INSTITUTO TECNOLÓGICO DE CIUDAD MADERO

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

POSGRADO EN CIENCIAS DE LA INGENIERÍA

"POR MI PATRIA Y POR MI BIEN"

TESIS

ESTRATEGIAS DE DEEP LEARNING PARA PRONÓSTICO DE CAMBIO CLIMÁTICO

Que para obtener el grado de:
Doctor en Ciencias de la Ingeniería

Presenta:
M.C.C. Erick Estrada Patiño
D10070518
No. CVU: 740069

Directora de Tesis:
Dra. Guadalupe Castilla Valdez
No. CVU: 281385

Co-Director de Tesis:
Dr. Juan Frausto Solís
No. CVU: 31308

Ciudad Madero, Tamaulipas. Febrero 2026

Educación
Secretaría de Educación Pública

TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Ciudad Madero
Subdirección Académica
División de Estudios de Posgrado e Investigación

Ciudad Madero, Tamaulipas, 26/enero/2026

Oficio No.: U10.008/2026
Asunto: Autorización de Impresión

C. ERICK ESTRADA PATIÑO
No. DE CONTROL D10070518
P R E S E N T E

Me es grato comunicarle que después de la revisión realizada por el Jurado designado para su Examen de Grado de Doctor en Ciencias de la Ingeniería, el cual está integrado por los siguientes catedráticos:

PRESIDENTE :	DRA.	GUADALUPE CASTILLA VALDEZ
SECRETARIO:	DR.	JUAN FRAUSTO SOLÍS
VOCAL 1:	DR.	JUAN JAVIER GONZÁLEZ BARBOSA
VOCAL 2:	DRA.	SILVIA BEATRIZ BRACHETTI SIBAJA
VOCAL 3:	DR.	LUCIANO AGUILERA VÁZQUEZ
SUPLENTE:	DR.	RUBÉN SALAS CABRERA

Se acordó autorizar la impresión de su tesis titulada:

"ESTRATEGIAS DE DEEP LEARNING PARA PRONÓSTICO DE CAMBIO CLIMÁTICO "

Es muy satisfactorio para esta División compartir con Usted el logro de esta meta. Espero que continúe con éxito su desarrollo profesional y dedique su experiencia e inteligencia en beneficio de México.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Por Mi Patria y Por Mi Bien" ®

RAFAEL CASTILLO GUTIÉRREZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN

EDUCACIÓN
INSTITUTO TECNOLÓGICO DE CIUDAD MADERO
DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

ccp. Archivo
RCG/NRV/jar

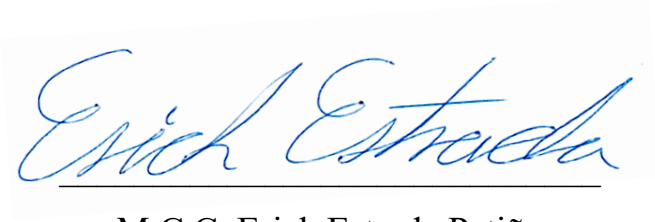
2026 año de Margarita Maza

DECLARACIONES DE ORIGINALIDAD, PROPIEDAD INTELECTUAL, CESION DE DERECHOS Y/O CONFIDENCIALIDAD

Declaro que este trabajo de investigación es original y ha sido realizado bajo mi autoría intelectual. Los contenidos aquí presentados no infringen derechos de terceros, y toda la información o datos tomados de otros autores han sido debidamente citados y referenciados en el apartado bibliográfico.

Por lo tanto, asumo la responsabilidad total sobre el contenido de este documento, deslindando de cualquier reclamación sobre propiedad intelectual a mis directores de tesis y al Instituto Tecnológico de Ciudad Madero.

Atentamente



M.C.C. Erick Estrada Patiño

AGRADECIMIENTOS

Agradezco profundamente al **Gobierno de México** por el apoyo económico brindado a través de la **Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI)**, así como al **Instituto Tecnológico de Ciudad Madero (ITCM)** y a su **División de Estudios de Posgrado e Investigación (DEPI)** por brindarme el espacio y las herramientas para culminar este grado.

A mi comité tutorial y mis maestros, quienes de manera directa o indirecta participaron en mi formación y en el desarrollo de este trabajo.

DEDICATORIA

A mis hijas, **Alicia Ailish** y **Bárbara Geany:** Ustedes son el motor que impulsa cada uno de mis sueños. Cada hora de estudio y cada esfuerzo en esta tesis fueron pensando en construir un mejor futuro para las dos. Gracias por ser mi luz, mi alegría y la razón por la que nunca me di por vencido. Espero que este logro les sirva de inspiración para saber que, con dedicación, no hay meta que no puedan alcanzar. Las amo con todo mi corazón.

A mis padres: Por su apoyo incondicional a lo largo de mi vida. Gracias por enseñarme el valor del trabajo y la perseverancia; este grado doctoral es también un fruto de su guía y sacrificio.

A mis compañeros de posgrado: Gracias por compartir las desveladas, los debates técnicos y la camaradería. El camino fue más ligero gracias a que compartimos la misma meta y los mismos retos.

Mención Especial: Expreso mi más profundo agradecimiento a la **Dra. Guadalupe Castilla Valdez** y al **Dr. Juan Frausto Solís**. Gracias por su invaluable guía, por compartir su sabiduría conmigo y por orientarme con paciencia y rigor académico durante todo este proceso. Sus consejos fueron fundamentales para dar forma a esta investigación.
A todos ustedes, **gracias totales.**

ESTRATEGIAS DE DEEP LEARNING PARA PRONÓSTICO DE CAMBIO CLIMÁTICO

M.C.C. Erick Estrada Patiño

Resumen

El presente trabajo aborda el problema del pronóstico de variables climáticas en el contexto del cambio climático, un fenómeno caracterizado por comportamientos altamente no lineales. El objetivo principal de este trabajo fue el diseño, implementación y evaluación de distintas estrategias de aprendizaje profundo y esquemas de ensamble optimizados, con el objetivo de mejorar la precisión y eficiencia de los modelos tradicionales y de otras estrategias de ensamble que han mostrado éxito en el área de pronóstico. Para ello, se desarrollaron cuatro metodologías: HELI, FORSEER, GEANY y TAE-Predict.

En cualquiera de las metodologías realizadas, se aplican enfoques modernos como los modelos de Deep learning y estrategias de ensamble de modelos de pronóstico, ponderadas con diferentes estrategias. Algunas de las estrategias propuestas se fundamentan en la descomposición de las series de tiempo en tendencia, estacionalidad y residuales, permitiendo un ajuste más eficiente de los modelos. En los enfoques multivariados se incorporó una estrategia de reducción de dimensionalidad basada en la unión de conjuntos, optimizando los tiempos de procesamiento sin afectar la calidad del pronóstico. La validación experimental se realizó con bases de datos climáticos de distintas zonas metropolitanas de México, bajo un esquema de evaluación estadística riguroso.

Los resultados muestran que las metodologías HELI y FORSEER, alcanzan un desempeño comparable al modelo SMYL, referente en competencias internacionales. Además, los modelos desarrollados presentan mayor estabilidad y menores tiempos de entrenamiento, incluso ante series de alta complejidad, lo que los posiciona como alternativas eficientes para el análisis y pronóstico de variables climáticas.

DEEP LEARNING STRATEGIES FOR CLIMATE CHANGE FORECASTING

M.C.C. Erick Estrada Patiño

Abstract

This paper addresses the problem of forecasting climate variables in the context of climate change, a phenomenon characterized by highly nonlinear behavior. The main objective of this work was to design, implement, and evaluate different deep learning strategies and optimized ensemble schemes, with the aim of improving the accuracy and efficiency of traditional models and other ensemble strategies that have proven successful in the area of forecasting. To this end, four methodologies were developed: HELI, FORSEER, GEANY, and TAE-Predict.

In each of the methodologies, modern approaches such as deep learning models and ensemble strategies for forecasting models are applied, weighted with different strategies. Some of the proposed strategies are based on the decomposition of time series into trend, seasonality, and residuals, allowing for more efficient model adjustment. In the multivariate approaches, a dimensionality reduction strategy based on set union was incorporated, optimizing processing times without affecting forecast quality. Experimental validation was performed with climate databases from different metropolitan areas in Mexico, under a rigorous statistical evaluation scheme.

The results show that the HELI and FORSEER methodologies achieve performance comparable to the SMYL model, a benchmark in international competitions. In addition, the models developed offer greater stability and shorter training times, even for highly complex series, positioning them as efficient alternatives for the analysis and forecasting of climate variables.

Índice General

Índice Tablas

Índice de Figuras

1 Introducción

El clima se considera una serie de indicadores y condiciones naturales, biológicas, meteorológicas, cósmicas y geográficas, entre otras, que constituyen el entorno biológico para el desarrollo, crecimiento, extinción o evolución de las especies. Estas condiciones están estrechamente relacionadas con regiones específicas, y a lo largo de la extensión terrestre se observa una gran variedad de entornos climáticos [1], [2].

Estos indicadores climáticos están altamente correlacionados entre sí. Las alteraciones en uno de ellos tienen un impacto directo en otros, lo que a su vez repercute en el indicador original [3]. Este ciclo genera una red compleja de causalidad, donde las variaciones atípicas pueden provocar anomalías a largo plazo, afectando directamente a las especies que se desarrollan en esos entornos [4].

A lo largo de la historia de nuestro planeta, este ha experimentado importantes ajustes climatológicos, favoreciendo en algunos casos el desarrollo de la vida y, en otros, contribuyendo a su extinción[4], [5], [6], [7]. Por lo tanto, el cambio climático es un fenómeno inherente a la Tierra, con ajustes que van desde eventos inmediatos hasta procesos a largo plazo. Estos fenómenos son multifactoriales y, en gran medida, inevitables debido a su naturaleza, con un impacto significativo en el desarrollo natural de las especies [8], [9].

A lo largo de la historia humana, estos fenómenos solían ser producto de situaciones naturales y sus impactos, en la mayoría de los casos, eran focalizados [10]. Sin embargo, a partir de la Revolución Industrial y la expansión de los asentamientos humanos sobre grandes áreas, ciertos indicadores climáticos comenzaron a mostrar variaciones con tendencias más evidentes. Desde entonces, las alteraciones climáticas se han intensificado y globalizado, manifestándose en fenómenos meteorológicos abruptos y violentos que afectan a toda forma de vida en la Tierra [11].

De acuerdo con la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC), los fenómenos contemporáneos del cambio climático tienen su origen en

actividades antropogénicas, especialmente aquellas relacionadas con el uso de combustibles fósiles como el petróleo y el carbón [12], [13]. Esto se refleja principalmente en variables como las emisiones de gases de efecto invernadero y en cambios en las variables de temperatura y condiciones atmosféricas directamente relacionadas con estos gases [14].

Otros organismos internacionales, como el Panel Intergubernamental sobre Cambio Climático (IPCC) y la Organización de las Naciones Unidas (ONU), reconocen la naturaleza multifactorial del cambio climático, pero coinciden en señalar la temperatura como un indicador clave [15]. En la Conferencia de las Partes (COP) de París en 2015, se estableció como meta limitar el aumento de la temperatura global, dada su relevancia en la dinámica del cambio climático [16]. El IPCC menciona que el incremento de los efectos asociados al cambio climático puede tener un impacto significativo en el desarrollo natural de las sociedades, particularmente en aquellas que son vulnerables. Procesos esenciales como la agricultura, la ganadería, el desarrollo de cadenas de suministro, y la obtención de energía y agua, se ven gravemente afectados, lo que ocasiona pérdidas económicas, sociales y humanas [17], [18], [19].

Por ello, este trabajo propone diversas técnicas de análisis de datos relacionados con el cambio climático, representados como series de tiempo. El objetivo es contribuir a la mejora de los modelos de pronóstico y apoyar en la toma de decisiones frente a estos fenómenos. Estas técnicas incluyen enfoques modernos de análisis de datos que permiten descomponer patrones temporales, estacionales en las variables clave y seleccionar aquellas variables que resultan más significativas en los modelos, lo que resulta fundamental para anticipar los impactos y gestionar riesgos asociados al cambio climático.

1.1 Planteamiento del problema

La predicción climática ha representado un reto para la humanidad moderna, cuyo objetivo es prever diferentes escenarios climáticos que puedan afectar el desarrollo social. Esta predicción se realiza mediante modelos matemáticos y meteorológicos altamente complejos**,** que basan sus estimaciones en variables atmosféricas actuales y pasadas para emitir un pronóstico. En general, la calidad del pronóstico de estos modelos es adecuada a corto plazo;

sin embargo, en el mediano y largo plazo pierden precisión, ofreciendo resultados más genéricos. Esto se debe, en gran medida, a la imprevisibilidad y alta volatilidad de los fenómenos asociados al cambio climático [20].

Si bien existe literatura que sugiere que la información para hacer pronósticos de una variable se encuentra contenida dentro de la misma variable, la exclusión de variables relacionadas con fenómenos antropogénicos en los modelos matemáticos puede impactar negativamente la precisión de las predicciones[21], [22]. Algunos modelos más recientes han comenzado a incluir estas variables, pero su incorporación aumenta significativamente la complejidad de los modelos, lo que dificulta tanto el proceso de pronóstico como incrementa el tiempo de cálculo [23], [24].

Por otra parte, se han desarrollado modelos estadísticos genéricos que se aplican en diversas áreas de predicción. A pesar de su naturaleza generalista, estos modelos han demostrado alta adaptabilidad y, en general, ofrecen buen desempeño en términos de pronóstico. Técnicas como las propuestas por Holt, Winters, y los enfoques de Box y Jenkins proporcionan resultados competitivos con un costo computacional bajo y requieren menos datos para el ajuste de los modelos [25], [26].

Aunado a lo anterior, en el campo del aprendizaje automático (machine learning) se han desarrollado múltiples estrategias que permiten el análisis de grandes volúmenes de datos, incluso cuando estos son incompletos o presentan inconsistencias, como una alta cantidad de valores atípicos [27]. Los modelos de pronóstico o de selección de variables en el aprendizaje automático permiten el descubrimiento de patrones y la generalización de la información, adaptándose a los datos para generar resultados más robustos y competitivos [28], [29]. Aunque estas técnicas pueden tener una mayor complejidad, la evolución de las tecnologías de procesamiento de datos y la optimización de los modelos las posiciona como alternativas viables, competitivas y confiables [30], [31], [32].

En este trabajo de investigación se propone el desarrollo de enfoques de ensamble para el pronóstico de indicadores del cambio climático, combinando técnicas avanzadas de análisis

y selección de series de tiempo con métodos de aprendizaje automático. El objetivo es obtener resultados competitivos con los reportados en el estado del arte, contribuyendo al avance del conocimiento sobre los fenómenos climáticos. A través de este enfoque, se busca minimizar los efectos negativos del cambio climático y, al mismo tiempo, promover la corrección de comportamientos dañinos hacia el medio ambiente.

1.2 Objetivo general

Desarrollar un modelo para pronóstico de cambio climático basado en métodos de aprendizaje automático estadísticamente superior o equivalente a los reportados en el estado del arte.

1.3 Objetivos específicos

- Realizar una revisión exhaustiva de las técnicas de aprendizaje profundo, soft-computing, y métodos clásicos para pronóstico.
- Desarrollar métodos de pronóstico: clásicos, machine learning y/o de soft-computing.
- Estudiar y analizar técnicas de aprendizaje basadas en ensamble para seleccionar las mejores estrategias.
- Desarrollar una nueva técnica de aprendizaje basada en ensamble para mejorar el desempeño de las metodologías de la literatura.
- Analizar y evaluar diferentes métricas de desempeño.
- Evaluar experimentalmente los métodos desarrollados utilizando técnicas estadísticas.
- Realizar un análisis del estado del cambio climático para México y regiones similares y contrastantes de acuerdo con métricas de riesgo climático, desarrollo industrial y densidad demográfica.

1.4 Justificación del estudio

El pronóstico de series de tiempo ha tomado gran relevancia en diversas áreas donde es aplicable, ya que permite hacer estimaciones precisas sobre fenómenos específicos. En el ámbito climático, adquiere especial importancia cuando se combina con estrategias de

pronóstico meteorológico, ya que puede contribuir a mejorar la calidad de las predicciones, especialmente en escenarios donde los fenómenos de cambio climático están presentes.

En el pasado, las técnicas de aprendizaje automático estaban limitadas a fines experimentales, con resultados que eran inferiores frente a estrategias estadísticas, las cuales resultaban menos costosas en términos de tiempo y recursos computacionales, y ofrecían mejores resultados en ciertos contextos. Esto se debe a que los modelos de aprendizaje automático requieren grandes volúmenes de información y demandan recursos computacionales significativos para lograr tiempos de cómputo razonables [33].

Con la llegada de nuevas técnicas de procesamiento de información, el aumento constante en la velocidad de los procesadores y el desarrollo del cómputo acelerado, junto con la evolución de las estrategias de aprendizaje automático especializadas en distintos tipos de problemas, el aprendizaje automático se ha consolidado como una opción viable debido a su capacidad de generalizar la información. Estos modelos se ajustan de manera más precisa a las curvas de pronóstico, realizando cálculos más complejos en menos tiempo [34], [35], [36].

Gracias a este incremento en la capacidad de procesamiento, las técnicas de aprendizaje profundo juegan un papel fundamental en la ciencia de datos, incentivando el desarrollo de nuevas y mejores técnicas de clasificación y pronóstico de modelos. Aprovechar los recursos computacionales actuales y combinar estas técnicas con métodos clásicos de pronóstico puede resultar en un proceso de aprendizaje más eficiente, preciso y rápido, lo que permitirá obtener mejores resultados en procesos de clasificación y predicción.
Enfocar estos esfuerzos en la detección de patrones, causales y agrupamientos de información, así como en la predicción de fenómenos relacionados con el cambio climático, facilitará una comprensión más profunda de estos eventos y de su impacto, permitiendo anticiparse a sus efectos y proponer estrategias de mitigación más efectivas.

1.5 Organización de la tesis

El presente documento se organiza en cinco capítulos y una sección de anexos, cuya estructura responde a una progresión lógica desde la contextualización del problema hasta la discusión de los resultados y las conclusiones finales.

En el Capítulo 1 se introduce de manera general la problemática que motiva este estudio. Se describen las técnicas propuestas para su abordaje, se formulan los objetivos generales y específicos de la investigación y se delimitan las principales restricciones y alcances que condicionan el desarrollo del trabajo.

El Capítulo 2 desarrolla el marco teórico que sustenta la investigación. En esta sección se presentan los conceptos fundamentales necesarios para la comprensión del modelo propuesto y se incluye una revisión crítica de los trabajos relacionados más relevantes, los cuales han influido de manera directa en las decisiones metodológicas y en el diseño de la solución planteada.

En el Capítulo 3 se expone la metodología empleada. Se describe con detalle la arquitectura del modelo propuesto, así como la naturaleza, estructura y características de los datos de entrada y salida. Asimismo, se especifican los formatos utilizados y los procedimientos necesarios para la correcta implementación y reproducción del modelo.

El Capítulo 4 está dedicado al análisis experimental y a la presentación de resultados. En él se documenta el proceso de experimentación, se reportan los resultados obtenidos y se ofrece una interpretación detallada de los mismos. Además, se presentan las pruebas estadísticas e hipótesis empleadas para comparar el desempeño del modelo propuesto con otros enfoques representativos del estado del arte.

Finalmente, en el Capítulo 5 se presentan las conclusiones derivadas del estudio. Se evalúa el grado de cumplimiento de los objetivos planteados, se discuten las principales aportaciones del trabajo y se proponen posibles líneas de investigación futura que permitan extender o mejorar los resultados obtenidos.

La sección de Anexos complementa el documento mediante la inclusión de una liga a un repositorio que contiene el código o pseudocódigo utilizado para la obtención de los resultados presentados, con el objetivo de garantizar la transparencia y replicabilidad del estudio.

Capítulo 2

2 Marco Teórico

A continuación, se detallan, de forma general, las principales estrategias y modelos utilizados o que sirvieron como inspiración para nuestra propuesta metodológica. El objetivo es ofrecer un contexto general del funcionamiento de estos, proporcionando una base teórica sólida para comprender la justificación detrás de las decisiones metodológicas adoptadas en este trabajo. Se abarcarán tanto técnicas clásicas como métodos más recientes basados en aprendizaje automático, permitiendo así una visión completa de las herramientas disponibles para el análisis y pronóstico de series de tiempo.

Estos métodos han sido seleccionados y descritos tomando como referencia los avances más relevantes y recientes de la literatura. En particular, se destacan aquellos que forman parte del estado del arte y han demostrado un desempeño notable en problemas similares al abordado en esta investigación.

2.1 Clima

De acuerdo con la literatura, el clima se define como el conjunto de condiciones atmosféricas, geográficas y naturales promedio que caracterizan una región determinada a lo largo de un periodo prolongado de tiempo.[23] En otras palabras, el clima está conformado por un conjunto de variables atmosféricas, como la temperatura, la presión, la humedad, el viento, la nubosidad y el rocío, entre otras, que se mantienen relativamente estables a lo largo de los años dentro de una región geográfica con características similares. [21]

Es importante establecer la diferencia entre clima y tiempo atmosférico. Mientras el clima representa un promedio estadístico de largo plazo, el tiempo atmosférico describe el estado momentáneo o a corto plazo de la atmósfera, pudiendo incluir las mismas variables, pero observadas en periodos cortos o inmediatos. El tiempo atmosférico es altamente sensible a variaciones estacionales o eventos puntuales, como tormentas, frentes fríos o fenómenos de escala global, por ejemplo, El Niño y La Niña. En contraste, el clima, al definirse en escalas temporales mucho más extensas, ofrece un marco de referencia general y estable, que permite comprender cómo estos eventos se insertan dentro de un comportamiento promedio a largo plazo. [20]

2.1.1 Cambio climático

El hecho de que el clima no sea sensible a eventos puntuales no implica que permanezca estable a lo largo del tiempo. El clima de las distintas regiones del planeta evoluciona de manera continua, aunque sus transformaciones suelen ser graduales y con impactos pequeños, lo que permite que los entornos naturales se adapten progresivamente mediante mecanismos evolutivos. Este proceso de cambio lento y sostenido de las condiciones climáticas recibe el nombre de cambio climático.

El cambio climático es un fenómeno multifactorial, es decir, puede originarse por diversos factores que actúan de manera conjunta o independiente. Entre ellos se encuentran los factores geográficos, como la modificación de la posición o altitud de una región; los factores cósmicos, como las variaciones en la radiación solar o la actividad volcánica; y los factores naturales, que incluyen procesos biológicos como la migración o expansión de determinadas especies. No obstante, en la actualidad, la evidencia científica ha demostrado que las principales causas del cambio climático contemporáneo están estrechamente vinculadas a la actividad humana o antropogénica. [15], [23]

Las actividades humanas, especialmente el uso intensivo e indiscriminado de recursos naturales y la emisión de gases de efecto invernadero, han acelerado los procesos de transformación climática. A diferencia de los cambios naturales, que ocurren en escalas de tiempo muy amplias, los cambios inducidos por la acción humana se desarrollan con una

velocidad inusitada, generando alteraciones bruscas y eventos extremos en el sistema climático global.

2.1.2 Indicadores climáticos

El clima está conformado por una amplia variedad de variables que interactúan entre sí. Entre ellas se encuentran variables o factores geográficos, atmosféricos, estacionales, naturales, de origen humano, entre otros. Estas variables en conjunto determinan el comportamiento general del clima en una región específica. Cuando alguna de estas variables cambia, puede afectar no solo a otras variables, sino también al equilibrio climático en su totalidad, ya sea de forma directa o a través de cadenas de efectos indirectos. [21]

En el estudio del cambio climático, estas variables pueden considerarse como indicadores clave. Su análisis permite entender cómo se comportan otras variables relacionadas, hacer pronósticos sobre su evolución y obtener una visión más clara de la situación climática de una región determinada.

