mano de obra), la logística de transporte e importación dentro del costo de cada UC.
- La base de datos de costos de la UC seleccionada para el proceso de análisis y entrenamiento de los modelos de actualización se obtuvo a partir de la integración de diferentes fuentes de costos a nivel nacional e internacional. Se evaluaron costos provenientes de reguladores internacionales como la de Perú y de referentes nacionales como las provistas por proveedores y regulaciones anteriores como la circular 090 de 2021, circular CREG 076 de 2024 y la circular CREG 092 de 2024s
- La metodología propuesta permitió establecer como variables significativas aquellas que mostraron un nivel de fiabilidad por encima de 0.8. Adicionalmente, el análisis factorial resultó adecuado para integrar aquellas variables con alto grado de

correlación entre ellas, permitiendo pasar de 14 variables a tan solo 4 factores a considerar dentro del modelo aquí propuesto.

- La posibilidad de incorporar alternativas como *redes neuronales* y *random forest* permitió evaluar cada conjunto de datos disponibles por cada UC y seleccionar aquellos modelos con mejor desempeño para cada caso. En este ejercicio se logró consolidar la experiencia de los diferentes actores que participaron dentro de este ejercicio de actualización, permitiendo integrar variables que recopilan la dinámica representativa en la variación de los costos de las UC.
- El parámetro de desempeño de los modelos *MAPE* (*Mean Absolute Percentage Error*) general mostrado por los modelos entrenados con el conjunto de datos de costos de las UC fue de 4.5%. Lo anterior evidencia la pertinencia de la metodología propuesta para el ajuste de los modelos para cada tipo de UC.
- Las variables macroeconómicas demostraron brindar información significativa acerca de la variabilidad de los costos de las UC (ver Imagen 14). Esto se reflejó en mayor medida en los años 2021, 2022 y 2023 en donde se presentó un incremento en los costos de las UC debido a efectos relacionados, posiblemente, por la pandemia y conflictos políticos internacionales que impactaron los precios de algunos commodities, costos de importación y supuestos considerados por el equipo consultor. Cabe resaltar que estos efectos no fueron objeto de análisis en este estudio. Aun cuando el error para esos tres años llegó a alcanzar un valor máximo de hasta el 8% respecto a los datos de costos de referencia, los modelos lograron seguir la tendencia de los datos. Lo anterior, resulta relevante en la medida que, valida la metodología como una herramienta flexible para establecer los valores de referencia de las UC, solucionando la asimetría de información del mercado que representa una restricción significativa para este tipo de estudios obteniendo unos resultados en un rango aceptable para los propósitos del regulador.
- Como resultado de este análisis propuesto, se logró la implementación de una metodología de actualización de costos que le permitirá a la CREG actualizar e incluir nuevas UC de manera ágil y sencilla. La aplicación permite el ensamble y entrenamiento de nuevos modelos para la actualización de nuevas UC con granularidad mensual y/o anual convirtiéndose en una característica clave para las funciones regulatorias de la comisión, logrando aumentar la frecuencia de la actualización de esta infraestructura.

9. Instalación e implementación de la aplicación para la actualización de costos de UC.

El procedimiento de instalación y puesta en marcha de la aplicación para la actualización de costos de UC se encuentra detallado en el documento “Anexo#1_Informe#3_Manual de Usuario_ModeloActualizacionUC.pdf”.