En este trabajo se emplean indicadores reportados por agencias climáticas gubernamentales o aeroportuarias. A partir de estos datos, se realiza una selección estadística para identificar aquellos indicadores que tienen mayor influencia dentro del conjunto analizado. Aunque existen muchos otros indicadores relevantes para el estudio del cambio climático, su inclusión queda fuera del alcance de este trabajo debido a limitaciones en la disponibilidad y acceso a los datos.

2.2 Series de tiempo

Las series de tiempo son un conjunto de observaciones cuantitativas sobre un fenómeno específico. Cuando estas observaciones cumplen ciertas características y presentan autocorrelación, se considera que los datos conforman una serie de tiempo. [37], [38]

Las series de tiempo deben estar compuestas por observaciones registradas en intervalos de tiempo regulares y presentan una relación fuerte con los valores previos de la serie. Sin

embargo, esto no implica necesariamente un comportamiento lineal, ya que puede existir cierto grado de aleatoriedad en cada observación; no obstante, la autocorrelación sigue siendo el factor predominante en la estructura de los datos.

El análisis de series de tiempo permite representar fenómenos, identificar patrones o comportamientos y realizar predicciones sobre su evolución futura. [39] En términos generales, una serie de tiempo puede estar compuesta por tres principales patrones de comportamiento: tendencia, estacionalidad y componente residual.

Las series de tiempo pueden presentar comportamientos más complejos, como cambios estructurales o comportamientos atípicos. Un cambio estructural ocurre cuando la relación entre las observaciones se altera significativamente en un punto determinado, lo que puede deberse a factores externos como crisis económicas, políticas de gobierno, avances tecnológicos, Sin embargo, las series de tiempo pueden presentar comportamientos más complejos, como cambios estructurales o valores atípicos. Un cambio estructural ocurre cuando la relación entre las observaciones se altera significativamente en un punto determinado, lo que puede deberse a factores externos como crisis económicas, políticas de gobierno, avances tecnológicos o fenómenos naturales. Por otro lado, los valores atípicos son observaciones que pierden completamente la autocorrelación con la serie y no se ajustan a ninguno de los componentes descritos anteriormente. Estos valores pueden deberse a eventos extremos o errores en la medición y recolección de datos, lo que puede afectar la calidad del análisis y la precisión de los modelos de pronóstico.

2.2.1 Tendencia

La tendencia en las series de tiempo puede definirse como el comportamiento general que esta presenta a lo largo del tiempo. Es importante considerar que una serie de tiempo puede mostrar fluctuaciones inherentes al fenómeno que describe, como las variaciones estacionales. [37] Sin embargo, más allá de estas fluctuaciones, la serie puede mostrar una dirección predominante: un aumento, una disminución o mantenerse relativamente estable. Esto puede observarse en la Figura 2.1, donde se muestra una serie de tiempo con una tendencia creciente.

Existen tres tipos de tendencia. Una tendencia positiva indica un incremento general en los valores de la serie a lo largo del tiempo, mientras que una tendencia negativa representa una disminución progresiva en dichos valores. Cuando la serie de tiempo no muestra un cambio significativo y sus valores se mantienen alrededor de un nivel constante, se dice que tiene un comportamiento estacionario.

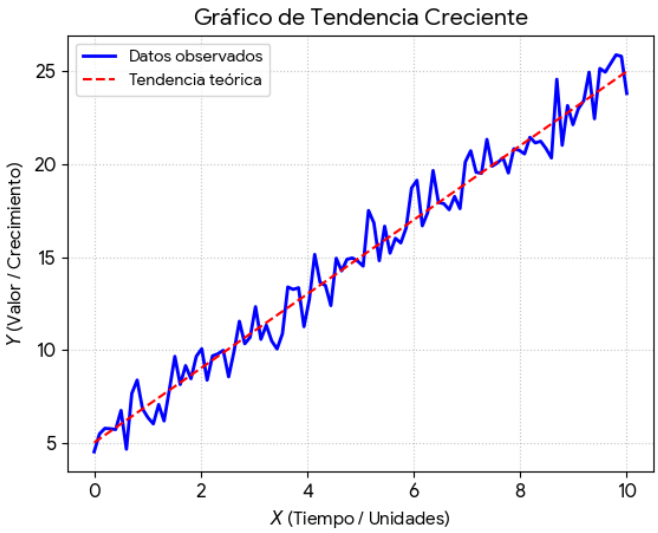


Figura 2.1. Comportamiento de tendencia creciente en una serie de tiempo

Es importante destacar que la tendencia no siempre es lineal, ya que en muchos casos puede presentar comportamientos más complejos, como tendencias exponenciales o logarítmicas. Además, una serie de tiempo puede exhibir diferentes tendencias en distintos períodos, lo que se conoce como tendencia cambiante.

2.2.2 Estacionalidad

En el análisis de series de tiempo, dependiendo del fenómeno que se estudie, pueden observarse patrones o comportamientos que se repiten de forma periódica. [37] Esta periodicidad depende del contexto del problema y puede presentarse en intervalos que van desde segundos hasta varios años. A este comportamiento se le conoce como estacionalidad.

La estacionalidad puede definirse como un patrón recurrente y predecible dentro de una serie temporal. Dicho patrón se repite con cierta regularidad y puede mantenerse estable o mostrar ligeros cambios con el paso del tiempo. En series con alta variabilidad, la identificación de estos patrones puede resultar más difícil, ya que el ruido o las fluctuaciones pueden ocultar su estructura real.

Cuando el patrón estacional se mantiene constante en el tiempo, es decir, no cambia su amplitud ni su forma, se dice que existe una estacionalidad aditiva. En este tipo de series, el componente estacional se suma al comportamiento general sin aumentar ni disminuir su intensidad. Un ejemplo común son las estaciones del año, donde las variaciones climáticas suelen repetirse con una magnitud similar cada año.

Por otro lado, cuando el patrón estacional conserva su forma y su frecuencia, pero aumenta o disminuye su amplitud con el tiempo, se habla de una estacionalidad multiplicativa. Este tipo de estacionalidad suele ser más compleja, ya que el efecto estacional depende del nivel de la serie. Un ejemplo típico se presenta en los ciclos económicos, donde las variaciones estacionales tienden a amplificarse o reducirse según la inflación o el crecimiento económico.

2.2.3 Residuales

Son valores que no se ajustan a los otros dos componentes. Este componente está integrado por la aleatoriedad presente en el fenómeno. Dicha aleatoriedad puede deberse a múltiples factores impredecibles, como fluctuaciones espontáneas en el sistema, errores en la medición o la presencia de eventos atípicos que no siguen un patrón definido. [40] Aunque la aleatoriedad es inherente a muchas series de tiempo, es crucial diferenciar entre variabilidad natural y anomalías que podrían distorsionar el análisis y que pueden ser corregidas con técnicas de preprocesamiento de los datos.

2.2.4 Descomposición de series de tiempo

Una de las estrategias más utilizadas en el análisis de series de tiempo es la descomposición de tendencia y estacionalidad [37], [41] . Este enfoque permite extraer de la serie temporal

el componente de tendencia y el componente estacional, dejando como residuo el ruido presente en la serie.

$$y_t = S_t + T_t + R_t \quad (2.1)$$

$$y_t = S_t \times T_t \times R_t \quad (2.2)$$

Dependiendo del comportamiento de la serie, esta puede descomponerse utilizando las ecuaciones (2.1) y (2.2) que describen los enfoques aditivo y multiplicativo, respectivamente; donde y_t corresponde al valor de la serie de tiempo, S_t es el valor del componente estacional, T_t representa el componente de tendencia y R_t corresponde al valor del componente residual, todos para un tiempo específico t [37]. El enfoque aditivo es adecuado para series cuyos cambios y estacionalidad se mantienen constantes a lo largo del tiempo, lo que significa que, aunque haya variaciones estacionales, estas se repiten con la misma magnitud a lo largo del tiempo. Por otro lado, los modelos multiplicativos son aplicables cuando las variaciones estacionales crecen o decrecen en proporción a la tendencia, es decir, cuando las fluctuaciones estacionales aumentan o disminuyen junto con la tendencia [42].

En la Figura 2.2 se muestra el proceso de descomposición de una serie temporal, donde el resultado son tres componentes más simples (tendencia, estacionalidad y residuo), que juntas forman la serie original. Este proceso requiere un análisis de la amplitud estacional, generalmente realizado mediante el cálculo de los coeficientes de autocorrelación, un suavizamiento para extraer la tendencia, y la extracción de la estacionalidad mediante medias estacionarias [43].

La determinación del periodo estacional se puede hacer calculando los coeficientes de autocorrelación. A medida que el desfase (*lag*) aumenta, el valor del coeficiente de correlación tiende a disminuir. Sin embargo, cuando la serie presenta estacionalidad, el coeficiente vuelve a aumentar en cierto punto. El valor más alto antes de que el coeficiente

vuelva a decrecer corresponde al periodo estacional, y la cantidad de periodos entre ese intervalo cerrado determina el tamaño de la estacionalidad.

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{t=1}^{T}(y_t-\bar{y})^2} \quad (2.3)$$

La ecuación (2.3) muestra el cálculo de los coeficientes de correlación [44]. Donde T es la cantidad de observaciones totales en la serie de tiempo, k es el valor de lag o rezago, $\bar{y}$ representa la media de la serie de tiempo, por último y representa cada observación de la serie de tiempo en el índice que corresponda.

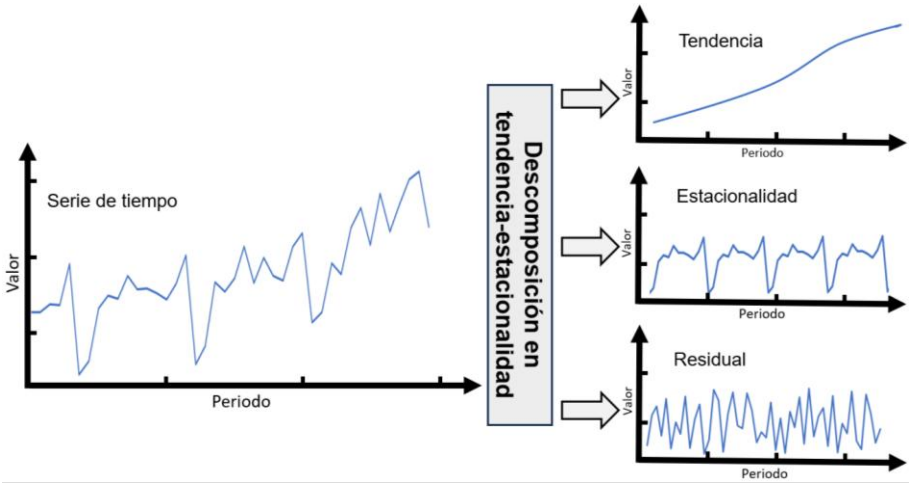


Figura 2.2. Proceso de descomposición en componentes de tendencia-estacionalidad

Luego de determinar el periodo estacional, es necesario decidir si la serie es aditiva o multiplicativa. Esto se puede hacer analizando la varianza de los datos en relación con el componente estacional. Si la varianza permanece constante, la serie es aditiva. Si la varianza aumenta o disminuye con el tiempo, estamos frente a una serie multiplicativa.

$$SMA = \frac{A_1+A_2+A_3+A_n}{n} \quad (2.4)$$

Con el periodo estacional y el tipo de serie determinados, es posible realizar la descomposición [45]. El primer componente a extraer es la tendencia, lo cual se realiza

mediante una técnica de medias móviles simples (SMA), como se describe en la ecuación (2.4), donde A_n es una observación en el periodo n, el cual representa la amplitud del periodo estacional.

$$Y'_{des} = (Y_s + Y_R) - Y_T \quad (2.5)$$

$$Y'_{des} = \frac{(Y_s * Y_R)}{Y_T} \quad (2.6)$$

Como resultado de esta técnica, se obtiene una curva suavizada que representa el componente de tendencia de la serie. Este componente es extraído de la serie original, y dependiendo del tipo de serie, se aplican la ecuación (2.5) para series aditivas o la ecuación (2.6) para series multiplicativas; donde Y'_{des} es la serie de tiempo desestacionalizada, Y_s es el componente estacional, Y_R es el componente residual y Y_T representa al componente de tendencia

$$S_n = \frac{1}{k}\sum_{i=1}^{k} Y'_{des_{i*t}} \quad (2.7)$$

La serie resultante Y'_{des} ya no contiene el componente de tendencia; está formada únicamente por la estacionalidad y el ruido. Para extraer el componente estacional, se utiliza la técnica de medias estacionales, descrita en la ecuación (2.7), donde S_n es el valor del componente estacional en el periodo n, k es la cantidad de bloques estacionales en la serie, y t es la longitud del periodo estacional.

$$Y_R = Y'_{des} - Y_s \quad (2.8)$$

$$Y_R = \frac{Y'_{des}}{Y_s} \quad (2.9)$$

Calculado un bloque estacional, este se replica tantas veces como k bloques tenga la serie. Después de obtener el componente estacional, el siguiente paso es obtener el componente residual que no se ajusta a los componentes anteriores. Esto se realiza con las ecuaciones (2.8) y (2.9), correspondientes a series aditivas y multiplicativas, respectivamente [44].

2.2.5 Preprocesamiento de series de tiempo

Los datos provenientes de fuentes reales suelen presentar diversas inconsistencias, lo cual se debe a la naturaleza de los datos, la dificultad para obtener las observaciones, errores de captura, fallos humanos, fallos de máquinas, sensores descalibrados, entre otros tipos de problemas [46], [47]. Por ello, es fundamental aplicar técnicas de preprocesamiento de la información, que permiten adecuar los datos a un intervalo apropiado para su procesamiento, rellenar o eliminar observaciones con información faltante y eliminar outliers o datos atípicos. El resultado del preprocesamiento es una serie de tiempo ajustada a un intervalo definido y suavizada.

Este proceso mejora la calidad de la información, lo que a su vez incrementa la eficiencia de los algoritmos o modelos de predicción, selección o descomposición que se aplican posteriormente [46].

2.2.5.1 Imputación de datos faltantes

Los datos faltantes son aquellos que no están presentes en la serie de tiempo; esto puede deberse a errores en procesos humanos o fallos en los sistemas de captura de datos [48]. Los datos faltantes pueden convertirse en un problema cuando se agrupan en grandes cantidades, ya que esto podría afectar los patrones estacionales, dificultando los procesos de descomposición, selección y pronóstico. Eliminar los periodos con datos faltantes puede generar problemas en fases futuras del análisis, por lo que la opción más adecuada es la imputación de los datos, proceso mediante el cual se calcula la información faltante.

Existen diversas técnicas para imputar datos faltantes, las cuales se aplican según las características de la serie de tiempo y la cantidad de datos ausentes [49]. Estas técnicas pueden ser tan simples como el uso del promedio de los periodos circundantes, o más complejas, como la aplicación de métodos de pronóstico. En este trabajo de investigación se describen la interpolación lineal y la interpolación cuadrática.

$$y = y_0 + \frac{y_1 - y_0}{t_1 - t_0} * (t - t_0) \quad (2.10)$$

$$y = \frac{y_0(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)} + \frac{y_1(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)} + \frac{y_2(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)} \quad (2.11)$$

La interpolación lineal se expresa mediante la ecuación (2.10) donde y representa el valor faltante, y_0 y y_1 son los valores conocidos anteriores y posteriores, respectivamente, y t_0 y t_0 son los periodos correspondientes a esos valores. Esta técnica se utiliza para casos donde los datos faltantes son únicos y aislados [50] [51].

Por otro lado, la interpolación cuadrática se describe mediante la siguiente ecuación (2.11), y en este caso, t_0, t_1 y t_2 son periodos cercanos, y y_0, y_1 y y_2 son las observaciones correspondientes a esos periodos. A diferencia de la interpolación lineal, los valores t no necesitan ser contiguos a los datos faltantes; en su lugar, deben seleccionarse tiempos cercanos para realizar un ajuste adecuado y obtener una buena aproximación. Este método se utiliza principalmente cuando los datos faltantes son contiguos, especialmente si están ubicados en puntos críticos de estacionalidad [50].

2.2.5.2 Reducción de valores atípicos

También conocidos como outliers, son datos individuales o en conjunto que presentan un comportamiento atípico y no corresponden con las observaciones esperadas en la serie temporal. Estos valores suelen dificultar el proceso de generalización de los modelos de pronóstico, lo que complica el aprendizaje y afecta directamente el tiempo de ajuste y la calidad del pronóstico [52], [53].

En series de tiempo, eliminar estos valores puede tener un impacto directo en el procesamiento de la estacionalidad, ya que podría causar un desfase en la amplitud estacional, especialmente cuando se presentan múltiples valores atípicos en el mismo periodo estacional o cuando los periodos estacionales son cortos. Por lo tanto, es fundamental tratarlos adecuadamente.

Una técnica eficiente para abordar este problema es la aplicación de la regla de 3 sigma, donde cualquier valor que se encuentre por encima o por debajo de 3 desviaciones estándar con respecto a la media es ajustado a ese límite. Aunque esta técnica no elimina completamente el valor atípico, lo limita, permitiendo que otras estrategias de reducción de ruido puedan manejarlo con mayor eficacia [54], [55].

Por otro lado, es importante señalar que, en series de tiempo con estacionalidad multiplicativa, esta técnica por sí sola no es del todo eficaz. En estos casos, puede aplicarse de manera estacional, es decir, fragmentando la serie en N periodos estacionales y aplicando la técnica a cada uno de ellos. De esta forma, se compara el valor con sus respectivos fragmentos estacionales de periodos anteriores o futuros.

2.2.5.3 Reducción de ruido

Los conjuntos de datos, independientemente de su origen y naturaleza, pueden contener ruido o elementos indeseables que no reflejan fielmente las observaciones reales [55]. Este fenómeno es común y puede deberse a fallas en la captura, errores en la lectura de los datos o incluso fluctuaciones no controladas o aleatorias. Para mitigar estos problemas, se aplican diversas técnicas de reducción de ruido, cuyo objetivo es suavizar la serie de tiempo sin alterar significativamente el comportamiento de los datos. Estas técnicas permiten mantener la integridad de las observaciones clave al eliminar el ruido y mejorar la precisión de los análisis posteriores.

2.2.5.4 Descomposición en valores singulares

La descomposición en valores singulares (SVD) permite reducir los componentes principales de una serie de tiempo, descartando aquellos que resultan poco significativos y que contribuyen principalmente como ruido en la serie [56], [57], [58].

Dada una serie de tiempo $Y = y_1, y_2, y_3, \dots, y_N$, donde y_i representa cada observación y N es el tamaño total de la serie, es necesario aplicar un proceso conocido como Hankelización. En este proceso, la serie de tiempo se transforma en una matriz de dimensión $m \times n$, donde N

es la longitud total de la serie, n es el tamaño de la ventana elegida para construir la matriz y debe cumplir con $n < N$. El valor de m, el número de filas, se define como $m = N - n + 1$. La serie hankelizada tiene la representación que se muestra en la ecuación (2.12

$$A = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \\ y_2 & y_3 & \cdots & y_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_m & y_{m+1} & \cdots & y_N \end{bmatrix} \quad (2.12)$$

La matriz generada $A \in \mathbb{R}^{m \times n}$ describe las relaciones entre cada observación y sus observaciones vecinas. Posteriormente, esta matriz se descompone en el producto de tres matrices como se describe en la ecuación (2.13 [59].

$$A = U \cdot \Sigma \cdot V^T \quad (2.13)$$

Donde $U \in \mathbb{R}^{m \times m}$, $V^T \in \mathbb{R}^{n \times n}$ son matrices de vectores singulares izquierdos y derechos, respectivamente. $\Sigma \in \mathbb{R}^{m \times n}$ es una matriz diagonal que contiene los valores singulares σ, ordenados de mayor a menor, cumpliendo que $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Esta descomposición permite reducir la cantidad de valores singulares en la matriz Σ, haciendo cero aquellos que son poco significativos sin cambiar las dimensiones de la matriz original. Este proceso suaviza los datos y reduce los valores atípicos y el ruido presente en la serie.

$$Y' = y_{1,1} + y_{1,2} + \cdots + y_{1,n} + y_{2,n} + \cdots + y_{m.n} \quad (2.14)$$

El proceso de reconstrucción genera una matriz $A' \in \mathbb{R}^{m \times n}$ con reducción de ruido. A continuación, se realiza la des-hankelización, donde se toman los valores que conforman el borde superior y derecho de la matriz para reconstruir la serie de tiempo original. Esto se expresa en la ecuación (2.14.

2.2.5.5 Ventanas móviles

Las ventanas móviles o deslizantes son una estrategia comúnmente utilizada en el análisis de series de tiempo, ya sea con fines de suavizamiento, imputación de datos faltantes o análisis local del comportamiento de las variables.[26] Esta técnica consiste en seleccionar subconjuntos consecutivos de datos, denominados ventanas, que se desplazan a lo largo del conjunto total de observaciones. Cada ventana permite analizar un segmento específico de la serie, lo que facilita la detección de patrones locales, tendencias temporales o cambios en la estructura de los datos.

Además, el uso de ventanas móviles permite obtener una representación más estable del comportamiento general dentro de cada intervalo, reduciendo el impacto de valores atípicos (outliers) y facilitando el cálculo de estimaciones cuando existen datos ausentes.

En la ecuación anterior podemos encontrar el comportamiento general de esta técnica donde $X^{(t)}$ es el subconjunto o ventana móvil, x_t es un valor observado en el instante t, t es un índice temporal contenido en T, T es la longitud total de los datos y w el tamaño de la ventana móvil.

2.2.5.6 Técnicas de remuestreo

Los conjuntos de datos observados pueden estar muestreados en diferentes intervalos de tiempo, tales como series horarias, diarias, semanales, quincenales, mensuales, bimestrales, trimestrales, cuatrimestrales, anuales, entre otros, incluyendo intervalos personalizados.

Para homogeneizar estas series de tiempo, puede ser necesario aplicar una transformación mediante remuestreo**.** Las técnicas de remuestreo utilizadas en este trabajo consisten en generar series con intervalos de tiempo más amplios, lo que conlleva la necesidad de suavizar las series originales. En este caso, el suavizamiento se llevó a cabo mediante el cálculo de la media de los datos dentro de cada intervalo ampliado. Esta estrategia fue seleccionada debido a su bajo impacto en la reducción de información relevante, ya que disminuye el ruido presente en la serie y mejora la eficiencia del aprendizaje del modelo [60].

En la Figura 2.3 se presenta una representación gráfica del proceso de remuestreo, donde los datos diarios se agrupan en periodos más largos, como trimestrales, cuatrimestrales, semestrales y anuales. Sin embargo, este método no está limitado solo a estos intervalos y puede adaptarse a otras frecuencias.

Figura 2.3. Ejemplo grafico del proceso de remuestreo

2.2.5.7 Técnicas de estandarización

Las series de tiempo representan un conjunto de observaciones de sucesos, ya sean reales o sintéticos, cuya característica común es que están medidas con distintas métricas, y la amplitud de los datos puede estar acotada en intervalos variables, lo que dificulta su comprensión. Esta característica representa un problema particular para los algoritmos de pronóstico, ya que algunos de ellos no aceptan valores negativos o cercanos a cero debido a la naturaleza de su estructura. Además, las métricas de error que evalúan el desempeño de estas estrategias pueden verse afectadas cuando los valores son cero o se encuentran muy cerca de cero.

Para solucionar esta problemática, existen diversas técnicas que permiten acotar la amplitud de la serie, ubicándola dentro de un intervalo adecuado para el modelo utilizado.

Aunque existen múltiples técnicas para la transformación de datos, en este trabajo se emplean dos: la normalización y la estandarización. Es importante resaltar que, en ambos casos, los valores resultantes pueden incluir ceros, y en el caso de la estandarización, pueden generarse valores negativos. Para abordar esta problemática, se puede aplicar un desplazamiento sumando un valor constante a todos los datos transformados, lo que desplaza el rango de los valores y resuelve el inconveniente de trabajar con números negativos o ceros.

$$x' = \frac{x - X_{Min}}{X_{Max} - X_{Min}} \quad (2.15)$$

$$Z = \frac{x - \mu}{\sigma} \quad (2.16)$$

La ecuación (2.15 presenta el proceso de normalización, donde x representa cada valor individual del conjunto de datos, X_{Min} y X_{Max} son los valores mínimo y máximo, respectivamente en el conjunto de datos [61]. Este procedimiento ajusta los valores dentro de un rango específico, típicamente entre 0 y 1. Por otro lado, la ecuación (2.16 describe la estandarización, donde x corresponde a cada dato, μ es la media del conjunto, y σ es su desviación estándar [62]. Este método convierte los datos en una distribución con media cero y desviación estándar de uno.

2.3 Métodos de selección de variables relevantes

En el análisis de datos es común buscar capturar la mayor cantidad de información posible, bajo la idea de que un volumen grande de datos puede mejorar la capacidad de generalización del modelo, favorecer una mejor exploración del espacio de soluciones y, en consecuencia, mejorar el proceso de aprendizaje.[63] Sin embargo, este principio solo se cumple bajo ciertas condiciones, particularmente cuando las variables son independientes o débilmente correlacionadas entre sí y cuando su información resulta verdaderamente relevante respecto de la variable objetivo o de predicción.

En la práctica, muchas variables pueden presentar una alta correlación o aportar información redundante, lo que introduce ruido y aumenta la complejidad del modelo sin generar beneficios reales. Cada variable incorporada representa una nueva dimensión dentro del espacio de análisis, lo que incrementa el costo computacional, dificulta la optimización del modelo y puede deteriorar su capacidad de generalización. Este fenómeno, conocido como la “maldición de la dimensionalidad”, afecta directamente la precisión y estabilidad de los modelos predictivos, especialmente en escenarios donde el número de variables supera la cantidad efectiva de observaciones informativas.

La selección de variables relevantes tiene como propósito reducir la dimensionalidad del problema, identificando y conservando únicamente aquellas variables que contienen información significativa sobre el fenómeno en estudio. Este proceso permite eliminar atributos redundantes, irrelevantes o ruidosos, obteniendo un subconjunto de características más compacto y eficiente. Aunque este procedimiento puede implicar una ligera pérdida de varianza explicada, en la práctica suele compensarse con una mayor interpretabilidad, menor riesgo de sobreajuste y un mejor desempeño computacional.

Existen diversas estrategias para realizar la selección de variables relevantes. Los métodos estadísticos evalúan la relación entre cada variable y la variable objetivo mediante medidas de correlación, pruebas de hipótesis o información mutua. En otros casos, se aplican enfoques de agrupamiento o clustering para identificar grupos de variables con comportamientos similares, conservando únicamente los representantes más informativos de cada grupo. Por último, los métodos basados en modelos de aprendizaje automático integran la evaluación de la importancia de las variables dentro del propio proceso de entrenamiento, como ocurre con técnicas de regularización tipo LASSO, modelos de árboles de decisión o algoritmos evolutivos.

La selección de variables relevantes constituye una etapa fundamental en el desarrollo de modelos predictivos, ya que contribuye a mejorar la calidad de la información, reducir la complejidad del espacio de características y fortalecer la estabilidad de los resultados. En el caso de las series de tiempo multivariadas, esta tarea adquiere una importancia aún mayor, pues permite identificar las variables que realmente influyen en la dinámica temporal del sistema, optimizando así la capacidad de pronóstico del modelo.

2.3.1 Análisis de componentes principales

Es una estrategia del álgebra lineal, la cual puede entenderse como una extensión directa de la descomposición en valores singulares y es utilizada para reducir la dimensionalidad de un sistema. En términos generales, este enfoque tiene dos vertientes principales: la selección de características y la extracción de características, ambas orientadas a reducir la

dimensionalidad del problema, tratando de conservar la información más relevante contenida en los datos.

En sus fases iniciales, el método consiste en descomponer la matriz de variables o características $A = U\Sigma V^T$, donde A representa la matriz original, U es una matriz ortogonal cuyos vectores corresponden a los vectores singulares izquierdos, Σ representa una matriz diagonal con valores singulares ordenados de forma descendente y estrictamente positivos, y V^T contiene los vectores singulares derechos. Los valores singulares almacenados en Σ están directamente relacionados con la varianza explicada por cada componente, lo que permite interpretar su importancia en los datos. A partir de aquí, si se seleccionan únicamente aquellos vectores asociados a los mayores valores singulares, es decir, a la mayor varianza explicada, se está realizando un proceso de selección de características. En este caso, el conjunto original no se transforma, sino que conserva aquellas que aportan mayor información al modelo. Esto resulta útil en escenarios donde la interpretación de las variables es un aspecto clave.

Por otra parte, los vectores singulares también pueden interpretarse como nuevas direcciones que rotan los ejes del sistema original haciéndolo ortogonal. Así, se generan nuevas características que concentran la mayor parte de la varianza explicada, mientras que se descartan aquellas direcciones con aportaciones poco significativas. La selección de estos nuevos ejes define un nuevo sistema de variables que ya no guarda una relación directa con las variables originales. Aunque este proceso implica una pérdida en la interpretación individual de las variables, se obtiene a cambio un sistema más compacto y eficiente.

2.3.2 Técnicas de árboles de decisión

Los árboles de decisión pueden entenderse como modelos simples que resultan útiles para identificar y ponderar variables relevantes de un conjunto de datos.[64] A partir de reglas de partición, estos modelos seleccionan de forma automática aquellas variables que mejor segmentan la información, lo que permite estimar valores de objetivo y, al mismo tiempo, obtener una primera aproximación sobre la importancia relativa de cada predictor. No obstante, cuando se emplean de forma individual, los árboles suelen ser débiles y pueden

perder capacidad de generalización, sobre todo con variables altamente no lineales o ruidosas.

Cuando se usan grandes conjuntos de estos árboles implementado estrategias de ensamble, estas estrategias se vuelven robustas, ya que la combinación de múltiples modelos reduce la dependencia de decisiones locales y favorece la identificación de variables relevantes. En este sentido, los métodos de ensamble no solo mejoran el desempeño predictivo, sino que también aportan criterios más robustos para la selección de variables.

Bajo la estrategia de Bagging, el ensamble de árboles da lugar al modelo de Random Forest para regresión. Este método construye múltiples árboles de forma independiente, cada uno entrenado con subconjuntos aleatorios de los datos y de variables. A partir de esto, es posible estimar la importancia de cada variable en función de su contribución a la reducción del error. De esta manera, Random Forest se convierte en una herramienta eficaz para identificar predictores relevantes de forma estable, incluso en presencia de alta dimensionalidad.

Por otra parte, cuando los árboles se ensamblan mediante la estrategia de *Boosting*, se obtiene un modelo de XGBoost. En este caso, los árboles se entrenan de forma secuencial, enfocándose en corregir errores residuales de los modelos previos. Esta estrategia permite asignar mayor peso a las variables que más contribuyen al modelo de forma significativa a en cada iteración. Así, XGBoost no solo logra un alto desempeño predictivo, sino que también permite seleccionar variables relevantes en el modelo.

2.4 Métodos de pronóstico de series de tiempo

Como se ha mencionado en apartados anteriores, las series de tiempo son secuencias de observaciones ordenadas cronológicamente que registran el comportamiento de un fenómeno en particular.[44] Por ello, resulta de interés pronosticar su evolución futura con el objetivo de prever, corregir, modificar, atender o, simplemente, observar las circunstancias que rodean el fenómeno estudiado.

A medida que las series de tiempo comenzaron a ganar relevancia en el análisis de datos, surgieron diferentes estrategias estadísticas para abordar su complejidad. Inicialmente, métodos simples como la regresión lineal permitían realizar estimaciones sobre los datos. Sin embargo, estas técnicas se volvieron insuficientes ante la complejidad de los conjuntos de datos que presentaban comportamientos no lineales. En respuesta a estas limitaciones, se desarrollaron técnicas más avanzadas como los métodos de Holt o Jenkins, que fueron probados con éxito para realizar pronósticos en series de tiempo más complejas.

Con el tiempo, el volumen de información creció de manera exponencial y los fenómenos fueron observados con mayor precisión, lo que generó series de tiempo aún más complejas. Esto impulsó la aplicación de métodos de aprendizaje automático como una alternativa. Estos métodos se basan en ajustar funciones matemáticas a grandes conjuntos de datos, con el fin de generar pronósticos más precisos.

Ambos enfoques tienen sus ventajas y desventajas, además de que no existe una superioridad universal entre ellas, y su aplicabilidad depende de los datos y el fenómeno a pronosticar. En consecuencia, ambas estrategias siguen utilizándose con éxito. Además, se ha demostrado que la combinación de enfoques, mediante estrategias de hibridación o ensamble, puede mejorar el rendimiento predictivo. En general, los modelos que combinan técnicas clásicas con métodos de aprendizaje automático tienden a superar a los modelos que emplean únicamente el aprendizaje automático.

2.4.1 Métodos clásicos

Los métodos clásicos de pronóstico son un conjunto de estrategias estadísticas que funcionan analizando datos históricos y ajustando coeficientes que describen el comportamiento de la serie de tiempo. Estas estrategias representan la base conceptual para algunos modelos de machine learning, ya que buscan identificar patrones, tendencias y componentes estacionales sin necesidad de grandes volúmenes de datos para su correcta aplicación.

Una de las principales ventajas de los métodos clásicos es su eficiencia computacional, lo que permite su implementación con recursos limitados o cuando se requiere obtener

resultados en tiempos reducidos. Además, su interpretación suele ser más sencilla en comparación con modelos más complejos, lo que facilita el análisis y comprensión del comportamiento de la serie.

Diversos trabajos han demostrado que cuando se aplican de forma individual, estos métodos pueden ofrecer resultados competitivos; sin embargo, su mayor potencial se observa cuando se combinan con algoritmos de machine learning, dando lugar a estrategias híbridas o de ensamble muy robustas. De esta manera, su integración con técnicas modernas se presenta como una alternativa en problemas de predicción de series de tiempo.

2.4.1.1 Suavizamiento exponencial

El suavizamiento exponencial es una familia de métodos de pronóstico que se basa en la asignación de pesos o ponderaciones a las observaciones pasadas, de tal forma que los valores más recientes tienen mayor influencia en la estimación futura. Esta característica lo hace especialmente adecuado para series de tiempo donde la información reciente refleja mejor el comportamiento inmediato del proceso. [25]

El suavizamiento exponencial simple se emplea en series sin tendencia ni estacionalidad, y asume que la serie fluctúa alrededor de un nivel aproximadamente constante. El suavizamiento exponencial doble, también conocido como método de Holt, extiende este enfoque al incorporar explícitamente un componente de tendencia, permitiendo modelar crecimientos o decrecimientos sistemáticos en el tiempo. Finalmente, el suavizamiento exponencial triple, o método de Holt-Winters, añade un componente estacional, lo que lo convierte en una alternativa eficaz para series que presentan patrones periódicos bien definidos, tanto aditivos como multiplicativos.

2.4.1.2 Métodos autorregresivos

Los métodos autorregresivos constituyen un conjunto de técnicas clásicas y ampliamente estudiadas dentro del análisis y pronóstico de series de tiempo. Se fundamentan en la hipótesis de que el valor actual de una variable puede ser explicado, al menos de manera

parcial, a partir de sus propios valores pasados inmediatos. Bajo este supuesto, la serie de tiempo se modela como una combinación lineal de observaciones desfasadas, a las cuales se les asocia un conjunto de coeficientes que capturan la dependencia temporal existente en los datos.[65]

De forma general, un modelo autorregresivo de orden p se denota como $AR(p)$, la cual expresa el valor del instante t como función de los p valores anteriores inmediatos, añadiendo un termino de error natural de las series de tiempo. Estos modelos asumen, que la serie de tiempo presenta propiedades de estacionariedad a lo largo del tiempo. Bajo estas condiciones, los métodos autorregresivos ofrecen interpretaciones claras de la estructura temporal y permiten analizar de forma directa la influencia de observaciones pasadas sobre el comportamiento futuro de la serie.

A pesar de su simplicidad, los modelos autorregresivos han demostrado un desempeño competitivo en series de tiempo con patrones relativamente estables y dependencias lineales bien definidas. Sin embargo, su capacidad de modelado se ve limitada frente a dinámicas altamente no lineales o ante la presencia de estacionalidades complejas, lo que ha motivado su extensión hacia modelos más generales y su integración con enfoques de machine learning.

2.4.2 Métodos de machine learning

Dentro del campo de la inteligencia artificial, el machine learning es una rama enfocada en el desarrollo de estrategias capaces de adaptarse y responder de forma eficiente ante situaciones tanto conocidas como desconocidas sobre un problema específico.[66] Estas estrategias son utilizadas principalmente para realizar estimaciones, generar pronósticos y apoyar en la toma de decisiones en diversos contextos.

El funcionamiento de estas técnicas se basa en un proceso de aprendizaje, mediante el cual se ajustan los parámetros de los algoritmos o modelos a partir de la información disponible. Este ajuste suele realizarse con datos previamente observados o etiquetados, lo que permite que el modelo reconozca patrones y mejore su capacidad de predicción o clasificación con el

tiempo. A medida que el modelo se expone a más ejemplos, su desempeño mejora progresivamente, siempre y cuando los datos sean adecuados.

En este sentido, la calidad y cantidad de los datos utilizados juegan un papel crucial. Un conjunto de datos bien estructurado, representativo y libre de errores incrementa significativamente la precisión y generalización del modelo. Por el contrario, si los datos son ambiguos, contradictorios o irrelevantes, pueden surgir problemas como el sobreentrenamiento que se da cuando el modelo memoriza los datos sin aprender los patrones reales o el subentrenamiento, donde el modelo no alcanza a capturar relaciones significativas. En ambos casos, el desempeño del modelo frente a nuevos escenarios puede verse seriamente comprometido.

En el presente trabajo se describen brevemente las estrategias de pronóstico empleadas, presentándolas en su versión base para mantener claridad y enfoque. No obstante, cabe señalar que muchas de estas estrategias cuentan con múltiples variantes y extensiones, las cuales pueden adaptarse según la naturaleza del problema y las características de los datos disponibles.

2.4.2.1 Bosques aleatorios

También conocida como Random Forest Regression (RFR) es una estrategia de machine learning cuyo comportamiento se basa en la idea de la "*sabiduría de las masas*", generando múltiples modelos individuales de árboles de decisión para mejorar el rendimiento general frente a estrategias que solo incorporan uno [67]. De forma más específica, este enfoque consiste en construir un conjunto de árboles de decisión independientes, cada uno entrenado en un subconjunto aleatorio de datos y características, lo que le permite generar predicciones robustas a partir de una diversidad de opiniones o modelos simplificados [64]. La aleatoriedad en la selección de características y en la profundidad de cada árbol genera una mayor variedad en la topología y los patrones de cada árbol, lo que, en conjunto, contribuye a minimizar el riesgo de sobreajuste en los datos de entrenamiento [68].

Aunque los árboles de decisión de forma individuales pueden ser débiles al manejar series de tiempo debido a su falta de capacidad para captar tendencias y patrones estacionales complejos, su simplicidad es beneficiosa en este contexto. En este enfoque, cada árbol se especializa en una porción particular del conjunto de datos y se construye de manera que refleja una variabilidad significativa respecto a otros árboles, tanto en la selección de variables como en su estructura [69].

El resultado final de la regresión se obtiene a partir de la combinación de los resultados de todos los árboles individuales. En la mayoría de los casos, se calcula como una media aritmética de las predicciones de cada árbol, aunque también existen otras técnicas de agregación como la media ponderada o la mediana, dependiendo de las características del problema y de la necesidad de controlar el impacto de árboles individuales atípicos.

La Figura 2.4 muestra el funcionamiento general de la estrategia, destacando cómo el ensamble de múltiples modelos permite lograr una predicción robusta mediante la combinación de diversos árboles de decisión. Esta representación gráfica ayuda a visualizar la naturaleza colectiva del modelo, donde cada árbol contribuye a la estimación final, resultando en una predicción global más confiable, precisa y con un riesgo de sobre ajuste bajo.

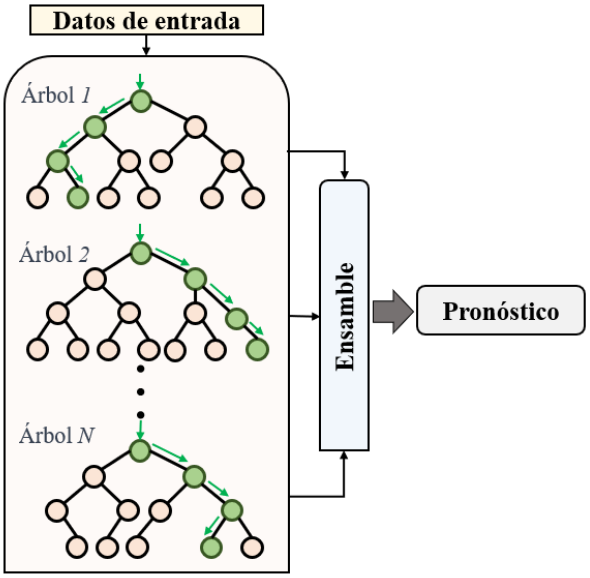


Figura 2.4. Descripción grafica del comportamiento de Random Forest

2.4.2.2 Redes neuronales

Es un enfoque de machine learning que guarda cierta similitud con el proceso neuronal, por lo que puede considerarse una estrategia bioinspirada. [70] Las redes neuronales están formadas por un conjunto de neuronas organizadas en capas, las cuales procesan los datos de entrada y propagan los resultados a través de la estructura de la red hacia enfrente. Este tipo de modelos ha demostrado ser especialmente efectivo en tareas de pronóstico de series de tiempo.

En su funcionamiento básico, una neurona se entiende como una función matemática que recibe valores de entrada, ya sea de los datos de entrada o de otras capas neuronales. Estas entradas se combinan y procesan para generar un valor de salida, el cual es propagado a todas las neuronas con las que mantiene conexión. Generalmente, estas conexiones son hacia delante y suelen ser densas, lo cual es una de las principales razones de su buen desempeño. De manera análoga a los sistemas nerviosos, cada conexión cuenta con un peso o ponderación específica que determina la importancia del valor recibido. Una vez que la información ha atravesado todas las capas, se obtiene el resultado final del modelo.

El proceso de aprendizaje, o ajuste de la red, se realiza en sentido inverso, es decir, desde la capa de salida hacia las capas de entrada. Este proceso consiste en la evaluación de una función de aprendizaje, que en modelos tradicionales suele ser la regla delta. A medida que la información se propaga hacia atrás, el ajuste se va distribuyendo entre las conexiones, de tal forma que cada una se actualiza de acuerdo con su grado de participación en el resultado final.

Con el paso del tiempo, este modelo ha evolucionado, dando lugar a arquitecturas de redes neuronales con propósitos específicos, diseñadas para atender limitaciones puntuales de las redes tradicionales o para adaptarse mejor a ciertos tipos de problemas.

2.4.2.3 Redes neuronales LSTM

Las redes neuronales Long-Short Term Memory (LSTM) son un enfoque de las redes neuronales recurrentes que ha ganado popularidad debido a su capacidad para abordar problemas clásicos de aprendizaje con dependencias sucesivas, que suelen ser complicados en las redes neuronales recurrentes tradicionales [71]. Las LSTM superan a estos enfoques, siendo ideales para tareas como el pronóstico de señales, el manejo de flujos de datos o la generación de texto [72].

En general, las redes neuronales tradicionales no están diseñadas para manejar dependencias temporales en series de tiempo, ya que su proceso de ajuste está limitado a la instancia actual de los datos, sin aprovechar las relaciones con instancias pasadas o futuras [73]. Por esta razón, las redes neuronales recurrentes representan una solución más adecuada, ya que su arquitectura permite la retroalimentación de información, almacenando un recuerdo o memoria de los valores anteriores, lo que proporciona cierto grado de influencia del pasado en el valor actual.

No obstante, un problema inherente a las redes recurrentes tradicionales es que la influencia de las observaciones pasadas tiende a desvanecerse rápidamente a medida que el horizonte temporal aumenta [74]. Esto significa que las redes recurrentes son eficientes en la captura de dependencias a corto plazo, pero pierden efectividad cuando se requiere recordar información relevante a mediano o largo plazo. Esta limitación es especialmente problemática en series de tiempo con estructuras estacionales, donde los patrones a largo plazo son cruciales para un pronóstico preciso [75].

Para abordar esta deficiencia, las LSTM ofrecen una solución robusta. Estas redes incluyen celdas de memoria en su arquitectura que permiten retener y procesar información relevante a corto, mediano y largo plazo sin experimentar una degradación significativa en el desempeño. Esto se debe a que las celdas LSTM están diseñadas para mantener información relevante durante intervalos de tiempo más largos y gestionar de manera eficiente qué información debe almacenarse, actualizarse o descartarse.

En términos estructurales, las LSTM se caracterizan por contener un tipo especial de celda dentro de una red neuronal recurrente estándar. Cada celda LSTM incluye un sistema complejo de gestión de la memoria compuesto por cuatro componentes principales: tres puertas sigmoideas (puerta de entrada, puerta de olvido y puerta de salida) y una función de activación de tangente hiperbólica. Las puertas sigmoideas controlan la cantidad de información que debe entrar, mantenerse o salir de la celda, permitiendo que la red decida qué información es relevante para las predicciones futuras. La Figura 2.5 muestra un esquema de la conectividad interna de una celda LSTM.

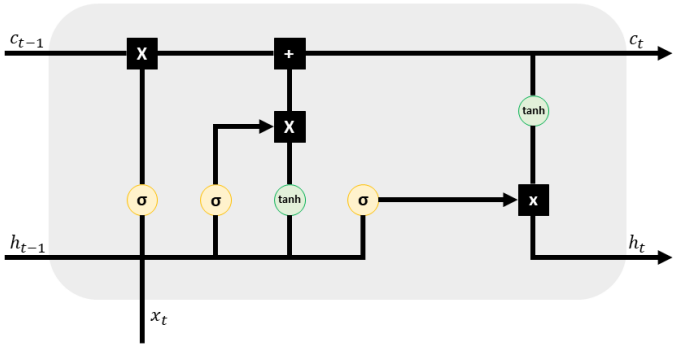


Figura 2.5. Estructura interna de una celda LSTM

Además, la arquitectura de las redes LSTM permite que estas celdas trabajen de manera individual, en paralelo, o apiladas en capas. Este apilamiento, mostrado en la Figura 2.6, incrementa la profundidad de la red y su capacidad de generalización. Al apilar múltiples capas de celdas LSTM, cada capa se especializa en tareas más pequeñas y específicas, lo que mejora la calidad de la predicción final al procesar la información de manera más efectiva a lo largo de la red.