10. Referencias

- Aldhyani, T. H. H., Alrasheedi, M., Alqarni, A. A., Alzahrani, M. Y., & Bamhdi, A. M. (2020). Intelligent Hybrid Model to Enhance Time Series Models for Predicting Network Traffic. *IEEE Access*, 8, 130431–130451. <https://doi.org/10.1109/ACCESS.2020.3009169>
- Alves da Silva, A. P., Ferreira, V. H., & Velasquez, R. M. G. (2008). Input space to neural network based load forecasters. *International Journal of Forecasting*, 24(4), 616–629. <https://doi.org/10.1016/j.ijforecast.2008.07.006>
- Atiya, A. F., El-Shoura, S. M., Shaheen, S. I., & El-Sherif, M. S. (1999). A comparison between neural-network forecasting techniques - Case study: River flow forecasting. *IEEE Transactions on Neural Networks*, 10(2), 402–409. <https://doi.org/10.1109/72.750569>
- Awolusi, T. F., Oke, O. L., Akinkurolere, O. O., Sojobi, A. O., & Aluko, O. G. (2019). Performance comparison of neural network training algorithms in the modeling properties of steel fiber reinforced concrete. *Heliyon*, 5(1), e01115. <https://doi.org/10.1016/j.heliyon.2018.e01115>
- Azadeh, A. Ā., Ghaderi, S. F., & Sohrabkhani, S. (2008). A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. *Energy Policy*, 36, 2637–2644. <https://doi.org/10.1016/j.enpol.2008.02.035>
- Barbounis, T. G., Theocharis, J. B., Alexiadis, M. C., & Dokopoulos, P. S. (2006). Long-term Wind Speed and Power Forecasting Using Local Recurrent Neural Network Models. *IEEE Transactions on Energy Conversion*, 273–284.
- Bataineh, M., & Marler, T. (2017). Neural network for regression problems with reduced training sets. *Neural Networks*, 95, 1–9. <https://doi.org/10.1016/j.neunet.2017.07.018>
- Bauer, A., Zufle, M., Herbst, N., Zehe, A., Hotho, A., & Kounev, S. (2020). Time Series Forecasting for Self-Aware Systems. *Proceedings of the IEEE*, 108(7), 1068–1093. <https://doi.org/10.1109/JPROC.2020.2983857>
- Beccali, M., Cellura, M., Brano, V. Lo, & Marvuglia, A. (2004). Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management*, 45, 2879–2900. <https://doi.org/10.1016/j.enconman.2004.01.006>
- Bilbao, E. S. de I. de. (2008). *Redes Neuronales y Sus Aplicaciones*. EHU 2008/12. <http://ocw.ehu.es/enseñanzas-tecnicas/redes-neuronales-artificiales-y-sus-aplicaciones/contenidos/pdf/libro-del-curso>
- Calvetti, D., & Somersalo, E. (2007). Introduction to Bayesian Scientific Computing. In *Introduction to Bayesian Scientific Computing*. <https://doi.org/10.1007/978-0-387-73394-4>
- Chouikhi, N., & Alimi, A. M. (2018). *Adaptive Extreme Learning Machine for Recurrent Beta-basis Function Neural Network Training*.
- Dalvand, M. M., Bahram, S., Azami, Z., & Tarimoradi, H. (2011). Long-term Load Forecasting of Iranian Power Grid Using Fuzzy and Artificial Neural Networks. *Universities Power Engineering Conference, 2008*.
- Darbellay, G. A., & Slama, M. (2000). *Forecasting the short-term demand for electricity Do neural networks stand a better chance ?* 16, 71–83.

- Díaz Monroy, L. G. (2007). *Estadística multivariada: inferencia y métodos*. Universidad Nacional de Colombia. <https://repositorio.unal.edu.co/handle/unal/79907>
- Eberhart, R., & Shi, Y. (2007). *Computational Intelligence concepts to implementations* (pp. 17–25).
- Ebrahimi, M., Safari Sinegani, A. A., Sarikhani, M. R., & Mohammadi, S. A. (2017). Comparison of artificial neural network and multivariate regression models for prediction of Azotobacteria population in soil under different land uses. *Computers and Electronics in Agriculture*, 140, 409–421. <https://doi.org/10.1016/j.compag.2017.06.019>
- Errachdi, A., & Benrejeb, M. (2018). Performance Comparison of Neural Network Training Approaches in Indirect Adaptive Control. *International Journal of Control, Automation and Systems*, 16(3), 1448–1458. <https://doi.org/10.1007/s12555-017-0085-3>
- Fakhreddine O. Karray, & Silva, C. de. (2004). *Soft Computing and Intelligent Systems Design*.
- Fan, S., & Chen, L. (2006). Short-Term Load Forecasting Based on an Adaptive Hybrid Method. *IEEE Transactions on Power Systems*, 21(1), 392–401.
- Fernando, T. M. K. G., Maier, H. R., & Dandy, G. C. (2009). Selection of input variables for data driven models: An average shifted histogram partial mutual information estimator approach. *Journal of Hydrology*, 367(3–4), 165–176. <https://doi.org/10.1016/j.jhydrol.2008.10.019>
- Flórez, R., & Fernández, J. M. (2008). *Las Redes Neuronales Artificiales* (pp. 16–41). Netbiblo.
- G, H. S., Rastad, M., & Nazarzadeh, J. (2006). *A Hybrid Nonlinear Model for the Annual Maximum Simultaneous Electric Power Demand*. 21(3), 1069–1078.
- Gabriel I. S. Ruas. (2008). *Electrical Energy Demand Prediction Using Artificial Neural Networks and Support Vector Regression*.
- Garg, B., Kirar, N., Menon, S., & Sah, T. (2016). A performance comparison of different back propagation neural networks methods for forecasting wheat production. *CSI Transactions on ICT*, 4(2–4), 305–311. <https://doi.org/10.1007/s40012-016-0096-x>
- Godfrey, L. B., & Gashler, M. S. (2017). Neural decomposition of time-series data. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017, 2017-Janua(7)*, 2796–2801. <https://doi.org/10.1109/SMC.2017.8123050>
- Gracia, A. M., Delgado, A. V., Usón, A. A., Bribián, I. Z., & Scarpellini, S. (2006). *Disminución de Costes Energéticos en la Empresa* (F. Editorial, Ed.).
- Gu, L., Tok, D. K. S., & Yu, D. L. (2018). Development of adaptive p-step RBF network model with recursive orthogonal least squares training. *Neural Computing and Applications*, 29(5), 1445–1454. <https://doi.org/10.1007/s00521-016-2669-x>
- Hacibeyoglu, M., & Ibrahim, M. H. (2018). A Novel Multimean Particle Swarm Optimization Algorithm for Nonlinear Continuous Optimization: Application to Feed-Forward Neural Network Training. *Scientific Programming*, 2018. <https://doi.org/10.1155/2018/1435810>