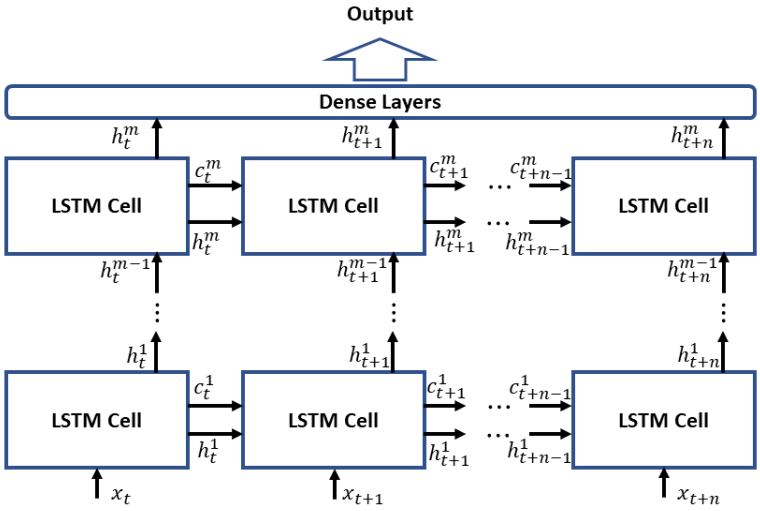


Figura 2.6. Red LSTM organizada en capas apiladas

2.4.2.4 Redes neuronales CNN

En los últimos años, las redes neuronales convolucionales (CNN) han demostrado ser eficaces en la clasificación de imágenes, aprovechando su capacidad para identificar y extraer características mientras disminuyen la dimensionalidad de los datos [74], [76]. Este enfoque es útil en imágenes o elementos matriciales, pero también se ha explorado su aplicabilidad en series de tiempo, debido a su capacidad para capturar y representar patrones secuenciales y relaciones internas entre datos [77].

La naturaleza de las series de tiempo es mediante representaciones vectoriales, y debido a que las redes CNN tradicionales procesan matrices de datos, lo que ha llevado a estrategias como la hankelización para adaptar series de tiempo a este formato. La hankelización convierte una serie temporal en una matriz Hankel, que representa un subconjunto de datos mediante una ventana de tamaño l tal que $l \leq t$, donde t es la longitud de la serie. Esta ventana se desplaza secuencialmente, creando una matriz simétrica en la cual cada valor tiene una relación directa con sus predecesores en el tiempo. La Figura 2.7. Descripción grafica de la matriz de hankelización representa este hankelización, donde la matriz resultante facilita el aprendizaje de características temporales en la CNN al mantener la estructura secuencial de los datos.

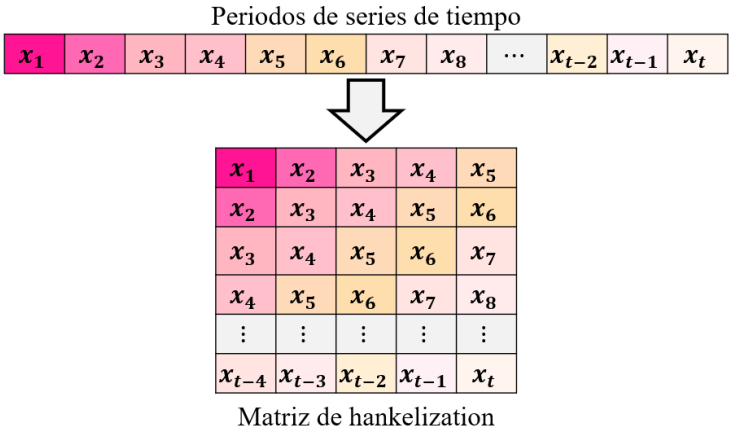


Figura 2.7. Descripción grafica de la matriz de hankelización

Para ilustrar la arquitectura típica de una CNN aplicada a series de tiempo, la Figura 2.8 muestra un modelo compuesto por dos capas convolucionales, cada una seguida de una capa de pooling. Esta estructura permite identificar patrones y reducir la dimensionalidad antes de alimentar la información a una red MLP (Multilayer Perceptron) que actúa como la capa de salida con una sola neurona.

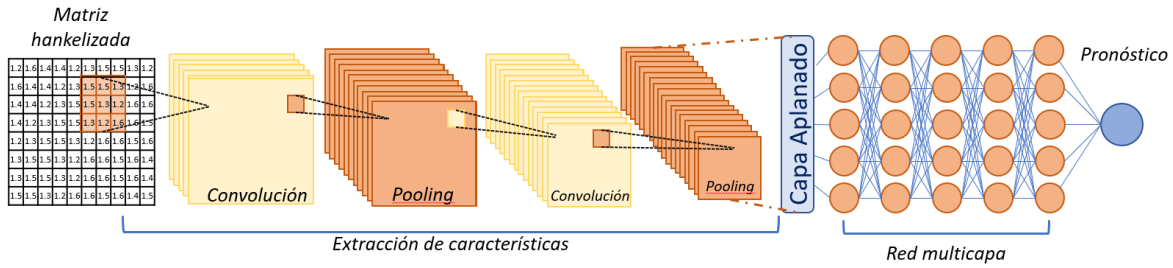


Figura 2.8. Descripción grafica del proceso convolucional de una red CNN

Las capas convolucionales aplican filtros, cuyo tamaño y número son parámetros ajustables. En cada capa, se realiza un barrido de la matriz, conocido como *stride*, que también es sintonizable. En general, el número de filtros se incrementa con la profundidad de la red, duplicándose en cada capa. Las capas de pooling agrupan información y reducen dimensionalidad mediante ventanas rectangulares o cuadradas, manteniendo la integridad de las características esenciales de los datos.

El proceso concluye con una capa de aplanado, que transforma la salida matricial en un vector, permitiendo que las capas finales del MLP procesen la información secuencialmente para emitir una predicción final. Esta arquitectura resulta particularmente adecuada para capturar patrones estacionales y tendencias inherentes a las series temporales, proporcionando así una solución robusta y adaptable para problemas de pronóstico y análisis secuencial.

2.4.2.5 Máquinas de soporte de regresión

Las máquinas de soporte vectorial constituyen un conjunto de modelos de machine learning originalmente formulados para tareas de clasificación. [78] Consisten en la construcción de una frontera de decisión óptima, la cual se obtiene mediante seleccionado un kernel capaz de transformar el espacio original de características en uno de mayor dimensionalidad, estirándolo. Esta transformación permite que patrones que no son linealmente separables puedan diferenciarse mediante hiperplanos.
Las máquinas de soporte vectorial para regresión (SVR) representan el enfoque anterior orientado hacia problemas de pronóstico. En la regresión se busca aproximar una función desconocida que describe la relación entre las variables de entrada y una variable continua de salida. El enfoque de SVR se centra en identificar un kernel y un conjunto de coeficientes que permitan aproximar, de la forma más fiel posible, la función generadora de la serie, manteniendo el error dentro de una banda de tolerancia previamente definida.

Una característica relevante es el concepto de intervalo de confianza, el cual permite controlar la complejidad del modelo. De esta manera, el modelo no solo ajusta los valores observados, sino que también reduce la sensibilidad al ruido.

Las SVR han sido ampliamente utilizadas en tareas de pronóstico, destacando su desempeño en series de tiempo con comportamientos estacionales y tendencias bien definidas. Su capacidad para modelar relaciones no lineales, junto con un control explícito del compromiso entre ajuste y generalización, las convierte en una alternativa sólida frente a métodos

tradicionales y en un componente recurrente dentro de enfoques híbridos y comparativos en estudios de predicción temporal.

2.5 Métodos de ranqueo

Cuando se trabaja con varios modelos de pronóstico, es común que cada uno ofrezca un desempeño distinto. En tales casos, puede ser necesario realizar una selección de métodos o determinar una ponderación inicial para un modelo de ensamble.[79]

Utilizar la métrica de error como único mecanismo de ponderación podría ser un procedimiento con resultados no óptimos, ya que los modelos de pronóstico pueden haber sufrido sobreajuste al conjunto de validación utilizado. Esto puede llevar a que los modelos ofrezcan pronósticos inexactos cuando se aplican a otros conjuntos de datos.
Los métodos de ranqueo evalúan otros aspectos adicionales a las métricas de desempeño, permitiendo generar un listado que clasifica los modelos de pronóstico según su idoneidad. A continuación, se describen los métodos de ranqueo empleados en este trabajo.

2.5.1 Criterio de información de Akaike

El criterio de información de Akaike (AIC) es una herramienta para evaluar métodos de pronóstico. [79] El objetivo es encontrar un equilibrio entre la precisión del modelo, medida por una función de verosimilitud, y la complejidad del modelo, la cual está determinada por el número de hiperparámetros ajustados para realizar el pronóstico.

$$AIC = 2k - 2\ln(L) \qquad (2.17)$$

El valor del AIC disminuye a medida que los modelos obtienen mejores resultados o la cantidad de hiperparámetros a configurar disminuye. Este comportamiento se refleja en la ecuación (2.17, donde k representa la cantidad de hiperparámetros del modelo, y L denota la función de verosimilitud. En este contexto, la función de verosimilitud se construye a partir de una métrica de error evaluado bajo el supuesto de normalidad de los errores.

2.5.2 Criterio de información Bayesiano

El criterio de información bayesiano (BIC) es una herramienta estadística para evaluar y comparar diversos métodos de pronóstico.[79] Este enfoque se basa en la búsqueda de un equilibrio entre la precisión del pronóstico del modelo, evaluada a través de la función de verosimilitud, y la complejidad de ajuste del modelo. Un aspecto distintivo del BIC es su capacidad para penalizar aquellos modelos que se benefician de muestras pequeñas, contribuyendo así a una evaluación más robusta y generalizable.

$$BIC = n \cdot \ln(L) + k \cdot \ln(n) \qquad (2.18)$$

En la ecuación (2.18 se presenta el criterio donde, n representa el tamaño del conjunto de datos, L la función de verosimilitud y k el número de hiperparámetros en el modelo. A diferencia del AIC, el BIC multiplica el número de parámetros por el logaritmo del tamaño de la muestra, un peso que acentúa la penalización a medida que aumenta la complejidad del modelo.

2.6 Métodos de ensamble

El enfoque de ensamble o combinación de métodos es una estrategia ampliamente utilizada y exitosa que consiste en combinar los resultados de dos o más modelos entrenados de manera individual para generar un resultado único. Este resultado ensamblado debería ser igual o superior al mejor resultado obtenido por los modelos individuales participantes [80]. La construcción de este ensamble suele consistir en una combinación ponderada de los métodos involucrados [40]. Cabe destacar que la suma de las ponderaciones no siempre debe ser igual a la unidad, especialmente en pronósticos que combinan componentes de series de tiempo, ya que algunos pueden ser aditivos, multiplicativos o estar en diferentes escalas con respecto a los demás componentes [81].

$$\hat{y}_e = p_1\hat{y}_{m_1} + p_2\hat{y}_{m_2} + p_3\hat{y}_{m_3} + \cdots + p_n\hat{y}_{m_n} \qquad (2.19)$$

En la ecuación (2.19, $\hat{y}_e$ representa el pronóstico ensamblado, p_i es la ponderación asignada a cada modelo m_i que participa en el ensamble, y $\hat{y}_{m_i}$ es el pronóstico individual de cada uno de los modelos. Generalmente, las ponderaciones p_i son uniformes, lo que implica que cada modelo contribuye de manera equitativa. Sin embargo, esta aproximación puede dar lugar a resultados de menor calidad que los obtenidos por el mejor modelo individual en algunos casos. Esto ocurre porque los distintos modelos tienden a generalizar aspectos diferentes de los datos, y la calidad del pronóstico varía en función de factores como su configuración, los datos utilizados y las particularidades de cada método.

Debido a esta variabilidad, existen diversas estrategias para seleccionar las ponderaciones en los ensambles, como el uso de métricas de desempeño, métodos de ranqueo o técnicas de optimización combinatoria para asignar el peso más apropiado a cada modelo.

Además, existen diferentes enfoques de ensamble, que determinan la forma en que se realiza la combinación de modelos. En este trabajo se abordan dos técnicas clave que han mostrado gran efectividad: Boosting y Bagging. Estas técnicas serán descritas a continuación y representan enfoques avanzados dentro de la combinación de modelos, con características particulares que les permiten superar a otros métodos de ensamble en términos de precisión y robustez.

2.6.1 Estrategias de boosting

El boosting es una estrategia secuencial en la que múltiples modelos se entrenan de manera progresiva, de modo que cada nuevo modelo intenta corregir los errores cometidos por los modelos anteriores. [82] Este enfoque se ha vuelto ampliamente popular debido a su robustez y a su eficiencia computacional, a pesar de que su estructura de entrenamiento es secuencial.

El boosting consiste en asignar un mayor peso a los valores que fueron mal predichos por los modelos previos, concentrando el aprendizaje en los casos más difíciles que en lo general se traducen como errores. En la primera etapa, el modelo inicial se entrena utilizando el conjunto completo de datos. Posteriormente, los siguientes modelos enfocan su entrenamiento en los errores residuales del anterior, ajustando sus parámetros para mejorar

la precisión global del ensamble. Este proceso continúa de manera iterativa, produciendo un sistema en el que cada modelo complementa las deficiencias del previo.

Aunque la naturaleza secuencial del boosting impide el paralelismo total, cada modelo posterior requiere entrenar sobre un subconjunto de información más pequeño, lo que reduce el costo computacional por iteración o modelo nuevo. Esta característica permite que el método mantenga un equilibrio entre precisión y eficiencia, resultando especialmente útil en contextos donde se busca minimizar el error sin sacrificar interpretabilidad o estabilidad.

2.6.2 Estrategias de bagging

Las estrategias de ensamble que incorporan técnicas de bagging (bootstrap aggregating) se han consolidado como estrategias ampliamente utilizadas en el aprendizaje automático debido a que promueven una forma natural de cooperación e integración entre modelos individuales. [82] Su principio fundamental consiste en integrar múltiples modelos base para obtener una predicción conjunta más estable y precisa que la generada por cualquiera de ellos de manera individual.

El enfoque del bagging parte de la idea de combinar modelos simples, que en algunos casos en lo individual se consideran estrategias débiles, aunque en la práctica pueden ser también modelos robustos entrenados sobre diferentes subconjuntos de datos. La lógica detrás de este proceso es que, al combinar sus resultados, los errores individuales tendrán una tendencia a compensarse entre sí, mejorando así la capacidad de generalización del sistema. En esta integración, cada modelo contribuye con un determinado grado de participación o ponderación en el ensamble final, el cual puede definirse mediante distintas estrategias de ajuste. Estas ponderaciones pueden ser uniformes, cuando todos los modelos aportan el mismo peso; basadas en desempeño, cuando se priorizan aquellos con menor error o mayor precisión; o ajustadas inteligentemente, cuando se emplean técnicas heurísticas que optimizan la contribución de cada modelo según su comportamiento.

A diferencia del Boosting, donde los modelos se entrenan de manera secuencial y cada uno intenta corregir los errores del anterior, el bagging se caracteriza por permitir un

entrenamiento completamente paralelo, ya que los modelos individuales son independientes entre sí. Esta independencia favorece la eficiencia computacional individual y la escalabilidad del proceso, aunque también puede generar un efecto de redundancia, al no existir interacción o intercambio de información entre los modelos, estos pueden aprender patrones similares de los datos, lo que incrementa el riesgo de sobreajuste, especialmente cuando los conjuntos de entrenamiento no presentan suficiente variabilidad o son los mismos.

Entre las principales ventajas de las técnicas de bagging se encuentran su alta capacidad de paralelización, la reducción del sesgo y la mayor estabilidad de los resultados, derivada de la agregación de múltiples modelos. No obstante, sus desventajas incluyen la posibilidad de sobreajuste en conjuntos con alta correlación entre los modelos base, así como un costo computacional elevado cuando se entrena con modelos robustos, producto del entrenamiento simultáneo de varios modelos.

En conjunto, los métodos basados en bagging constituyen una herramienta poderosa para la mejora del desempeño, especialmente en contextos donde la variabilidad de los datos puede afectar la estabilidad de los modelos individuales. Su adecuada configuración y balance entre independencia, diversidad y ponderación son factores clave para obtener ensambles de calidad.

2.6.3 Optimización del ensamble

El proceso de ensamble consiste en combinar distintos modelos con el objetivo de aprovechar las fortalezas de cada uno de ellos. [40] Esta combinación puede llevarse a cabo mediante estrategias como bagging, boosting u otras técnicas similares. En general, todos los modelos participantes contribuyen al resultado final, y con frecuencia esta participación es uniforme, es decir, cada modelo tiene el mismo peso en la combinación.

No obstante, repartir el peso de forma uniforme no siempre conduce a una configuración óptima. Aunque el resultado final incluye el aporte de todos los modelos, esto no garantiza que el desempeño sea mejor al mejor de los modelos por separado. Por esta razón, algunas estrategias como los criterios BIC y AIC pueden utilizarse para ordenar los modelos y

asignarle una participación proporcional a su calidad. Sin embargo, incluso estas estrategias no aseguran obtener la mejor combinación posible.

Por lo tanto, el problema de definir cuánto debe aportar cada modelo dentro del ensamble puede entenderse como un problema de optimización combinatoria. El objetivo es encontrar la combinación de participaciones que minimice el error del pronóstico, permitiendo así alcanzar el mejor desempeño posible. Esta optimización puede realizarse mediante búsquedas exhaustivas o mediante técnicas heurísticas, y en general permite obtener resultados comparables o incluso superiores a los del mejor modelo individual.

2.7 Métodos de optimización

En el ámbito computacional, muchos problemas de optimización presentan una gran cantidad de soluciones factibles, donde el objetivo es identificar aquella que optimice una función específica, ya sea minimizando o maximizando su valor dentro de este conjunto. Este conjunto de soluciones constituye el espacio de soluciones, y su complejidad aumenta exponencialmente con el número de variables y restricciones del problema, lo que hace que la exploración exhaustiva sea infactible. La búsqueda exhaustiva de la solución óptima es un problema de optimización combinatoria que, dada la cantidad de posibles combinaciones, suele requerir altos tiempos de procesamiento.

Para tratar de solucionar este problema es común recurrir a algoritmos heurísticos y metaheurísticos que optimizan el proceso de búsqueda. Estas técnicas, en lugar de explorar cada solución posible, realizan una búsqueda inteligente para identificar soluciones cercanas al óptimo en un tiempo razonable. [40] Aunque estos métodos no siempre garantizan el hallazgo del máximo o mínimo global, proporcionan soluciones de buena calidad dentro de un tiempo polinomial.

Entre las múltiples estrategias aplicables, la elección del método depende del contexto del problema. En este trabajo, se ha seleccionado un enfoque de algoritmo evolutivo poblacional. Este tipo de algoritmo permite no solo integrar soluciones preseleccionadas que podrían encaminar la búsqueda, sino también añadir soluciones aleatorias. Este balance facilita una

amplia exploración del espacio de soluciones y promueve la diversidad en la población, evitando la convergencia prematura hacia óptimos locales y mejorando las probabilidades de obtener soluciones de alta calidad.

2.7.1 Algoritmo genético

Los algoritmos genéticos forman parte de las estrategias de optimización evolutiva y están inspirados en los procesos de evolución de las especies. [83] Esta estrategia ha demostrado su éxito en diversas áreas de aplicación, como la selección de hiperparámetros.

Estos algoritmos están constituidos por una población de soluciones codificadas en genotipos que, a través de generaciones, tienen la oportunidad de cruzarse y mutar para generar nuevas soluciones. Estas nuevas soluciones pueden sustituir algunos elementos de la población existente o ser descartadas. La población inicial suele generarse de forma aleatoria, asegurando la factibilidad de las soluciones; sin embargo, también puede ser creada mediante alguna otra estrategia con el fin de incorporar soluciones conocidas al inicio y así encaminar el algoritmo hacia una convergencia más rápida. El tamaño de la población debe definirse en función de la cantidad total de soluciones posibles y su variabilidad. Poblaciones demasiado pequeñas pueden provocar una convergencia prematura, mientras que poblaciones muy grandes pueden ralentizar el proceso de optimización.

La Figura 2.9 ilustra el proceso evolutivo de cada generación. En cada iteración, se seleccionan pares de genotipos que actuarán como padres. Estos genotipos se combinan mediante una función de cruza, y los nuevos genotipos resultantes tienen una probabilidad de mutación, lo que introduce diversidad en la población. En la fase final, se evalúan los individuos generados para determinar si sustituyen a algún miembro de la población o si deben ser descartados.

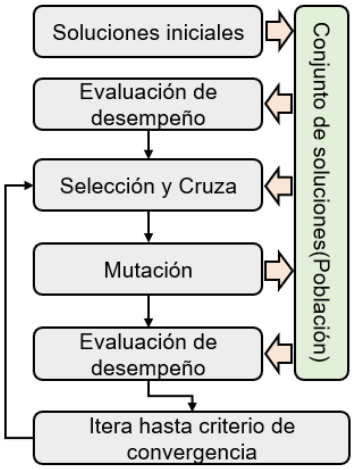


Figura 2.9. Ejemplo del proceso iterativo de un algoritmo evolutivo

El proceso de selección de padres en este trabajo se lleva a cabo mediante una estrategia de ruleta ponderada, en la cual se evalúa el fitness o desempeño de cada genotipo. Aquellos con mejor desempeño tienen un área mayor en la ruleta, lo que les otorga una mayor probabilidad de ser seleccionados. Sin embargo, dado que la selección es aleatoria, los individuos con menor desempeño aún tienen una probabilidad, aunque reducida, de ser elegidos. Es importante destacar que ningún individuo debe ser excluido de la selección, ya que el elitismo en este proceso puede generar problemas como la convergencia prematura o el estancamiento en la población.

El proceso de cruza depende del contexto del problema a resolver, ya que existen diversos métodos, como la combinación de bloques de genes de cada padre o el promedio de los valores de los genes. En este trabajo, se utiliza la cruza dispersa, en la cual cada gen es seleccionado de forma aleatoria de los padres. Este método garantiza la factibilidad de las soluciones generadas.

En cuanto a la mutación, su propósito es aportar diversidad a la población, evitando que todos los individuos converjan a características similares, lo que podría llevar a un estancamiento prematuro. Sin embargo, la mutación debe aplicarse con precaución, evitando una probabilidad excesiva que convierta el proceso en una caminata aleatoria por el espacio de

soluciones. En este trabajo, se emplea una estrategia de mutación doblemente aleatoria, en la que el individuo seleccionado para mutar altera una cantidad aleatoria de sus genes, asignándoles valores también aleatorios, asegurando siempre que la solución resultante sea factible. La probabilidad de mutación es dinámica y aumenta conforme avanza el proceso evolutivo.

Si bien la estrategia planteada tiene un número límite de generaciones, en algunas ocasiones puede no ser necesario completar todas las iteraciones, ya que la probabilidad de mejora disminuye con el número de generaciones y la mejor solución encontrada deja de cambiar. Por ello, se ha definido un criterio de convergencia adicional: si la mejor solución no presenta cambios después de un número determinado de generaciones, se considera que el algoritmo ha convergido y, en consecuencia, el proceso evolutivo se detiene.

2.7.2 Optimización por enjambre de partículas

El algoritmo de optimización por enjambre de partículas (PSO) es una técnica evolutiva inspirada en el comportamiento colectivo de sistemas de la naturaleza, tales como bandadas de aves o grupos de peces.[84] Su comportamiento básico consiste en la búsqueda cíclica de soluciones óptimas mediante un conjunto de partículas que exploran el espacio de soluciones, ajustando su posición a partir de su experiencia individual y de la información compartida por el enjambre. Esta característica permite encontrar soluciones de calidad con un costo computacional moderado, regularmente inferior al de otras técnicas evolutivas.

En este trabajo, PSO se emplea como un mecanismo de optimización de pesos dentro del proceso de ensamble, permitiendo combinar de forma los modelos individuales de pronóstico de forma eficiente. Al evaluar de forma iterativa el desempeño de cada combinación, el algoritmo converge hacia una solución que equilibra las contribuciones de los distintos modelos, aprovechando sus fortalezas y reduciendo el impacto de sus debilidades en el desempeño global del sistema, asegurando resultados al menos tan buenos como el mejor método individual.

Capítulo 3

3 Metodología

A lo largo de este trabajo de investigación, se han explorado una amplia gama de modelos, metodologías y estrategias, las cuales se presentan y detallan en esta sección. Cada enfoque es descrito en términos de su arquitectura, configuración y los parámetros clave utilizados para su implementación, proporcionando un análisis exhaustivo de las decisiones técnicas que respaldan los experimentos realizados.