- Hannah Jessie Rani, R., & Aruldoss Albert Victoire, T. (2018). Training radial basis function networks for wind speed prediction using PSO enhanced differential search optimizer. In *PLoS ONE* (Vol. 13, Issue 5). <https://doi.org/10.1371/journal.pone.0196871>
- Huang, S., Member, S., & Shih, K. (2003). Short-Term Load Forecasting Via ARMA Model Identification Including Non-Gaussian. *IEEE Transactions on Power Systems*, 18(2), 673–679.
- Jabbar, R. A., Junaid, M., Masood, M. A., Zaka, A., Jahangir, H., Rafique, A., & Kaleem, I. (2010). Neural Network (NN) based Demand Side Management (DSM). *Universities Power Engineering Conference (AUPEC), 2010 20th Australasian*.
- Jaramillo, J., Velasquez, J. D., & Franco, C. J. (2017). Research in Financial Time Series Forecasting with SVM: Contributions from Literature. *IEEE Latin America Transactions*, 15(1), 145–153. <https://doi.org/10.1109/TLA.2017.7827918>
- Jiang, S., Xiao, R., Wang, L., Luo, X., Huang, C., Wang, J.-H., Chin, K.-S., & Nie, X. (2019). Combining Deep Neural Networks and Classical Time Series Regression Models for Forecasting Patient Flows in Hong Kong. *IEEE Access*, 7, 118965–118974. <https://doi.org/10.1109/access.2019.2936550>
- Jimenez, J., Navarro, L., Quintero M., C. G., & Pardo, M. (2021). Multivariate statistical analysis for training process optimization in neural networks-based forecasting models. *Applied Sciences (Switzerland)*, 11(8). <https://doi.org/10.3390/app11083552>
- Jimenez, J., Pertuz, A., Quintero, C., & Montana, J. (2019). Multivariate statistical analysis based methodology for long-term demand forecasting. *IEEE Latin America Transactions*, 17(1), 93–101. <https://doi.org/10.1109/TLA.2019.8826700>
- Jimenez Mares, J., Charris, D., Pardo, M., & Quintero M, C. G. (2023). An Architecture to Improve Energy-Related Time-Series Model Validity Based on the Novel rMAPE Performance Metric. *IEEE Access*, 11, 36004–36014. <https://doi.org/10.1109/ACCESS.2023.3264713>
- Konar, A. (2005). *Computational Intelligence* (pp. 1–8). Elsevier.
- Kumar, R., Aggarwal, R. K., & Sharma, J. D. (2015). Comparison of regression and artificial neural network models for estimation of global solar radiations. *Renewable and Sustainable Energy Reviews*, 52, 1294–1299. <https://doi.org/10.1016/j.rser.2015.08.021>
- Li, C. H., Yang, L. T., & Lin, M. (2014). Parallel training of an improved neural network for text categorization. *International Journal of Parallel Programming*, 42(3), 505–523. <https://doi.org/10.1007/s10766-013-0245-x>
- Liang, R.-H., & Cheng, C.-C. (2000). Combined regression-fuzzy approach for short-term load forecasting. *IEE Proceedings - Generation, Transmission and Distribution*, 147(4), 261. <https://doi.org/10.1049/ip-gtd:20000507>
- Liang, Y., Jiang, J., Chen, Y., Zhu, R., Lu, C., & Wang, Z. (2019). Optimized Feedforward Neural Network Training for Efficient Brillouin Frequency Shift Retrieval in Fiber. *IEEE Access*, 7, 68034–68042. <https://doi.org/10.1109/ACCESS.2019.2919138>
- Lin, J. W., Chao, C. T., & Chiou, J. S. (2018). Determining Neuronal Number in Each Hidden Layer Using Earthquake Catalogues as Training Data in Training an Embedded

- Back Propagation Neural Network for Predicting Earthquake Magnitude. *IEEE Access*, 6, 52582–52597. <https://doi.org/10.1109/ACCESS.2018.2870189>
- Liu, F., Lu, Y., & Cai, M. (2020). A hybrid method with adaptive sub-series clustering and attention-based stacked residual LSTMs for multivariate time series forecasting. *IEEE Access*, 8, 62423–62438. <https://doi.org/10.1109/ACCESS.2020.2981506>
 - Liu, S., Gu, S., & Bao, T. (2017). An automatic forecasting method for time series. *Chinese Journal of Electronics*, 26(3), 445–452. <https://doi.org/10.1049/cje.2017.01.011>
 - Liu, S., Gu, S., & Peng, J. (2017). Self-adaptive Processing and Forecasting Algorithm for Univariate Linear Time Series. *Chinese Journal of Electronics*, 26(6), 1147–1153. <https://doi.org/10.1049/cje.2017.09.027>
 - Miguel, A., Gonza, E., & Carmona-ferna, D. (2007). *Forecasting of the electric energy demand trend and monthly fluctuation with neural networks*. 52, 336–343. <https://doi.org/10.1016/j.cie.2006.12.010>
 - Mizutani, A., Yukawa, T., Numa, K., & Kuze, Y. (2005). *Improvement of Input-Output Correlations of Electric Power Load Forecasting By Scatter Search*. 429–433.
 - Moreira de Andrade, L. C., & Nunes da Silva, I. (2009). Very Short-Term Load Forecasting Based on ARIMA Model and Intelligent Systems. *2009 15th International Conference on Intelligent System Applications to Power Systems*, 1–6. <https://doi.org/10.1109/ISAP.2009.5352829>
 - Moshkbar-Bakhshayesh, K. (2020). Performance study of bayesian regularization based multilayer feed-forward neural network for estimation of the uranium price in comparison with the different supervised learning algorithms. *Progress in Nuclear Energy*, 127(May), 103439. <https://doi.org/10.1016/j.pnucene.2020.103439>
 - Nagasaka, K., & Mamun, M. Al. (2004). *Long-term Peak Demand Prediction of 9 Japanese Power Utilities Using Radial Basis Function Networks*. 1–8.
 - Oğcu, G., Demirel, O. F., & Zaim, S. (2012). Forecasting Electricity Consumption with Neural Networks and Support Vector Regression. *Procedia - Social and Behavioral Sciences*, 58, 1576–1585. <https://doi.org/10.1016/j.sbspro.2012.09.1144>
 - OLANREWAJU. (2011). *Energy Efficiency Assessment Using Artificial Neural Network Combined with Data Envelopment Analysis*.
 - Olaya, J. (2007a). Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, 2, 187–202.
 - Olaya, J. (2007b). Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, 2, 187–202.
 - Oniscu, G. C., Adair, A., & Hidalgo, E. (2009). Severe preprocurement blunt trauma to the liver: Is there a need for back-table cutdown? *Liver Transplantation*, 15(11), 1647–1648. <https://doi.org/10.1002/lt.21880>
 - Pacheco, C., Sabino, L., Neto, R., Maria, M., Vellasco, B. R., Aure, M., & Souza, F. J. De. (2004). *Electric load forecasting: evaluating the novel hierarchical neuro-fuzzy BSP model*. 26, 131–142. [https://doi.org/10.1016/S0142-0615\(03\)00060-7](https://doi.org/10.1016/S0142-0615(03)00060-7)