Además, se describen en detalle los conjuntos de datos empleados en este estudio, incluyendo las características estadísticas relevantes que definen su comportamiento. Esta información es esencial para comprender el contexto en el que se aplican los modelos, así como los desafíos que presentan los datos en términos de su complejidad y variabilidad.

Asimismo, se presentan las métricas de error utilizadas para evaluar la precisión de los modelos propuestos y otras métricas que permiten cuantificar el rendimiento predictivo. Estas métricas son fundamentales para comparar la efectividad de los diferentes enfoques y seleccionar el más adecuado para las tareas específicas.

3.1 Condiciones de experimentación

En el marco del proceso experimental, se alojaron todos los datos utilizados en este trabajo en archivos con formato CSV, disponibles en la sección de anexos. Este tipo de archivo, comúnmente utilizado en el ámbito de machine learning y pronóstico, es ideal para datos numéricos y compatible con todas las plataformas computacionales actuales.

Por otro lado, las metodologías evaluadas fueron desarrolladas en el lenguaje Python sobre la plataforma de Anaconda. La programación en Python proporciona una amplia gama de librerías, procedimientos y métodos optimizados para el área de machine learning, además de ser ampliamente adoptada por la comunidad científica. Esto facilita la reproducibilidad de las estrategias propuestas y asegura compatibilidad multiplataforma sin necesidad de ajustes sustanciales en la estructura del código. Para esta investigación, se utilizaron librerías como NumPy, Keras y Scikit-learn, optimizadas para reducir el tiempo de codificación e, incluso, en algunas ocasiones, el tiempo de experimentación.

En cuanto al equipo de cómputo, se empleó una PC con procesador Ryzen 7 5700X de 8 núcleos y frecuencia de 3.70 GHz, 32 GB de memoria RAM y una tarjeta gráfica discreta Nvidia 4060 OC. Cuando fue posible y siempre que la naturaleza de las metodologías lo permitió, se empleó el cómputo acelerado en GPU para mejorar la eficiencia del proceso experimental.

3.2 Métricas de desempeño

En el ámbito del pronóstico mediante técnicas clásicas o de aprendizaje automático, la elección de métricas de desempeño para ajustar los modelos resulta fundamental, ya que cada de las métricas captura un aspecto específico de los errores de predicción y permite interpretar de distintas formas en el desempeño del modelo. La literatura menciona que no existe una métrica única que pueda ofrecer una evaluación completa y universalmente aplicable para modelos de pronóstico en problemas de regresión. Por ello, en este trabajo se usa un conjunto de métricas, incluyendo el error cuadrático medio (MSE), la raíz del error

cuadrático (RMSE), el error porcentual absoluto medio (MAPE) y el error porcentual absoluto medio simétrico (sMAPE), seleccionadas para proporcionar una visión integral de los modelos propuestos y su precisión.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (3.1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (3.2)$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3.3)$$

$$sMAPE = \frac{100}{N}\cdot\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}} \quad (3.4)$$

Cada una de estas métricas aporta un enfoque complementario para la evaluación. El MSE y el RMSE descritos en la ecuación (3.1) y(3.2 respectivamente, al penalizar errores más grandes, son especialmente útiles en modelos donde es necesario reducir al mínimo las desviaciones significativas, mientras que el RMSE proporciona la métrica en la misma escala que los datos, facilitando una interpretación intuitiva de los errores. Por otro lado, MAPE permite analizar los errores de manera relativa, lo cual es útil cuando se necesita observar el rendimiento en términos de proporciones o porcentajes. Sin embargo, dado que MAPE descrito en la ecuación (3.3 puede amplificar los errores en valores bajos, sMAPE se emplea como una alternativa simétrica que mitiga este sesgo, proporcionando una evaluación comparativa más balanceada, este se describe en la ecuación (3.4. Para todos los casos anteriores, y_i refiere a la observación real en el periodo, $\hat{y}_i$ refiere a la observación estimada para ese periodo y N refiere a la cantidad total de periodos evaluados

Este conjunto de métricas permite una evaluación robusta de los modelos, abordando los distintos aspectos del error y proporcionando un amplio panorama de la capacidad predictiva del sistema. Al utilizar sMAPE, en particular, se obtiene una perspectiva adicional que permite identificar sesgos y fallas sustanciales en el MAPE, brindando una valoración más

equilibrada y precisa de las capacidades de cada modelo de pronóstico en el contexto del cambio climático.

3.3 Conjuntos de datos

A continuación, se describen los conjuntos de datos utilizados en el proceso experimental, los conjuntos de datos usados provienen de fuentes públicas, aunque su acceso puede verse obstaculizado, ya que algunas bases de datos oficiales solo están disponibles temporalmente o requieren permisos de acceso específicos. Para facilitar la reproducibilidad este trabajo, los datos se compilaron y alojaron en un repositorio, asegurando la disponibilidad de los mismos. A continuación, se detalla la fuente original de cada conjunto de datos y en la sección de anexos se proporciona el enlace directo al repositorio que contiene esta compilación para su acceso y uso en futuros análisis o validaciones experimentales.

3.3.1 Temperaturas máximas y mínimas del sur de la CDMX

Para el análisis y pronóstico de la temperatura en el sur de la Ciudad de México, se utilizaron registros diarios de temperaturas máxima y mínima, cubriendo un periodo desde el 1 de febrero de 1950 hasta el 31 de agosto de 2017. Estos datos fueron obtenidos de estaciones meteorológicas operadas por el Servicio Meteorológico Nacional (SMN) y localizadas en las alcaldías de Milpa Alta, Tlalpan y Xochimilco [85]. Estas alcaldías representan áreas de clima templado subhúmedo con características topográficas similares, favoreciendo la homogeneidad de los datos en términos de condiciones ambientales.

Esta región, mostrada en la Figura 3.1 ha mostrado una baja influencia del fenómeno de las islas de calor, lo que permite que los registros de temperatura reflejen menos distorsión debida a factores urbanos como la acumulación de calor, típica en áreas con alta densidad de edificaciones y superficies de concreto. De acuerdo con datos del Instituto Nacional de Estadística y Geografía (INEGI), estas zonas mantienen un microclima más estable en comparación con otras áreas urbanas densas [86].

Figura 3.1. Región de observación de temperaturas para el set de la CDMX

La recopilación de datos en esta zona goza de una alta disponibilidad debido a su ubicación en la capital, aunque se identificaron valores faltantes en el conjunto de datos. Estos vacíos fueron corregidos mediante técnicas de imputación para asegurar la integridad y continuidad de la serie temporal, favoreciendo así la precisión en las etapas posteriores de análisis y modelado del pronóstico.

3.3.2 Temperaturas de Copernicus

El conjunto de datos empleado para evaluar el desempeño de la metodología propuesta está compuesto por observaciones históricas de temperatura media diaria (en grados Celsius) correspondientes a diez ciudades representativas de México.[87] Estas ciudades —Ciudad de México (CDMX), Ciudad Nezahualcóyotl, Guadalajara, Juárez, León, Monterrey, Puebla, Tampico, Tijuana y Toluca— fueron seleccionadas por su densidad poblacional, relevancia económica y vulnerabilidad frente a fenómenos asociados al cambio climático.

Los registros abarcan el periodo comprendido entre los años 1980 y 2020, con una resolución temporal diaria, proporcionando un total de 14,884 observaciones por ciudad. Los datos fueron extraídos del servicio Copernicus Climate Change Service, específicamente del producto ERA5-Land, el cual constituye una reconstrucción de alta fidelidad del estado

atmosférico global basada en un sistema de asimilación de datos que integra observaciones satelitales y modelos físicos del clima. La descarga de los datos se realizó a través de un repositorio consolidado en la plataforma Kaggle, que compila datos meteorológicos para más de 1000 ciudades del mundo.

3.3.3 Conjuntos multivariados en México

Este conjunto de datos utilizado corresponde a series multivariadas; es decir, series que tienen diferentes atributos sucediendo a par y que pueden otorgar información relevante al modelo. En este caso, el conjunto de series de tiempo multivariadas correspondiente a cuatro ciudades estratégicas de México: Monterrey, Guadalajara, Tijuana y Tampico.[88] La elección de estas ciudades responde a diversos criterios de relevancia: son metrópolis densamente pobladas, presentan climas contrastantes y han sido identificadas como altamente vulnerables ante los efectos del cambio climático, tales como eventos extremos de temperatura, huracanes, tormentas severas o periodos prolongados de sequía.

Según el Índice de Riesgo Climático Global 2023 publicado por Germanwatch, México ocupa la posición 31 a nivel mundial, reflejando un alto nivel de exposición y vulnerabilidad frente a fenómenos meteorológicos extremos. Esta situación se ve agravada por la complejidad geográfica del país, que cuenta con costas en dos océanos, una vasta extensión territorial y una notable diversidad de ecosistemas climáticos.

En este trabajo, las series de tiempo fueron construidas a partir de registros diarios recopilados por estaciones meteorológicas ubicadas en los aeropuertos de las ciudades seleccionadas. El periodo de observación abarca de enero de 2012 a diciembre de 2022.
Cada serie multivariada incorpora un total de 12 a 14 variables climáticas, seleccionadas por su relevancia tanto para la predicción meteorológica como para la evaluación de riesgos climáticos. Estas variables son: temperatura: máxima, mínima y promedio; punto de rocío: máximo, promedio y mínimo; humedad relativa: máxima y promedio; velocidad del viento: máxima, promedio y mínima; presión atmosférica: máxima, promedio y mínima, además de observaciones de gases de efecto invernadero.

3.4 Metodología HELI

El análisis del cambio climático y el impacto de este requiere de una comprensión profunda del comportamiento de las variables atmosféricas, siendo la temperatura una variable clave para evaluar las tendencias en este campo. Sin embargo, el comportamiento de las variables atmosféricas, y particularmente el de la temperatura, está influenciado por múltiples factores a corto plazo como la sensación térmica, el viento, la nubosidad, entre otros. Estos factores dificultan el aprendizaje de los modelos de pronóstico, incrementando el tiempo de ajuste y tiempo de procesamiento necesarios para alcanzar resultados precisos. Además, como se revisó anteriormente, la literatura sugiere que no existe un método de pronóstico universalmente superior; por lo tanto, el uso de un único modelo de regresión puede no capturar completamente el comportamiento de esta variable. Esto justifica el desarrollo de estrategias de ensamble que incorporen distintos enfoques de pronóstico para mejorar la precisión general.

En respuesta a esto, se desarrolló la metodología HELI, un enfoque de ensamble que incorpora múltiples modelos de pronóstico y los optimiza para obtener un resultado que iguale o supere al del mejor método individual. HELI emplea una serie de estrategias de regresión, cada una adaptada y ajustada al conjunto de datos, que luego son evaluadas y ranqueadas en función de su desempeño. Posteriormente, los modelos son ponderados y ensamblados a través de un esquema de optimización, logrando una robustez que minimiza la variabilidad inherente de las observaciones. [81]

La arquitectura de HELI, ilustrada en la Figura 3.2. Descripción general de la metodología HELI presenta un diseño modular y extensible, que permite integrar o eliminar o sustituir algoritmos de regresión según las necesidades del problema. Su estructura también posibilita una alta paralelización en la fase de ajuste, lo cual es particularmente ventajoso cuando se trabaja con grandes volúmenes de datos o cuando se requieren pronósticos en tiempo real. La alta modularidad y la capacidad de paralelización son dos de las características más destacadas de esta metodología, permitiendo tanto la adaptabilidad del modelo a diferentes contextos como la eficiencia en su procesamiento.

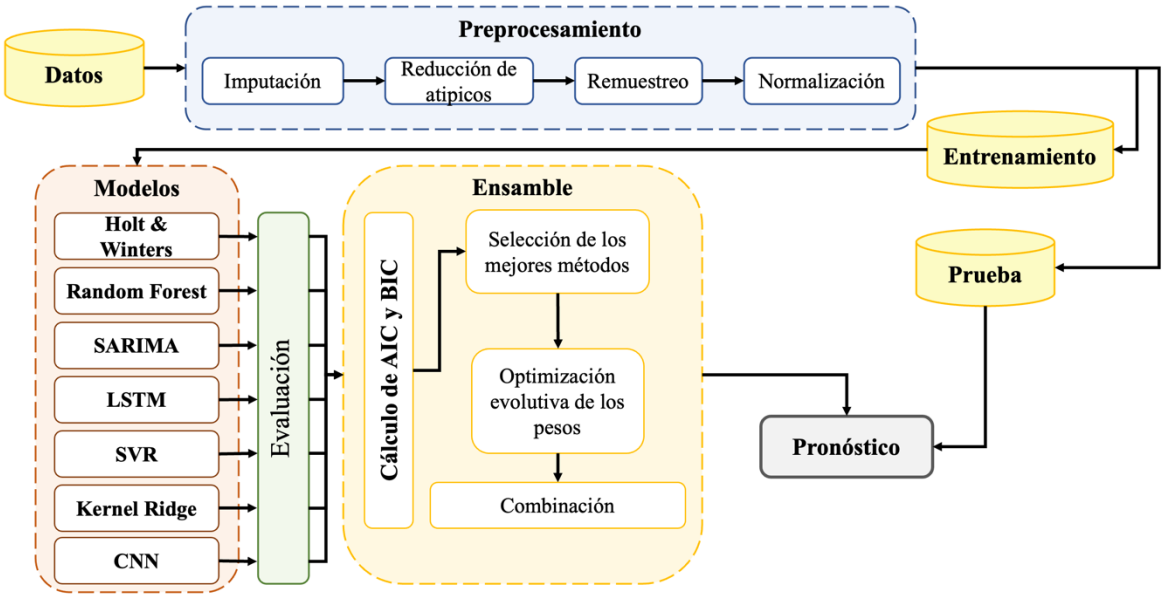


Figura 3.2. Descripción general de la metodología HELI

De esta manera, HELI se convierte en una herramienta versátil y potente para el análisis de fenómenos complejos como los relacionados con el cambio climático, al combinar la fortaleza de múltiples métodos de pronóstico y optimizar su desempeño en un esquema de ensamble robusto.

En la metodología desarrollada, el preprocesamiento de los datos de series de tiempo es esencial para asegurar la consistencia de las observaciones a lo largo del tiempo. Inicialmente, se realiza la imputación de valores faltantes mediante interpolación lineal, un enfoque que permite mantener la continuidad temporal sin introducir complejidad innecesaria. Además, para abordar los valores atípicos que pueden distorsionar el análisis, se emplea una estrategia estacional de 3-sigma, que garantiza que solo los valores inusuales dentro de su contexto estacional sean ajustados, minimizando el sesgo por periodos estacionales inusuales.

Con fines comparativos con metodologías reportadas en la literatura, se incluye un módulo opcional de remuestreo, que permite convertir la periodicidad de los datos de diaria a semanal, ajustando la granularidad de la serie de tiempo sin comprometer la

representatividad, este método esta descrito en la sección anterior. La última etapa del preprocesamiento normaliza los datos al intervalo [1, 2], desplazando la serie para evitar problemas en el ajuste de modelos de pronóstico y en el cálculo de métricas de desempeño. Esta normalización proporciona estabilidad numérica y facilita la comparación entre resultados de distintos modelos.

Posteriormente, se entrenan siete modelos de pronóstico que incluyen tanto técnicas clásicas como enfoques de machine learning, proporcionando a todos la misma oportunidad de participar en el proceso de ensamblado posterior. Cada uno de estos modelos puede requerir ajustes específicos de preprocesamiento, adaptándose según la arquitectura del modelo en cuestión. La Tabla 3.1 presenta una lista de los hiperparámetros óptimos identificados para cada modelo si aplica en estos, lo cual es crucial para garantizar un rendimiento adecuado. Cabe señalar que el comportamiento modular permite ajustar la cantidad de modelos según el problema en estudio, aunque es importante considerar que un número elevado de modelos puede incrementar la complejidad computacional sin necesariamente mejorar la precisión.

Tabla 3.1. Hiperparámetros de los algoritmos de pronóstico usados en HELI

Hiperparámetro	**Configuración**
Bosques Aleatorios	
Criterio	Friedman MSE
Estimadores	23000
Características máximas	Sqrt
Muestras máximas	0.8
Mínimo de muestras por hoja	10
Mínimo de divisiones por muestra	8
Máquinas de vectores de regresión	
Kernel	Polinomial
Valor epsilon	0.01
Valor C	5
Kernel Ridge	
Kernel	Polinomial
Grado	4
Valor alpha	1
Red neuronal convolucional	
Kernel	(15,15), (5,5)
Filtros	64
Función de pooling	Promedio
Función de activación	ReLU
Ratio de aprendizaje	0.01
Red neuronal LSTM	
Capas LSTM	1
Celdas LSTM por capa	35
Capas MLP (Perceptron multicapa)	4
Neuronas MLP por capa	200
Ratio de aprendizaje	0.005

Para seleccionar los modelos más eficaces, se aplican métricas de desempeño junto con criterios como el BIC y el AIC, que clasifican los modelos tanto por su precisión como por la eficiencia en el ajuste de hiperparámetros. Este ranking no pretende ser una solución final, pero sí ofrece una solución inicial encaminada a la optimización de pesos en el ensamble de los modelos.

Para alcanzar una combinación óptima, un algoritmo evolutivo explora el espacio de soluciones utilizando el ranking inicial y soluciones aleatorias, buscando configuraciones que minimicen el error, en la Tabla 3.2 se muestra la configuración se sus hiperparámetros. Al concluir este proceso, se obtiene un vector de pesos optimizado, que representa las combinaciones idóneas para el ensamble de los modelos. Estos pesos son empleados para generar el pronóstico en el conjunto de prueba, permitiendo una evaluación final del modelo y validando la precisión de la metodología mediante métricas de desempeño.

Tabla 3.2. Hiperparámetros del algoritmo evolutivo para la metodología HELI

Hiperparámetro	**Configuración**
Número de generaciones	Desconocido
Tamaño de la población	500
Tamaño del genotipo	7
Técnica de selección	Ruleta ponderada
Cantidad de padres seleccionados	200
Tipo de cruza	Scattered
Probabilidad de mutación	Adaptativa de 0.9 - 0.2
Tipo de mutación	Doblemente Aleatoria
Permite duplicados	No
Elitismo	50 individuos
Criterio de parada	Pendiente continua 20 generaciones
Función objetivo	Minimizar métrica de error

3.5 Metodología FORSEER

Como se mencionó anteriormente, el enfoque de pronóstico univariable se basa en el uso de una única variable para estimar su propio comportamiento en el futuro. Esto implica que toda la información necesaria para realizar el pronóstico está contenida en la variable, y como consiguiente es un problema de baja dimensionalidad frente a otros enfoques. Las estrategias univariables analizan el comportamiento histórico de la serie para ajustar un modelo de pronóstico, que luego se emplea para estimar valores futuros. Sin embargo, como se discutió en el capítulo anterior, no existe un modelo de pronóstico universalmente superior, y la complejidad inherente de algunas series de tiempo puede hacer que el ajuste de hiperparámetros y el entrenamiento requieran mayor tiempo y recursos.
Esta problemática es evidente en series de tiempo complejas, como las que reflejan la temperatura en el contexto del cambio climático. En estos casos, la serie de tiempo presenta una estacionalidad que puede verse distorsionada por los fenómenos asociados al cambio climático, además de contener un alto nivel de ruido residual. Estas características, en conjunto, hacen que el pronóstico de estas series sea una tarea compleja.

Para abordar esta dificultad, se desarrolló FORSEER (*Forecasting by Selective Ensemble Estimation and Reconstruction*), una metodología de ensamble de componentes en series de tiempo. FORSEER permite descomponer la serie original en múltiples componentes más simples y manejables, las cuales se pronostican de manera independiente mediante diferentes estimadores. [40] Posteriormente, los resultados de estos estimadores se ensamblan y optimizan para producir pronósticos de alta precisión. La metodología puede visualizarse en la Figura 3.3.

La metodología comienza con una fase de filtrado y limpieza de la serie de tiempo, en la cual se abordan los problemas comunes en este tipo de datos. Primero, los huecos presentes en la serie se rellenan mediante interpolación lineal, asegurando continuidad en los datos. Para la reducción de ruido, se utiliza la descomposición en valores singulares (SVD), como se detalló en el capítulo anterior. Este método permite conservar el 95% de la varianza de los

componentes principales, eliminando aquellos que aportan escasa información, típicamente considerados como ruido.

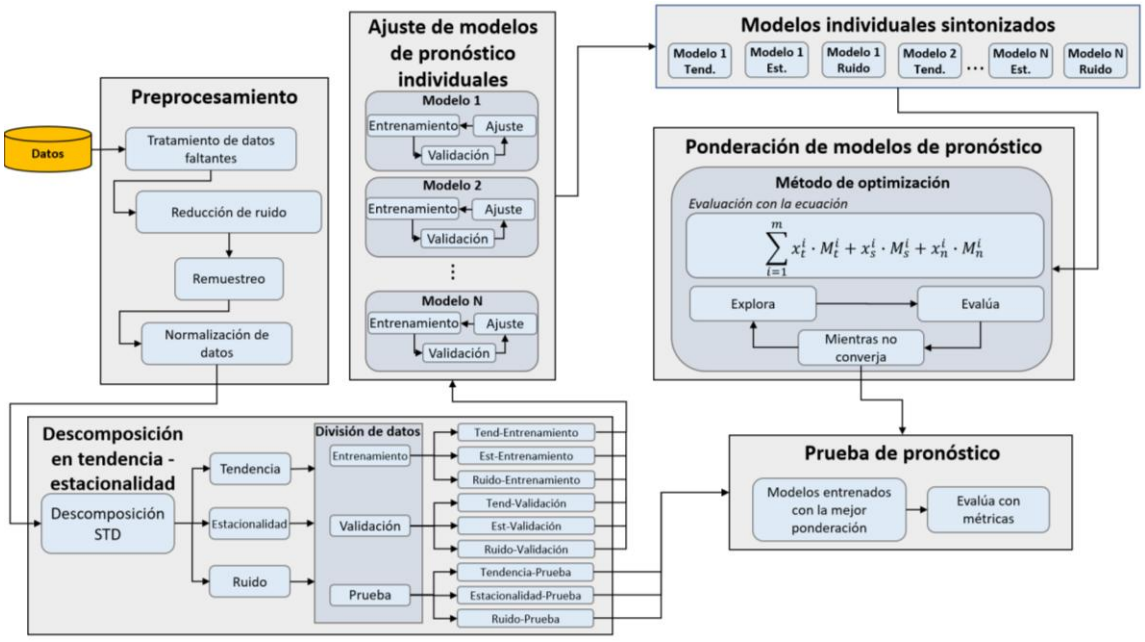


Figura 3.3. Descripción general de la Metodologia FORSEER

Con fines comparativos y para ajustar la periodicidad a diferentes enfoques del estado del arte, se incorpora un módulo de remuestreo que transforma datos diarios en observaciones semanales. Luego, se aplica una normalización en un intervalo desplazado $(1, 2)$ para prevenir posibles problemas en las métricas de evaluación y los modelos de pronóstico. Este proceso prepara la serie de tiempo para el ajuste del modelo sin afectar la estructura de sus valores.

Una vez completado este preprocesamiento, la serie es descompuesta en tres componentes fundamentales: tendencia, estacionalidad y el componente residual, cada uno traducido en una serie de tiempo individual. Estas tres series representan el comportamiento estructural de la serie original y son esenciales para el análisis en fases posteriores. A continuación, cada componente es segmentado en tres conjuntos: entrenamiento, validación y prueba, que permiten evaluar y optimizar el desempeño del modelo.

Para el proceso de pronóstico, se emplean tres tipos de algoritmos: una red neuronal de memoria a largo corto plazo (LSTM), un modelo de bosques aleatorios (RFR) y una máquina de vectores de regresión (SVR). La arquitectura es modular, permitiendo la inclusión o exclusión de modelos adicionales; sin embargo, incluir múltiples algoritmos no necesariamente garantiza un aumento en la eficiencia y puede incrementar significativamente el tiempo de ejecución. Los hiperparámetros optimizados para cada modelo de pronóstico se resumen en la Tabla 3.3 reflejando los ajustes específicos realizados. Este esquema permite alta paralelización en el entrenamiento y aprovecha los recursos de hardware disponibles, maximizando la eficiencia computacional.

Tabla 3.3. Hiperparámetros de los algoritmos de pronóstico usados en FORSEER.

	Tendencia	**Estacionalidad**	**Residual**
Bosques Aleatorios (RFR)			
Criterio	Error cuadrático		
Estimadores	10 000		
Características máximas	Todas		
Muestras máximas	0.8	0.7	0.8
Mínimo de muestras por hoja	1		
Mínimo de divisiones por muestra	2		
Red neuronal LSTM			
Capas LSTM	1		
Celdas LSTM	12	24	36
Capas MLP	1	4	4
Neuronas MLP	100		
Dropout MLP	-	0.2	
Función de activación	Tanh		
Optimizador	Adam		RMSProp
Ratio de Aprendizaje	0.001		
Épocas	500		
Tamaño del Lote	10		5
Parada prematura	35 épocas sin mejora		
Máquinas de soporte de regresión			
Kernel	Lineal	Polinomial	RBF
Grado	-	2	-
Epsilon	0.1		
Gamma	-	Escala	
C	1.0		

Una vez entrenados los modelos, cada regresor ajusta su curva de predicción para cada componente, generando así un total de 3*R modelos, donde R representa el número de regresores empleados. Sin embargo, no todos los modelos ni componentes poseen igual relevancia en el pronóstico final, por lo cual es crucial optimizar su peso en el ensamble. Para

ello, se utiliza un algoritmo evolutivo que asigna ponderaciones a cada modelo en un rango de [-100, 100], lo que permite explorar combinaciones no triviales de los resultados individuales. Este algoritmo evolutivo busca asegurar que el ensamble sea al menos tan bueno como el mejor método de pronóstico de individual entrenado.

Tabla 3.4. Hiperparámetros del algoritmo evolutivo usado en FORSEER

Hiperparámetro	Configuración
Número de generaciones base	1000
Tamaño de la población	500
Tamaño del genotipo	9
Técnica de selección	Ruleta ponderada
Cantidad de padres seleccionados	250
Tipo de cruza	Scattered
Probabilidad de mutación	Adaptativa de 0.9 - 0.2
Tipo de mutación	Doblemente Aleatoria
Permite duplicados	No
Elitismo	50 individuos
Criterio de parada	Pendiente continua
Función objetivo	Minimizar métrica de error

Los hiperparámetros de este algoritmo evolutivo están detallados en la Tabla 3.4. La condición de convergencia se establece cuando la mejor solución presenta una pendiente cercana a cero respecto a la generación previa durante cien generaciones consecutivas, indicando que se ha alcanzado un nivel de ajuste óptimo.

Con el proceso evolutivo terminado, se obtienen los pesos o la participación de los algoritmos optima o cercana a la óptima. El pronóstico final se genera utilizando el conjunto de prueba, permitiendo una evaluación robusta de la efectividad del modelo desarrollado.

3.6 Metodología TAE-Predict

De acuerdo con las conclusiones de las convenciones marco contra el cambio climático, la temperatura se reconoce como una de las variables fundamentales para describir la evolución de los fenómenos asociados al cambio climático. No obstante, esta variable no actúa de manera aislada, sino que se ve influenciada por un conjunto muy grande de factores o variables. Si bien existen enfoques de pronóstico univariables, como los descritos anteriormente, la incorporación de múltiples variables puede contribuir a la generación de modelos con mayor capacidad descriptiva y predictiva.

Sin embargo, la inclusión de múltiples variables también puede generar problemas adicionales. La naturaleza de los fenómenos de cambio climático involucra variables de origen natural, atmosférico y también variables humanas. Aunque no siempre se dispone de un conjunto amplio y confiable de variables, en muchos casos la cantidad de información disponible resulta considerable. Esta situación puede derivar en un incremento de la dimensionalidad del sistema, lo cual impacta directamente en el tiempo de ajuste de los modelos y, en consecuencia, dificulta el proceso de aprendizaje, sin garantizar que una mayor cantidad de variables genere pronóstico de calidad.

Ante este escenario, se propone la metodología TAE-Predict, la cual es un enfoque de pronóstico multivariable que incorpora un mecanismo de selección de variables relevantes, así como un esquema de ensamble de diferentes estrategias de pronóstico.[89] Dicho esquema se apoya en una estrategia evolutiva basada en enjambre de partículas, con el objetivo de mejorar la eficiencia del aprendizaje y la calidad del pronóstico obtenido garantizando obtener resultados al menos tan buenos como el mejor método de pronóstico individual.

Este proceso puede visualizarse en la Figura 3.4, donde se describen de manera secuencial cada uno de los bloques de procesamiento. Es importante destacar que la metodología propuesta sigue un enfoque modular, en el cual cada una de las secciones puede ser reemplazada o extendida total o parcialmente mediante otras estrategias reportadas en la

literatura, permitiendo su adaptación al problema o contexto especifico. De igual forma, el proceso de ajuste de los modelos individuales puede paralelizarse, con el objetivo de mejorar los tiempos de cómputo principalmente mediante el uso de técnicas de cómputo masivo.

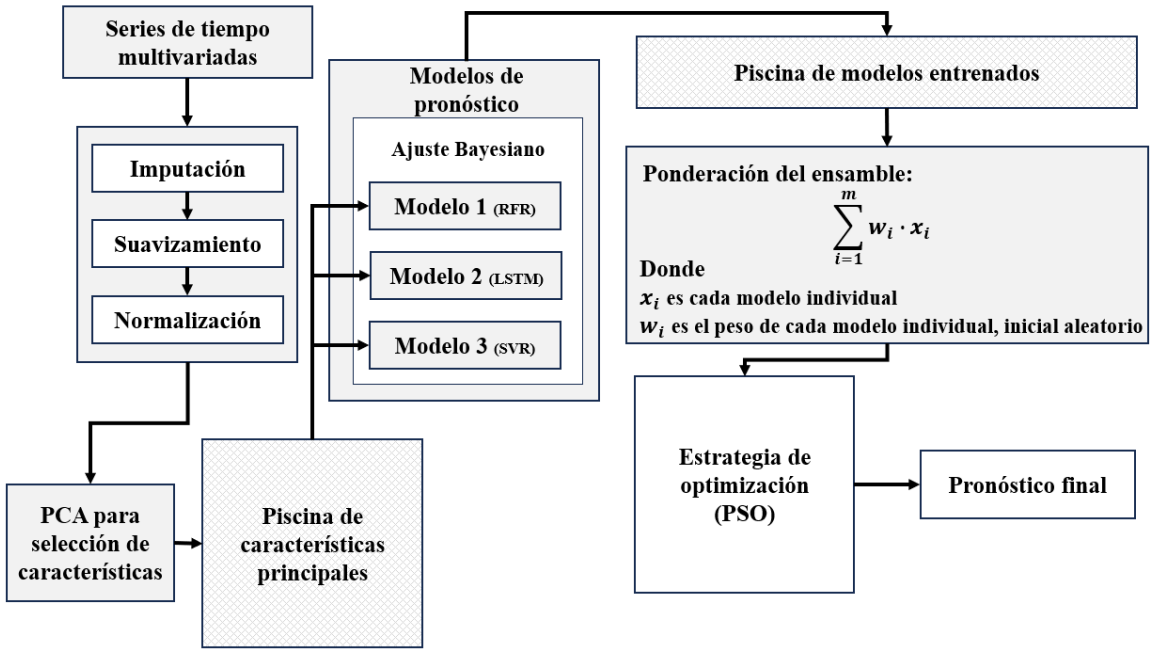


Figura 3.4. Descripción general de la metodologia TAE-Predict

La primera fase corresponde a un preprocesamiento simple. En esta etapa, los datos faltantes se rellenan mediante una estrategia de imputación cuadrática, la cual fue seleccionada a partir de un análisis previo de los datos utilizados en la experimentación, resultando en la técnica más adecuada. Posteriormente, se aplica un suavizamiento basado en ventanas deslizantes, que permite atenuar variaciones abruptas considerando el comportamiento de los valores vecinos. A continuación, el ruido presente en las series se reduce mediante descomposición en valores singulares. Finalmente, debido a la presencia de múltiples series en distintas escalas debido a la naturaleza de las variables, se aplica una estrategia de normalización con el fin de unificar los datos en una misma escala.

La estrategia de PCA utilizada es una variante de la tradicional, y tiene como objetivo conservar las variables relevantes, por lo que no se redefinen los ejes del sistema original. Aqui, se busca preservar al menos el 95 % de la varianza explicada, asegurando que la mayor parte de la información en los datos originales se mantenga en el conjunto reducido, considerando las variables descartadas como información que poco aporta al modelo o se encuentra repetida y su presencia se traduce en ruido. Como resultado, se construye un subconjunto de variables relevantes que posteriormente son utilizadas como entradas por los modelos de pronóstico en la siguiente fase, este conjunto constituye una piscina de variables relevantes.

En esta etapa se emplean tres estrategias de pronóstico: RFR, LSTM y SVR, las cuales pueden generalizar características diferentes en los datos. No se reportan hiperparámetros específicos, ya que para cada modelo se utiliza una estrategia de ajuste bayesiano para obtener una buena configuración del modelo de manera automática. Una vez completado el proceso de entrenamiento, se genera una piscina de modelos ajustados.
Finalmente, el método de ensamble combina los modelos individuales mediante un esquema de ponderación, con el objetivo de obtener un desempeño superior al alcanzado por cada modelo de forma individual y por otras estrategias de ensamble convencionales. En este caso, el algoritmo PSO se encarga de optimizar los pesos asociados a cada modelo, explorando el espacio de soluciones para identificar la combinación que maximiza el desempeño global del sistema de pronóstico. De esta manera, el ensamble aprovecha las fortalezas particulares de cada modelo y mitiga sus limitaciones individuales.

3.7 Metodología GEANY

Los fenómenos de cambio climático suelen abordarse mediante dos enfoques de predicción principales: las estrategias univariables y multivariables. En el enfoque univariable, la información a pronosticar depende principalmente del comportamiento pasado de la misma variable, sin integrar efectos externos. En el enfoque multivariable, sin embargo, el pronóstico no solo es influenciado por la propia variable en periodos anteriores, sino también por otros factores relacionados, comúnmente llamados indicadores en estudios de cambio

climático. Estos indicadores incluyen variables atmosféricas, de emisiones, geográficas y naturales, entre otros.

Para optimizar el ajuste del pronóstico, los algoritmos de predicción multivariable aprovechan la información de múltiples indicadores climáticos, lo cual podría sugerir que incluir una gran cantidad de estos factores mejoraría la precisión del pronóstico. Sin embargo, esto presenta tres desafíos, la disponibilidad limitada de datos, la inclusión de variables no significativas y el aumento en la dimensionalidad.

En un modelo multivariable, algunas variables pueden aportar poca o ninguna información útil, ya que muchas de ellas están altamente correlacionadas, lo que genera redundancia. Además, cada variable adicional incrementa la dimensionalidad del espacio de ajuste, complicando el proceso y aumentando los recursos computacionales necesarios.
Frente a esta problemática, proponemos GEANY (*Generalized Ensemble Application for Components Analysis*), una metodología para la selección de variables relevantes en el contexto del cambio climático. En la Figura 3.5, se ilustra el funcionamiento general de esta metodología.

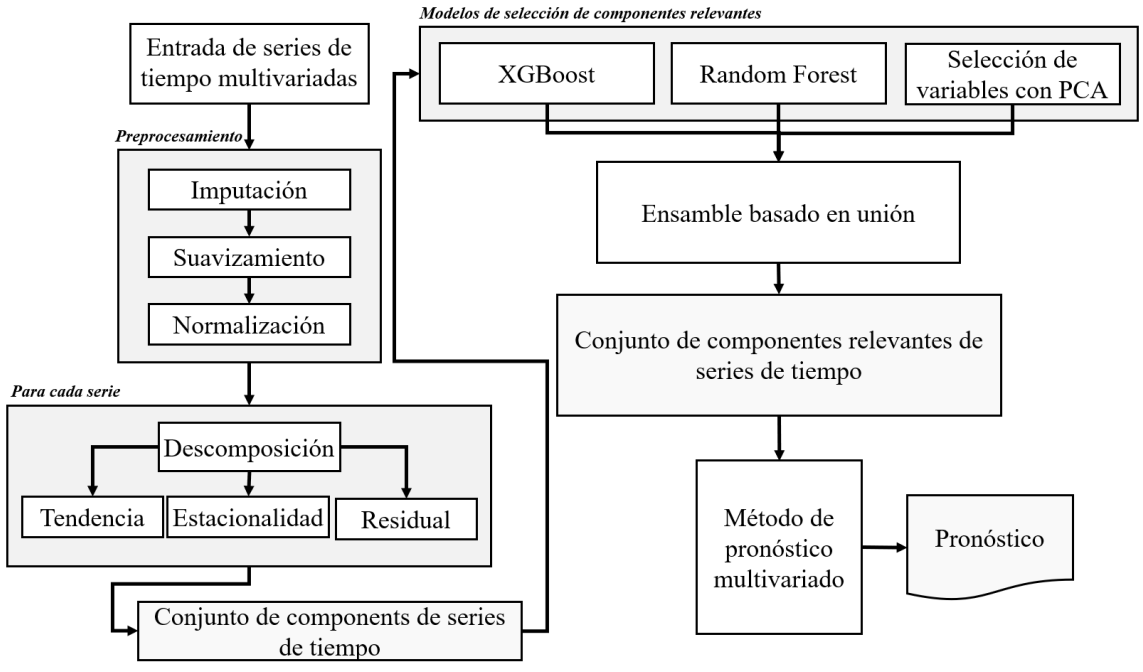


Figura 3.5. Descripción general de la metodología GEANY propuesta

En este caso, las series de tiempo multivariadas pasan por un proceso de filtrado para corregir datos faltantes, outliers y otros problemas que puedan comprometer la calidad de los datos. Este paso es fundamental, ya que, como se mencionó en la sección anterior, una serie de tiempo preprocesada facilita significativamente el ajuste del modelo de pronóstico.

La metodología propone un proceso de imputación de datos mediante interpolación lineal. Para el tratamiento de valores atípicos, se empleó la estrategia de 3-sigma junto con ventanas móviles para suavizar el comportamiento de la serie. La normalización se realizó con una escala de unidad desplazada entre 1,2. Estas fases son modulares y pueden ajustarse según el contexto específico del problema.

Para extraer los componentes de tendencia, estacionalidad y ruido de cada serie de tiempo, se aplica un proceso de descomposición. Este procedimiento incrementa tres veces el número de series, aumentando así la dimensionalidad del problema. Sin embargo, dicho incremento permite aislar los patrones específicos de cada componente, lo cual facilita la selección y el descarte de aquellos que no aportan al modelo en etapas posteriores.
Como resultado, se genera un pool de componentes de series de tiempo, tres veces mayor que el conjunto original, sobre el cual se realizará la selección de los componentes más relevantes para el sistema. En este punto debe definirse e integrarse en el pool la variable objetivo (Target), sobre la cual se evaluará la importancia de los demás componentes

Para esta selección, se implementan diferentes métodos de selección de variables. Dado que la metodología es modular y extensible, pueden aplicarse múltiples métodos de selección, con el correspondiente impacto en la complejidad del sistema. En este trabajo se emplean tres estrategias de selección: Random Forest, XGBoost y PCA, detalladas en la sección anterior. Los hiperparámetros y configuraciones específicas se encuentran en la Tabla 3.5. Hiperparámetros de los algoritmos de selección de variables en GEANY

Tabla 3.5. Hiperparámetros de los algoritmos de selección de variables en GEANY

Hiperparámetro	Configuración
PCA	
Umbral de varianza explicada	92%
Bosques aleatorios para selección de variables	
Estimadores	1000
Criterio	Error cuadrático
Profundidad máxima	10
Mínimo de divisiones por muestra	5
Mínimo de muestras por hoja	2
XGBoost para selección de variables	
Ratio de aprendizaje	0.1
Gamma	1
Profundidad máxima	10

Cada uno de estos modelos selecciona componentes relevantes conforme a sus criterios internos y la variable Target, generando un conjunto de componentes significativos para el modelo. Aunque es probable que los componentes más relevantes aparezcan en todos los conjuntos seleccionados, algunos pueden estar presentes únicamente en ciertos modelos. Para consolidar un único conjunto robusto de componentes, se integran los resultados de cada estrategia mediante una operación de unión, como se ilustra en la Figura 3.6.

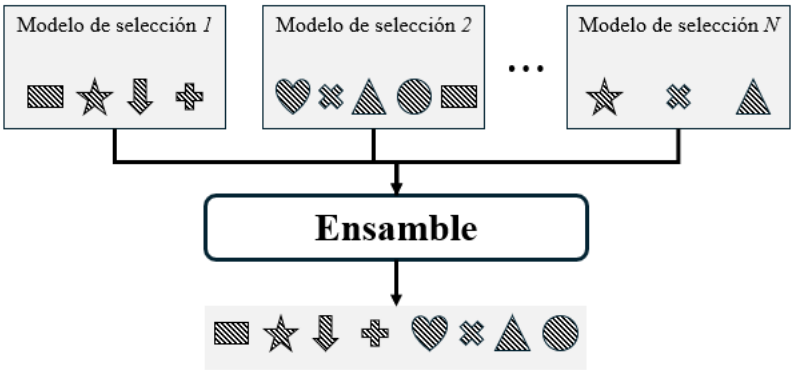


Figura 3.6. Descripción grafica del proceso de ensamble basado en unión

Dado que las metodologías pueden ofrecer conjuntos diferentes, ensamblar los resultados se vuelve esencial para obtener un conjunto robusto y significativo de componentes. Así, proponemos la unión de los componentes relevantes seleccionados por los métodos individuales, denotada por la expresión $\bigcup_{i=1}^{n} A_i = \{x | x \in A_i\}$ donde x representa a todos los elementos presentes en al menos uno de los conjuntos A

Como resultado del proceso de ensamble, se obtiene un pool reducido de componentes relevantes. Este conjunto reducido permite optimizar los recursos computacionales y reducir la complejidad del modelo, conservando la mayor parte de la información relevante de los datos originales. Sobre este nuevo conjunto, se implementa un proceso de pronóstico mediante una red neuronal LSTM, en esta fase de la metodología, se emplea únicamente un método de pronóstico robusto, con el objetivo de evaluar la eficacia de la selección de componentes relevantes y su impacto en la precisión del modelo.

Capítulo 4

4 Resultados

A partir de las metodologías propuestas, se llevó a cabo un proceso exhaustivo de experimentación en el que todos los modelos fueron evaluados con múltiples conjuntos de datos, específicamente seleccionados para representar la naturaleza del problema a resolver.

En esta sección, se describen los resultados obtenidos en cada caso, destacando el rendimiento de los modelos y presentando gráficos detallados que permiten visualizar el comportamiento de las predicciones, errores y métricas de ajuste para cada configuración.

Además de los resultados obtenidos mediante nuestras metodologías principales, se incluyen los resultados de otras estrategias y variaciones de los métodos desarrollados en este trabajo. También se comparan estos hallazgos con los resultados de estudios relevantes en el estado del arte, lo cual permite analizar el rendimiento de los algoritmos propuestos en un contexto más amplio.

Por último, en aquellos casos donde es posible, se presentan los resultados de pruebas de hipótesis que evalúan la efectividad de las metodologías propuestas. Estas pruebas estadísticamente fundamentadas aportan evidencia adicional sobre el desempeño de los algoritmos y su capacidad para mejorar la precisión del pronóstico en comparación con métodos estándar. Este conjunto de resultados y validaciones ofrece una visión integral sobre la eficacia de la selección de componentes relevantes y la robustez de los modelos propuestos.

4.1 Resultados obtenidos de la metodología HELI

A continuación, se presentan los resultados obtenidos tras la implementación de la metodología descrita en la sección anterior. Para garantizar la confiabilidad de los resultados, cada uno de los algoritmos fue ejecutado al menos 30 veces bajo las mismas condiciones experimentales y utilizando diferentes semillas aleatorias. Esta repetición permitió explorar la variabilidad de los resultados y evaluar su estabilidad. Cabe destacar que los resultados aquí reportados han sido también reportados en el artículo correspondiente asociado a este trabajo.

4.1.1 Conjunto de datos

En esta metodología se utilizan series de tiempo univariables, es decir, aquellas en las que la información a predecir proviene de una sola variable medida a lo largo del tiempo. En este caso, se trabajó con dos series correspondientes a la temperatura máxima y mínima, expresadas en grados Celsius, registradas en la zona sur de la Ciudad de México. La descripción detallada de estas series se presenta en la sección 3.3.

4.1.2 Resultados obtenidos

En esta sección se presentan los resultados obtenidos durante la primera fase experimental, cuyo objetivo fue ajustar y optimizar los métodos de pronóstico de forma individual. Para reportar su rendimiento se utilizó la métrica sMAPE, aplicada sobre un conjunto de validación que no se utilizó durante el entrenamiento. Esta evaluación permitió observar la capacidad de generalización de cada modelo. En la Tabla 4.1 se mostrarán los resultados obtenidos por cada modelo en términos de sMAPE para los dos conjuntos climáticos de temperatura máxima y mínima.

Tabla 4.1. Resultados obtenidos del mejor ajuste de los métodos individuales

Modelos	**Temperatura mínima**	**Temperatura máxima**
SARIMA	25.42%	37.40%
Holt & Winters	24.8%	24.45%
CNN	9.63%	7.24%
LSTM	**11.02%**	**6.93%**
Random Forest	**11.74%**	**6.87%**
SVR	**10.23%**	**6.66%**
Kernel Ridge	19.69%	8.41%

Los métodos basados en aprendizaje automático, como las redes CNN, las redes LSTM y RFR, presentaron en general un mejor desempeño frente a los métodos clásicos. Este resultado se debe a su capacidad de adaptarse a la complejidad de los datos y al ruido natural presente en las series de temperatura.

Entre los modelos evaluados, CNN y SVR fueron los que obtuvieron los mejores resultados en la mayoría de los casos. Sin embargo, esto no implica que sean los mejores para cualquier problema, ya que el teorema "No Free Lunch" nos recuerda que no existe un modelo universalmente superior para todos los tipos de series o contextos.

Un punto importante es que modelos como SVR y Kernel Ridge lograron un desempeño competitivo utilizando menos hiperparámetros, lo cual puede ser una ventaja al reducir el riesgo de sobreajuste y facilitar la implementación en escenarios con pocos recursos computacionales.

Posteriormente, se evaluaron distintas estrategias de ensamble. En la Tabla 4.2 se resumen los resultados de combinar los modelos individuales mediante métodos de ensamble, incluyendo un promedio uniforme, BIC, AIC y una combinación optimizada a través de un algoritmo evolutivo.

Tabla 4.2. Resultados obtenidos por los métodos de ensamble

Ponderaciones	**Temperatura mínima**	**Temperatura máxima**
Uniforme	19.45%	14.12%
AIC	11.69%	8.45%
BIC	11.52%	6.23%
HELI	**9.12%**	**5.87%**

Se observó que el promedio uniforme no mejora significativamente el rendimiento, e incluso en algunos casos lo empeora, debido a que no considera las diferencias en la capacidad de generalización entre modelos. El uso de otras estrategias de ponderación como BIC o AIC pueden mejorar el resultado, pero no garantizan una combinación optima. En cambio, la estrategia optimizada mediante ponderación sí logró mejorar los resultados al asignar pesos adecuados a cada modelo, superando tanto a los individuales como al ensamble uniforme.

También se realizó una comparación con el método SMYL, considerado una de las estrategias más sólidas del estado del arte y campeona del concurso de pronóstico M4. La Tabla 4.3 presenta el comparativo entre SMYL y HELI, tanto en términos de sMAPE como de tiempo de entrenamiento promedio.