- Pai, P., & Hong, W. (2005). Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74, 417–425. <https://doi.org/10.1016/j.epsr.2005.01.006>
- Park, D. C., El-Sharkawi, M. A., Marks, R. J., Atlas, L. E., & Damborg, M. J. (1991). Electric Load Forecasting Using an Artificial Neural Network. *IEEE Transactions on Power Systems*.
- Ramyar, S., & Kianfar, F. (2019). Forecasting Crude Oil Prices: A Comparison Between Artificial Neural Networks and Vector Autoregressive Models. *Computational Economics*, 53(2), 743–761. <https://doi.org/10.1007/s10614-017-9764-7>
- Rasouli, S., Tabesh, H., & Etminani, K. (2016). A Study of Input Variable Selection to Artificial Neural Network for Predicting Hospital Inpatient Flows. *British Journal of Applied Science & Technology*, 18(4), 1–8. <https://doi.org/10.9734/bjast/2016/30987>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
- Senjyu, T., Takara, H., Uezato, K., & Funabashi, T. (2002). One-Hour-Ahead Load Forecasting. *Trans. on Engineering, Computing and ...*, 17(1), 113–118.
- Song, K., Baek, Y., Hong, D. H., & Jang, G. (2005). *Short-Term Load Forecasting for the Holidays Using*. 20(1), 96–101.
- Tabaraki, R., & Khodabakhshi, M. (2020). Performance comparison of wavelet neural network and adaptive neuro-fuzzy inference system with small data sets. *Journal of Molecular Graphics and Modelling*, 100, 107698. <https://doi.org/10.1016/j.jmgm.2020.107698>
- Taylor, J. W., de Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1), 1–16. <https://doi.org/10.1016/j.ijforecast.2005.06.006>
- Tran, H. D., Muttill, N., & Perera, B. J. C. (2015). Selection of significant input variables for time series forecasting. *Environmental Modelling and Software*, 64, 156–163. <https://doi.org/10.1016/j.envsoft.2014.11.018>
- Wang, L., Yang, Y., Min, R., & Chakradhar, S. (2017). Accelerating deep neural network training with inconsistent stochastic gradient descent. *Neural Networks*, 93, 219–229. <https://doi.org/10.1016/j.neunet.2017.06.003>
- Wong, Y. J., Arumugasamy, S. K., & Jewaratnam, J. (2018). Performance comparison of feedforward neural network training algorithms in modeling for synthesis of polycaprolactone via biopolymerization. *Clean Technologies and Environmental Policy*, 20(9), 1971–1986. <https://doi.org/10.1007/s10098-018-1577-4>
- Xie, J., & Hong, T. (2018). Variable selection methods for probabilistic load forecasting: Empirical evidence from seven states of the United States. *IEEE Transactions on Smart Grid*, 9(6), 6039–6046. <https://doi.org/10.1109/TSG.2017.2702751>
- Yalcinoz, T., & Eminoglu, U. (2005). Short term and medium term power distribution load forecasting by neural networks. *Energy Conversion and Management*, 46, 1393–1405. <https://doi.org/10.1016/j.enconman.2004.07.005>

- Yang, Y., & Yang, Y. (2020). Hybrid method for short-term time series forecasting based on EEMD. *IEEE Access*, 8, 61915–61928. <https://doi.org/10.1109/ACCESS.2020.2983588>
- Yu, H., Ming, L. J., Sumei, R., & Shuping, Z. (2020). A Hybrid Model for Financial Time Series Forecasting-Integration of EWT, ARIMA with the Improved ABC Optimized ELM. *IEEE Access*, 8, 84501–84518. <https://doi.org/10.1109/ACCESS.2020.2987547>
- Zambri, N. A., Mohamed, A., & Wanik, M. Z. C. (2015). Performance comparison of neural networks for intelligent management of distributed generators in a distribution system. *International Journal of Electrical Power and Energy Systems*, 67, 179–190. <https://doi.org/10.1016/j.ijepes.2014.11.005>
- Zhang, G. P., & Kline, D. M. (2007). Quarterly time-series forecasting with neural networks. *IEEE Transactions on Neural Networks*, 18(6), 1800–1814. <https://doi.org/10.1109/TNN.2007.896859>
- Zheng, K., Qian, B., Li, S., Xiao, Y., Zhuang, W., & Ma, Q. (2020). Long-Short Term Echo State Network for Time Series Prediction. *IEEE Access*, 8, 91961–91974. <https://doi.org/10.1109/ACCESS.2020.2994773>