Tabla 4.3. Comparativa de la metodología HELI contra SMYL

Metodologías	**Temperatura mínima**	**Temperatura máxima**
SMYL	**8.78%**	**5.13%**
HELI	9.54%	6.02%

Aunque SMYL tuvo una ligera ventaja en algunas métricas, al aplicar la prueba estadística de Wilcoxon no se encontraron diferencias significativas. La Tabla 4.4 muestra los valores de esta prueba, donde se confirma que ambos métodos pueden considerarse equivalentes al no existir diferencia significativa entre ambos.

Tabla 4.4. Parámetros de la prueba de hipótesis aplicada

Alpha	0.05
z-statistic	2.126
p-value	0.0554

Un aspecto destacable de HELI es su eficiencia en términos de tiempo. Mientras que SMYL puede tardar hasta 25 minutos si no cuenta con información previa sobre la serie, HELI mantiene un tiempo de ejecución promedio constante de 20 minutos, sin depender de un historial previo. Esto lo convierte en una alternativa más práctica para escenarios donde se requieren resultados rápidos sin comprometer la calidad.

Además de las pruebas anteriores, se realizó un experimento para evaluar la capacidad del modelo HELI de realizar pronósticos a largo plazo. Para ello, se extendió la serie original hasta el año 2028, generando predicciones equivalentes a 521 semanas (diez años) a futuro.

En la Tabla 4.5 se muestran los valores reales observados entre 2018 y 2023, comparados con los valores estimados por el modelo.

Tabla 4.5. Comparativa entre los valores observados y estimados a seis años

Año	**Temperatura mínima**		**Temperatura máxima**	
	Observado	**Estimado**	**Observado**	**Estimado**
2018	10.2°	10.5°	23.5°	22.7°
2019	10°	10.6°	23.3°	22.9°
2020	10.1°	10.4°	23.2°	22.7°
2021	9.8°	10.3°	23°	22.8°
2022	9.9°	10.4°	23.1°	22.8°
2023	10°	10.2°	23.2°	23°

Los resultados muestran una buena coincidencia entre los valores estimados y los reales, lo cual confirma la capacidad del modelo para mantener la coherencia de los patrones a lo largo del tiempo. La Figura 4.1 mostrará la evolución de las temperaturas máxima y mínima estimadas, donde puede apreciarse una ligera tendencia a incrementar la diferencia entre ambas.

Figura 4.1. Pronóstico realizado con la HELI 10 años hacia el futuro

Este hallazgo puede ser indicio de cambios en los patrones climáticos del sur de la Ciudad de México, posiblemente relacionados con fenómenos asociados al cambio climático. Este tipo de observaciones refuerzan la utilidad de los modelos de pronóstico como herramientas de apoyo para el análisis ambiental y la toma de decisiones de largo plazo.

4.2 Resultados obtenidos de la metodología FORSEER

A continuación, se detallan los resultados obtenidos al aplicar la metodología FORSEER, expuesta previamente. Con el fin de asegurar la robustez de los hallazgos, cada uno de los modelos fue ejecutado al menos 30 veces bajo condiciones experimentales controladas, empleando distintas semillas aleatorias. Esta estrategia permitió no solo estimar la variabilidad inherente a cada algoritmo, sino también analizar la consistencia de su desempeño. Es importante mencionar que estos resultados coinciden con los presentados en el artículo científico derivado del presente estudio.

4.2.1 Conjunto de datos

Para evaluar el desempeño de la metodología propuesta, se empleó un conjunto de datos de temperatura atmosférica en México. En total, se analizaron doce series: diez asociadas a ciudades densamente pobladas susceptibles a fenómenos derivados del cambio climático, y dos series adicionales que representan los valores extremos (máximos y mínimos) de temperatura en la Ciudad de México.

Por una parte, las series correspondientes a Ciudad de México (CDMX), Ciudad Neza, Guadalajara, Juárez, León, Monterrey, Puebla, Tampico, Tijuana y Toluca provienen del servicio Copernicus Climate Change Service, y las series CDMX_MIN y CDMX_MAX, obtenidas del Servicio Meteorológico Nacional (SMN), descrito en la sección 3.3.

En todos los casos, la estructura del conjunto de datos fue homogénea. Cada serie se dividió en tres subconjuntos para entrenamiento, validación y prueba, utilizando una proporción del 60%, 20% y 20%, respectivamente.

4.2.2 Resultados obtenidos

En esta sección se presentan los resultados obtenidos a partir de la evaluación experimental de la metodología FORSEER, con el propósito de analizar de manera objetiva y estadísticamente robusta su desempeño en tareas de predicción. La validación se realiza en dos fases: la primera orientada a la reducción de ruido mediante SVD, y la segunda enfocada en la evaluación comparativa de los modelos de pronóstico seleccionados, incluyendo técnicas de ensamblado.

En la primera etapa, se aplicó la técnica de SVD con el fin de mitigar el ruido inherente en las series temporales. Para evaluar cuantitativamente la efectividad del proceso de reducción de ruido, se empleó la prueba de varianza de Levene. Esta prueba permite verificar si existen diferencias estadísticamente significativas en la varianza entre la serie original y la procesada, lo cual constituye un indicio claro de eliminación de ruido.

Las hipótesis planteadas fueron para la hipótesis nula H_0: No existe diferencia significativa entre las varianzas de las series (igualdad de varianzas), para la hipótesis alternativa H_1: Existe diferencia significativa entre las varianzas de las series (desigualdad de varianzas). El nivel de significancia fue establecido en $\alpha = 0.05$. Para el cálculo de la prueba se utilizaron $N = 1690$ observaciones y $k = 2$ grupos (serie original y los nuevos datos). Los resultados se resumen en la Tabla 4.6

Tabla 4.6. Evaluación estadística del proceso de reducción de ruido

Series de tiempo	**valor-p obtenido**
CDMX	$2.33e^{-8}$
Ciudad Nezahualcóyotl	$4.912e^{-14}$
Guadalajara	$1.876e^{-10}$
Ciudad Juárez	$5.781e^{-7}$
León	$8.453e^{-12}$
Monterrey	$3.290e^{-9}$
Puebla	$7.159e^{-13}$
Tampico	$9.014e^{-16}$
Tijuana	$2.667e^{-11}$
Toluca	$6.498e^{-15}$
Min_CDMX	0.0000115
Max_CDMX	$3.442e^{-15}$

Dado que en ambos casos el valor-p es menor al umbral de significancia, se rechaza la hipótesis nula. Esto permite concluir que las varianzas son estadísticamente distintas, validando así la efectividad del proceso de reducción de ruido mediante SVD.

Posteriormente, se realizó una evaluación comparativa de diversos modelos de regresión bajo condiciones estándar, es decir, sin aplicar previamente la descomposición. El objetivo de esta fase fue identificar las tres estrategias de modelado más eficaces, basándose principalmente en sMAPE.

Tabla 4.7. Evaluación de los métodos individuales en el conjunto de validación

Instancias	sMAPE							
	SE	**SARIMA**	**MLP**	**CNN**	**LSTM**	**RFR**	**SVR**	**Kernel Ridge**
CDMX	58%	22%	21%	12%	11%	12%	**11%**	19%
Ciudad Nezahualcóyotl	37%	48%	33%	15%	**9%**	9%	10%	14%
Guadalajara	78%	20%	28%	14%	11%	**9%**	9%	18%
Ciudad Juárez	95%	58%	63%	**10%**	12%	11%	13%	15%
León	68%	45%	55%	18%	**9%**	11%	12%	18%
Monterrey	46%	55%	39%	**9%**	12%	11%	12%	12%
Puebla	55%	49%	52%	10%	**9%**	10%	9%	10%
Tampico	80%	65%	37%	17%	**15%**	17%	15%	17%
Tijuana	71%	28%	31%	9%	**7%**	9%	8%	10%
Toluca	96%	43%	55%	16%	13%	**8%**	12%	11%
Min_CDMX	83%	23%	26%	8%	**8%**	8%	12%	15%
Max_CDMX	89%	44%	53%	5%	5%	4%	**3%**	5%
Media	71%	42%	41%	12%	**10%**	**11%**	**11%**	14%

Aunque se reportan varias métricas (RMSE, MSE, MAPE, sMAPE), el proceso de selección de modelos se basó principalmente en la métrica de sMAPE, aunque el ajuste de los algoritmos se realizó mediante MSE. A partir de estos resultados mostrados en la Tabla 4.7, se eligieron Random Forest Regression (RFR), Support Vector Regression (SVR) y LSTM como componentes del ensamble, debido a su sólido desempeño individual. Aunque CNN y Kernel Ridge muestran métricas comparables, fueron descartados por su mayor complejidad computacional o redundancia funcional.

Una vez seleccionados los modelos, se aplicó el esquema de descomposición-reconstrucción para validar el impacto de los modelos sobre cada componente: tendencia, estacionalidad, ruido y reconstrucción de la serie.

Tabla 4.8. Evaluación del porcentaje de error por componentes.

Instancias	sMAPE											
	LSTM				RFR				SVR			
	T	S	R	Serie	T	S	R	S	T	S	R	Serie
CDMX	2.1	93	**67**	11.0	1	128	204	11.2	**0.4**	**74**	314	11.2
Ciudad Neza.	**0.8**	**83**	**101**	9.6	1.4	89	187	8.4	**0.8**	105	276	10.55
Guadalajara	0.7	67	**179**	8.7	2.4	73	196	8.9	**0.3**	**29**	301	9.8
Ciudad Juárez	**1.2**	91	**106**	9.2	1.7	98	145	8.7	**1.2**	**63**	487	13.68
León	0.9	106	**128**	9.1	2.4	87	308	11.0	**0.5**	**81**	138	12.1
Monterrey	1.5	89	**246**	11.1	1.8	124	278	10.2	**0.7**	**67**	271	12.6
Puebla	0.7	108	**105**	9.2	1.6	106	269	7.5	**0.5**	**61**	236	9.45
Tampico	3.1	**99**	**87**	15.4	2.4	143	367	14.7	**1.4**	107	146	15
Tijuana	1.1	78	**103**	7.0	1.5	86	254	6.8	**0.4**	**53**	136	8.1
Toluca	**0.8**	**87**	**223**	12.6	1.1	106	305	8.57	0.9	102	258	12.1
Min_CDMX	0.3	**40**	**110**	7.7	0.38	174	223	8.16	**0.09**	275	289	14.1
Max_CDMX	0.1	**40**	**204**	3.2	0.26	100	332	3.13	**0.09**	84	271	7.72

En la Tabla 4.8 se observa que, en términos generales, existe una alta variabilidad en los errores asociados a cada componente de la serie de tiempo. Esta variabilidad está estrechamente relacionada con la estabilidad o varianza inherente a cada componente. En particular, el componente tendencia suele presentar una varianza baja debido a su comportamiento suavizado y gradual. En contraste, la estacionalidad, aunque puede tener una varianza relativamente baja, exhibe una dinámica oscilatoria que puede dificultar su modelado debido a patrones complejos o no lineales. Finalmente, el componente residual presenta una varianza considerablemente mayor, lo que refleja su carácter errático y aleatorio, convirtiéndolo en el más desafiante de modelar y predecir con precisión. La predicción de estos componentes puede observase con más detalle en la Figura 4.2.

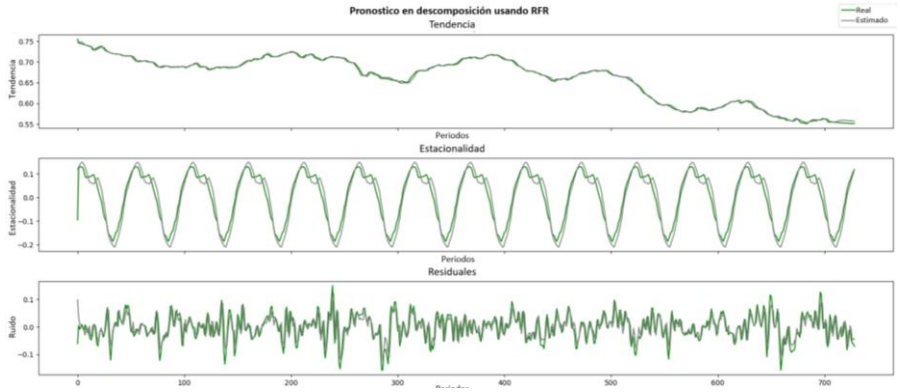


Figura 4.2. Ejemplo del pronóstico en descomposición usando un modelo de RFR

Es importante destacar que, a pesar de que el modelo utilizado es de naturaleza aditiva, los errores individuales de los componentes no deben sumarse directamente. La estimación del error global requiere un proceso de reconstrucción de la serie, en el cual primero se suman los valores pronosticados de cada componente para cada periodo de tiempo, y posteriormente se calcula el error comparando esta reconstrucción con los valores observados reales.

Una vez ajustados los hiperparámetros de los modelos seleccionados para cada componente, se probaron diversas estrategias de ensamble, incluyendo técnicas tradicionales como ponderación uniforme, AIC y BIC, y una estrategia optimizada basada en algoritmos genéticos, que da lugar a la metodología FORSEER.

Tabla 4.9. Comparativa de desempeño de diferentes estrategias de ensamble

Instancias	**sMAPE**			
	Uniforme	**AIC**	**BIC**	**FORSEER**
CDMX	24.1%	12.6%	11.6%	**5.1%**
Ciudad Nezahualcóyotl	33.7%	25.6%	27.1%	**4.8%**
Guadalajara	26.4%	19.7%	24.6%	**6.1%**
Ciudad Juárez	16.7%	12.6%	8.4%	**6.9%**
León	13.7%	15.6%	16.4%	**4.3%**
Monterrey	34.2%	20%	16.4%	**2.9%**
Puebla	24.6%	12.4%	19.1%	**4.6%**
Tampico	66.4%	23.5%	27.8%	**7.8%**
Tijuana	43.1%	12.9%	8.1%	**1.4%**
Toluca	35.6%	15.3%	14.7%	**8.3%**
Min_CDMX	18.42%	8.63%	9.32%	**3.39%**
Max_CDMX	15.12%	6.15%	6.45%	**1.74%**

En la Tabla 4.9 puede observarse que para cualquiera de las instancias evaluadas la metodología propuesta tiene un mejor desempeño frente a las estrategias tradicionales, además como característica propia del mecanismo de ensamble descrito en la sección anterior, en cualquier caso, la metodología propuesta obtendrá resultados al menos tan buenos como el mejor método individual. La superioridad en desempeño puede atribuirse al mecanismo de optimización y búsqueda de pesos óptimos de cada algoritmo.

Tabla 4.10. Comparativa de FORSEER frente a modelos del estado del arte

Instancias	sMAPE				
	RFR	**SVR**	**LSTM**	**SMYL**	**FORSEER**
CDMX	12.8%	13.7%	16.4%	8.99%	**7.74%**
Ciudad Nezahualcóyotl	13.86%	11.85%	15.7%	**7.33%**	7.63%
Guadalajara	12.4%	13.4%	15.1%	9.89%	**7.16%**
Ciudad Juárez	16.1%	14.3%	13.4%	**6.55%**	7.16%
León	14.4%	12.3%	10.3%	8.05%	**7.42%**
Monterrey	13.7%	14.3%	16.4%	7.39%	**7.11%**
Puebla	12%	13.8%	13.1%	**6.2%**	6.82%
Tampico	16%	18.14%	20.43%	**7.38%**	10.08%
Tijuana	11.7%	11.7%	13.8%	6.8%	**5.62%**
Toluca	12.4%	13.9%	10.4%	**5.90%**	7.72%
Min_CDMX	9.17%	9.86%	9.43%	8.78%	**8.53%**
Max_CDMX	6.49%	7.55%	6.28%	5.13%	**4.98%**

Finalmente, se llevó a cabo una comparación con métodos reportados en la literatura reciente, empleando la métrica sMAPE sobre la temperatura mínima y máxima en la CDMX. Los resultados se presentan en la Tabla 4.10. Donde la metodología FORSEER demuestra un rendimiento superior en los conjuntos de datos frente a estrategias individuales especializadas, resultando también competitiva frente a modelos sofisticados como el ensamble de Smyl, ampliamente reconocido en competencias de series de tiempo.

En el caso de Smyl podemos observar que para algunas instancias el algoritmo es ligeramente superior y en otras es ligeramente inferior, por lo que para diferenciar si existe alguna equivalencia entre las estrategias propuestas es conveniente realizar una evaluación estadística, en este caso mediante la prueba no paramétrica de Wilcoxon. Las hipótesis formuladas fueron: H_0: No existe diferencia estadísticamente significativa entre los métodos y H_1: Existe diferencia significativa entre los métodos, aplicado un nivel de significancia de $\alpha = 0.05$, se obtuvo un estadístico $Z = 2.126$ y un valor-p $= 0.0554$. Aunque el valor obtenido está ligeramente por encima del umbral crítico, los resultados indican que no se puede rechazar la hipótesis nula. Por tanto, se concluye que FORSEER presenta un desempeño estadísticamente equivalente al del método Smyl, lo cual respalda su validez y eficacia como metodología de pronóstico

Si bien en cuestión de precisión la metodología FORSEER resulta equivalente a Smyl, valdría la pena evaluar por ejemplo el tiempo de respuesta frente a la complejidad de las series de tiempo evaluadas.

Tabla 4.11. Análisis de complejidad de las series de tiempo y su relación temporal

Instancias	Exponente Hurst	SMYL (Minutos)	FORSEER (Minutos)
CDMX	0.3156	66.76	8.47
Ciudad Nezahualcóyotl	0.2734	68.35	9.53
Guadalajara	0.307	100.06	8.62
Ciudad Juárez	0.2513	104.1	9.47
León	0.278	120	8.5
Monterrey	0.2919	160.91	9.23
Puebla	0.2739	185.4	9.58
Tampico	0.2921	209.05	9.85
Tijuana	0.2906	253.21	8.45
Toluca	0.3	257.4	8.33
Min_CDMX	0.3018	275.17	12.9
Max_CDMX	0.2888	293.63	14.6

La Tabla 4.11 presenta el tiempo promedio de entrenamiento (en minutos) requerido por cada metodología, evaluado sobre doce instancias geográficas, así como el exponente de Hurst utilizado como medida de complejidad de las series temporales. El análisis de la varianza del tiempo de entrenamiento mostró que FORSEER mantiene una varianza notablemente baja. Esta diferencia sugiere que FORSEER ofrece una mayor estabilidad computacional, independientemente de la complejidad de la serie temporal evaluada.

En la Figura 4.3, se representa gráficamente la relación entre el exponente de Hurst y el tiempo de entrenamiento de cada modelo. Se observa que, conforme la complejidad de la serie temporal se incrementa acercándose a 0.5, el tiempo requerido por SMYL aumenta de forma significativa. Este comportamiento refleja la naturaleza intensiva de SMYL, cuyo diseño incorpora una piscina de redes neuronales con el objetivo de maximizar la precisión del pronóstico. Este enfoque, si bien efectivo, implica un costo computacional considerablemente alto a medida que las series se tornan más complejas o se va

incrementando el tamaño de la piscina. Por el contrario, FORSEER no requiere conocimiento previo o redes pre entrenadas.

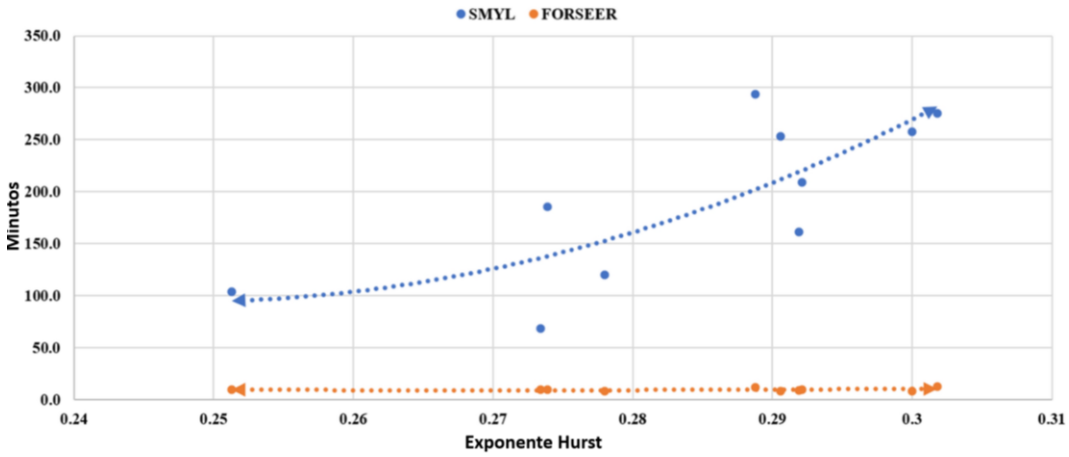


Figura 4.3. Relación entre la complejidad de la serie y el tiempo de entrenamiento

En conjunto, estos resultados permiten concluir que FORSEER representa una alternativa más eficiente desde el punto de vista computacional.

4.3 Resultados obtenidos de la metodología TAE-Predict

A continuación, se presentan los resultados obtenidos mediante la experimentación de la metodología TAE-Predict. Con el objetivo de garantizar la solidez y reproducibilidad de los hallazgos, cada experimento fue ejecutado al menos 30 veces, empleando diferentes semillas aleatorias. Esta estrategia permitió estimar de forma precisa la variabilidad estocástica.

4.3.1 Conjunto de datos

Para evaluar la metodología TAE-Predict, se utilizaron cuatro series de tiempo multivariables correspondientes a las ciudades de Monterrey, Guadalajara, Tijuana y Tampico. Estas series descritas en la sección 3.3 incluyen variables meteorológicas relevantes como temperatura, punto de rocío, humedad, velocidad del viento y presión atmosférica, registradas diariamente entre 2012 y 2022. Cada serie se dividió en subconjuntos de entrenamiento (55%), validación (30%) y prueba (15%) para asegurar una evaluación robusta del desempeño de los modelos.

4.3.2 Resultados obtenidos

En primera instancia, se evaluaron de forma individual los modelos individuales propuestos utilizando como variable objetivo la temperatura, considerada en sus tres formas: máxima, media y mínima.

Tabla 4.12. Evaluación de los modelos individuales y su dispersión

Ciudades	Temperatura	SVR	RFR	LSTM	σ SVR	σ RFR	σ LSTM
Guadalajara	Máxima	2.78%	1.61%	3.27%	0	2	0.04
	Promedio	2.94%	1.64%	3.73%	0	0.11	0.04
	Mínima	2.79%	1.58%	2.87%	0	0.01	0.11
Monterrey	Máxima	3.60%	2.86%	4.65%	0	0.95	0.65
	Promedio	4.18%	5.71%	5.66%	0	0.74	0.07
	Mínima	3.08%	2.66%	6.09%	0	0.46	0.37
Tampico	Máxima	3.16%	3.04%	3.97%	0	0.38	3.03
	Promedio	2.76%	2.89%	4.18%	0	0.05	0.30
	Mínima	3.16%	3.31%	3.97%	0	0.08	0.09
Tijuana	Máxima	5.02%	3.97%	4.07%	0	0.25	0.12
	Promedio	3.56%	2.99%	4.41%	0	0.12	0.01
	Mínima	4.87%	3.31%	4.22%	0	0.76	0.64

Los resultados obtenidos se presentan en la En primera instancia, se evaluaron de forma individual los modelos individuales propuestos utilizando como variable objetivo la temperatura, considerada en sus tres formas: máxima, media y mínima.
Tabla 4.12, donde se muestran los errores promedio de cada modelo evaluado sobre el conjunto de validación 2, el cual fue reservado exclusivamente para reportar resultados. El horizonte de predicción utilizado fue de 15 días.

En términos generales, los modelos SVR y RFR obtuvieron mejores resultados en promedio que el modelo LSTM. Es importante aclarar que el modelo SVR se comporta de forma determinista, es decir, que siempre genera los mismos resultados al entrenarse con los mismos datos y parámetros, por lo que su desviación estándar es cero. Por otro lado, los modelos RFR y LSTM son estocásticos y su rendimiento puede variar ligeramente entre ejecuciones. En ese sentido, el modelo LSTM mostró menor variabilidad que el RFR, lo cual indica una mayor estabilidad. No obstante, se observan dos excepciones importantes en la tabla: en Guadalajara (Máxima, RFR: $\sigma = 2.00$) y en Tampico (Máxima, LSTM: $\sigma = 3.03$),

donde se presentan desviaciones elevadas. Esto puede estar relacionado con la complejidad de los datos en esas ciudades o con la sensibilidad del modelo a ciertos patrones.

Tabla 4.13. Resultados del ensamble ponderado por PSO

Ciudades	Temperatura	Ensamble
Guadalajara	Máxima	1.61%
	Promedio	1.60%
	Mínima	1.58%
Monterrey	Máxima	2.59%
	Promedio	3.04%
	Mínima	2.52%
Tampico	Máxima	2.62%
	Promedio	2.39%
	Mínima	2.53%
Tijuana	Máxima	3.28%
	Promedio	2.87%
	Mínima	3.10%

Una vez evaluados los modelos por separado, se construyó un modelo ensamble utilizando un algoritmo de enjambre de partículas (PSO), el cual se entrenó con los conjuntos de entrenamiento y validación 1. Los resultados se presentan en la Tabla 4.13, la cual muestra el desempeño del ensamble sobre el conjunto de validación 2, que no fue utilizado en ningún momento durante el entrenamiento.

El modelo ensamblado logró un desempeño igual o mejor que el mejor modelo individual en todos los casos. Esto se debe a que el algoritmo PSO explora diversas combinaciones de pesos, aprovechando lo mejor de cada modelo base. Una vez encontrada la mejor combinación en los conjuntos de entrenamiento y validación 1, se mantuvo fija para su evaluación, garantizando así una evaluación objetiva y sin ajustes adicionales.

Tabla 4.14. Comparativa del mejor método individual frente al método de ensamble

Ciudad	Temperatura	Mejor modelo	MAPE mejor modelo	MAPE del ensamble	Mejora
Guadalajara	Máxima	RFR	1.61%	1.61%	0%
	Promedio	RFR	1.64%	1.60%	2.43%
	Mínima	RFR	1.58%	1.58%	0%
Monterrey	Máxima	RFR	2.86%	2.59%	9.44%
	Promedio	SVR	4.18%	3.04%	27.27%
	Mínima	RFR	2.66%	2.52%	5.26%
Tampico	Máxima	RFR	3.04%	2.62%	13.81%
	Promedio	SVR	2.76%	2.39%	13.26%
	Mínima	SVR	3.16%	2.53%	19.93%
Tijuana	Máxima	RFR	3.97%	3.28%	17.38%
	Promedio	RFR	2.99%	2.87%	4.01%
	Mínima	RFR	3.31%	3.10%	6.34%

En la Tabla 4.14 se comparan los resultados del modelo de ensamble frente al mejor modelo individual en cada caso, mostrando también el porcentaje de mejora obtenida. En promedio, el modelo ensamblado mejoró el desempeño en un 9.13% respecto al mejor modelo individual, confirmando su capacidad de generalización y su estabilidad frente a datos variables y multivariantes.

Tabla 4.15. Evaluación del ensamble en el conjunto de prueba

Ciudad	Temperatura	Ensamble
Guadalajara	Máxima	2.04%
	Promedio	3.92%
	Mínima	1.93%
Monterrey	Máxima	5.57%
	Promedio	4.32%
	Mínima	3.37%
Tampico	Máxima	2.44%
	Promedio	4.55%
	Mínima	4.12%
Tijuana	Máxima	4.24%
	Promedio	3.47%
	Mínima	5.31%

Los resultados de la evaluación final frente al conjunto de prueba se muestran en la Tabla 4.15, los cuales reflejan la capacidad del modelo para adaptarse a nuevos patrones o cambios bruscos en los datos, además de representar la capacidad del modelo para evitar sobreajuste u otros problemas típicos en el machine learning. Por otra parte, aunque el objetivo principal fue evaluar el desempeño en cuanto al error de predicción, también se registraron los tiempos de entrenamiento de los modelos. El modelo SVR tardó en promedio 3 segundos por serie, el modelo RFR 6.2 minutos, y el LSTM cerca de 8.4 minutos. La optimización del ensamble tomó aproximadamente 1.2 minutos por configuración.

La metodología propuesta destaca por su capacidad para combinar diferentes modelos, mejorando el rendimiento general, la robustez y la adaptabilidad ante datos climáticos variados.

4.4 Resultados obtenidos de la metodología GEANY

A continuación, se presentan algunos resultados obtenidos a partir del desarrollo y la experimentación de la metodología GEANY. Al igual que en los procesos experimentales anteriores, y con el fin de garantizar la confiabilidad de los resultados y la reproducibilidad de los experimentos, cada serie de tiempo se ejecutó al menos en 30 ocasiones. Los resultados que se muestran corresponden a los valores promedio obtenidos en dichas ejecuciones.

4.4.1 Conjunto de datos

Para evaluar la metodología, se utilizaron cuatro series de tiempo multivariables correspondientes a las ciudades de Monterrey, Guadalajara, Tijuana y Tampico. Estas series son exactamente las mismas utilizadas en la metodología TAE-Predict, por lo que de la misma forma están descritas en la sección 3.3 incluyen variables meteorológicas relevantes como temperatura, punto de rocío, humedad, velocidad del viento y presión atmosférica, registradas diariamente entre 2012 y 2022. Para esta estrategia, cada serie se dividió en subconjuntos de entrenamiento (55%), 2 conjuntos de validación iguales, sumando un total del (30%) y un conjunto de prueba (15%)

4.4.2 Resultados obtenidos

Esta metodología tuvo como objetivo identificar las variables relevantes dentro de un modelo multivariable o de varios indicadores climáticos en este caso. En este trabajo, el proceso experimental se limitó a la etapa de selección de variables y no al pronóstico, evaluando la importancia de la participación de cada variable con respecto a la variable objetivo.

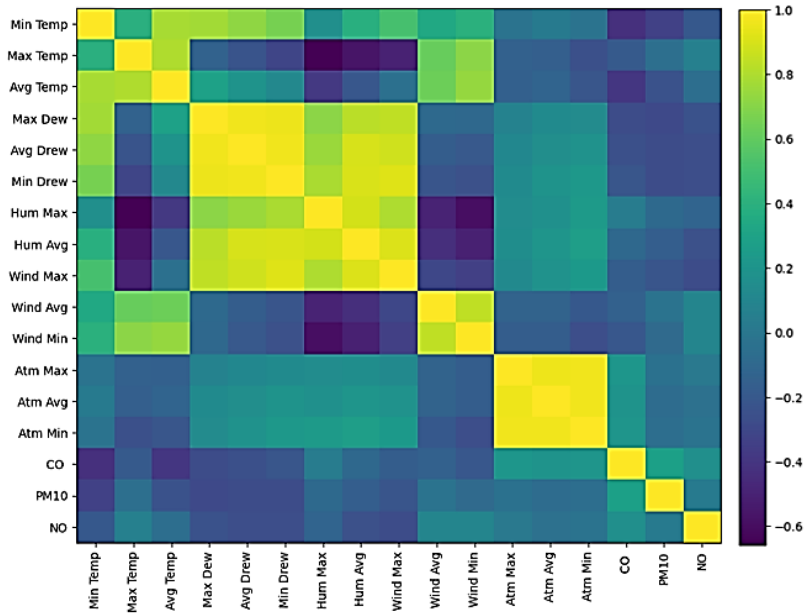


Figura 4.4. Comparativa de correlación de las variables originales

En este contexto, la primera fase realizada fue un análisis de correlación entre las variables, con el fin de identificar de forma visual aquellas cuyo comportamiento resulta similar positiva o negativamente. Si bien este análisis no implica un cambio directo en el sistema ni se involucra en el proceso de selección de variables relevantes, resulta importante para identificar posibles variables repetidas o con información redundante. Este proceso puede visualizarse en la Figura 4.4, donde se observa que algunas variables están altamente correlacionadas entre sí, como los valores de temperatura media, máxima y mínima, presentan una alta correlación debido a su comportamiento similar en el tiempo. Por otra parte, el resto de las series de tiempo muestran comportamientos poco correlacionados, lo que puede sugerir una alta complejidad del problema abordado.

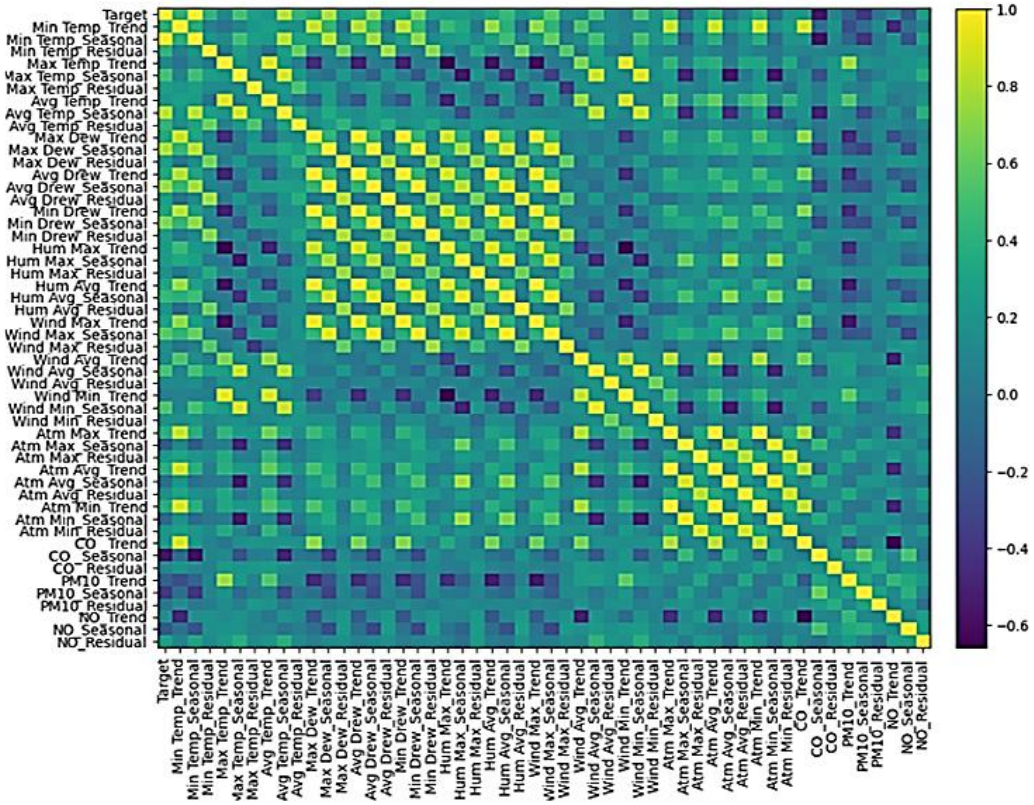


Figura 4.5. Comparativa de correlación después del proceso de descomposición

Posteriormente, se aplicó un proceso de descomposición, mediante el cual se extrajeron los componentes de tendencia, estacionalidad y ruido de los datos. En el gráfico presentado en la Figura 4.5 se incluye la variable objetivo; esta inclusión es necesaria debido a que el proceso de descomposición se aplica directamente sobre la variable a pronosticar, por lo que debe considerarse de manera explícita para la posteridad. De acuerdo con la figura, puede observarse que los componentes estacionales tienden a estar correlacionados con sus variables relacionadas. Esta descomposición permite aislar otros comportamientos que pueden aportar información relevante al modelo y cuya contribución podría verse atenuada o incluso eliminada por la preponderancia del componente estacional.

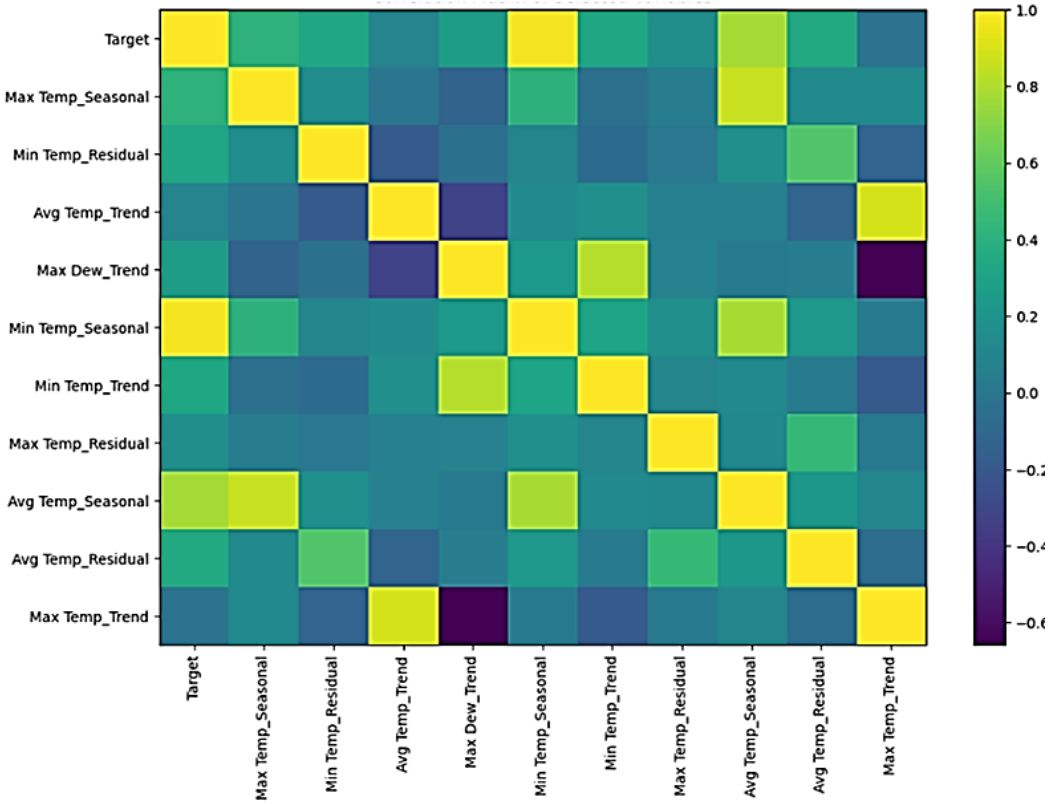


Figura 4.6. Correlación del conjunto de componentes reducidos

Una vez aplicado el modelo de ensamble para la selección de variables relevantes, se obtiene un conjunto reducido de variables. En la Figura 4.6 se muestra un ejemplo del resultado de este proceso. En este caso, y de acuerdo con la configuración del método, se seleccionan únicamente aquellas variables cuya contribución acumulada alcanza al menos el 95 % de la varianza explicada con respecto a la variable objetivo, garantizando así una buena configuración de desempeño entre reducción de dimensionalidad y preservación de la información relevante.

Capítulo 5

5 Conclusiones y Recomendaciones

A partir del desarrollo experimental y del análisis de los resultados obtenidos mediante las distintas metodologías propuestas en este trabajo de investigación, se presentan a continuación las siguientes conclusiones.

5.1 Conclusiones generales

En lo general, el trabajo realizado propone utilizar estrategias de pronóstico orientadas a ayudar a encontrar un pronóstico eficaz en el contexto de las variables involucradas en el cambio climático. En este trabajo se abordaron los dos esquemas de pronóstico, tanto monovariables como multivariables. De la misma forma se experimentó con diferentes estrategias estadísticas y modelos de machine learning, eligiendo aquellos que tenían un mejor comportamiento orientado al problema abordado. Abordando también técnicas novedosas como las estrategias de ensamble y de optimización inteligente.

El análisis de los resultados permite concluir que el uso de estrategias de ensamble, optimizadas mediante algoritmos evolutivos, constituye una herramienta eficaz para el pronóstico de variables o indicadores en el contexto de cambio climático. Se validó que la combinación ponderada de diversos modelos de pronóstico logra capturar de mejor forma los patrones o comportamientos en las series de tiempo, obteniendo resultados que igualan o superan el desempeño del mejor modelo individual en todos los escenarios evaluados.

Un aporte relevante de este trabajo es la competitividad alcanzada frente al estado del arte, particularmente en comparación con las metodologías HELI y FORSEER, las cuales mostraron una calidad de resultados estadísticamente equivalente al modelo SMYL, el cual es la metodología ganadora del certamen de pronóstico M4, cabe destacar que a pesar de tener la misma calidad de resultados, la metodología FORSEER se comprobó que es mas estable frente a series complejas, obteniendo en todos los casos un tiempo de ejecución menor. Así, y en lo general, las metodologías propuestas ofrecen una ventaja sustancial en cuanto a eficiencia computacional y estabilidad en los tiempos de entrenamiento.

Por otra parte, la integración de variables climáticas, atmosféricas o humanas mediante enfoques multivariados, como los utilizados en la metodología TAE-Predict, permitió fortalecer la capacidad de generalización de los modelos, integrando además estrategias de selección de variables relevantes. Asimismo, se comprobó que la reducción de la dimensionalidad aplicada en la metodología GEANY optimiza el uso de recursos y disminuye la complejidad del modelo, al emplear un ensamble basado en unión para la selección de componentes relevantes, sin comprometer la precisión del pronóstico final.

Por último, se comprobó que el uso de técnicas de descomposición para aislar componentes de tendencia, estacionalidad y residuales facilitó el ajuste de los algoritmos de aprendizaje automático, esto al especializarse únicamente en un comportamiento determinado en el que si bien, crece el número de modelos también estos son mas ligeros al tratar de generalizar comportamientos mas simples.

5.2 Recomendaciones y trabajos futuros

Se recomienda la exploración de arquitecturas de aprendizaje profundo emergentes y novedosos, tales como los modelos basados en Transformers, integrándolos en las secciones es modulares desarrolladas para evaluar posibles mejoras de desempeño. Se sugiere profundizar en la implementación de nuevas técnicas de cómputo acelerado en GPU. Por último, se considera conveniente ampliar los conjuntos de datos multivariados incorporando indicadores de humanos, de flora y fauna, lo cual podría favorecer la capacidad de los modelos en el contexto del pronóstico del cambio climático.

Bibliografía

[1] M. González Elizondo, E. Jurado Ybarra, S. González Elizondo, Ó. A. Aguirre Calderón, J. Jiménez Pérez, y J. de J. Návar Cháidez, «Cambio climático mundial : origen y consecuencias», *Cienc. UANL*, vol. 6, n.º 3, Art. n.º 3, 2003, Accedido: 18 de abril de 2022. [En línea]. Disponible en: http://eprints.uanl.mx/1287/

[2] J. Lelieveld, K. Klingmüller, A. Pozzer, R. T. Burnett, A. Haines, y V. Ramanathan, «Effects of fossil fuel and total anthropogenic emission removal on public health and climate», *Proc. Natl. Acad. Sci.*, vol. 116, n.º 15, pp. 7192-7197, abr. 2019, doi: 10.1073/pnas.1819989116.

[3] T. Gneiting, «Correlation functions for atmospheric data analysis», *Q. J. R. Meteorol. Soc.*, vol. 125, n.º 559, pp. 2449-2464, 1999, doi: 10.1002/qj.49712555906.

[4] T. Audesirk, G. Audesirk, y B. E. Byers, *Biología: la vida en la tierra*. Pearson Educación, 2003.

[5] M. L. DeMarche, D. F. Doak, y W. F. Morris, «Incorporating local adaptation into forecasts of species' distribution and abundance under climate change», *Glob. Change Biol.*, vol. 25, n.º 3, pp. 775-793, 2019, doi: 10.1111/gcb.14562.

[6] S. L. Brusatte *et al.*, «The extinction of the dinosaurs», *Biol. Rev.*, vol. 90, n.º 2, pp. 628-642, 2015, doi: 10.1111/brv.12128.

[7] A. J. Suggitt *et al.*, «Extinction risk from climate change is reduced by microclimatic buffering», *Nat. Clim. Change*, vol. 8, n.º 8, Art. n.º 8, ago. 2018, doi: 10.1038/s41558-018-0231-9.

[8] J. Kanipe, «Climate change: A cosmic connection», *Nature*, vol. 443, n.º 7108, pp. 141-144, sep. 2006.

[9] A. P. Williams *et al.*, «Observed Impacts of Anthropogenic Climate Change on Wildfire in California», *Earths Future*, vol. 7, n.º 8, pp. 892-910, 2019, doi: 10.1029/2019EF001210.

[10] H. Le Treut *et al.*, «Historical Overview of Climate Change Science. Chapter 1», sep. 2007, Accedido: 18 de abril de 2022. [En línea]. Disponible en: https://www.osti.gov/etdeweb/biblio/20962163

[11] J. F. B. Mitchell, «The "Greenhouse" effect and climate change», *Rev. Geophys.*, vol. 27, n.º 1, pp. 115-139, 1989, doi: 10.1029/RG027i001p00115.

[12] «UNFCCC». Accedido: 21 de octubre de 2022. [En línea]. Disponible en: https://unfccc.int/

[13] C. Lyon, «Complexity ethics and UNFCCC practices for 1.5°C climate change», *Curr. Opin. Environ. Sustain.*, vol. 31, pp. 48-55, abr. 2018, doi: 10.1016/j.cosust.2017.12.008.

[14] S. A. Montzka, E. J. Dlugokencky, y J. H. Butler, «Non-CO2 greenhouse gases and climate change», *Nature*, vol. 476, n.º 7358, pp. 43-50, ago. 2011, doi: 10.1038/nature10322.

[15] S. Legg, «IPCC, 2021: Climate Change 2021 - the Physical Science basis», *Interaction*, vol. 49, n.º 4, pp. 44-45, dic. 2021, doi: 10.3316/informit.315096509383738.

[16] R. Lal, «Beyond COP 21: Potential and challenges of the "4 per Thousand" initiative», *J. Soil Water Conserv.*, vol. 71, n.º 1, pp. 20A-25A, ene. 2016, doi: 10.2489/jswc.71.1.20A.

[17] J.-C. Ciscar *et al.*, «Physical and economic consequences of climate change in Europe», *Proc. Natl. Acad. Sci.*, vol. 108, n.º 7, pp. 2678-2683, feb. 2011, doi: 10.1073/pnas.1011612108.
[18] J. McCarthy, M. L. Minsky, N. Rochester, y C. E. Shannon, «A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955», *AI Mag.*, vol. 27, n.º 4, pp. 12-12, 2006.
[19] J. E. Olesen y M. Bindi, «Consequences of climate change for European agricultural productivity, land use and policy», *Eur. J. Agron.*, vol. 16, n.º 4, pp. 239-262, jun. 2002, doi: 10.1016/S1161-0301(02)00004-7.
[20] A. Robertson y F. Vitart, *Sub-seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*. Elsevier, 2018.
[21] C. D. Peters-Lidard *et al.*, «Indicators of climate change impacts on the water cycle and water management», *Clim. Change*, vol. 165, n.º 1, p. 36, mar. 2021, doi: 10.1007/s10584-021-03057-5.
[22] N. Nicholls, «Cognitive Illusions, Heuristics, and Climate Prediction», *Bull. Am. Meteorol. Soc.*, vol. 80, n.º 7, pp. 1385-1398, jul. 1999, doi: 10.1175/1520-0477(1999)080<1385:CIHACP>2.0.CO;2.
[23] M. Hulme *et al.*, «Climate change scenarios for global impacts studies», *Glob. Environ. Change*, vol. 9, pp. S3-S19, oct. 1999, doi: 10.1016/S0959-3780(99)00015-1.
[24] H. Ij, «Statistics versus machine learning», *Nat Methods*, vol. 15, n.º 4, p. 233, 2018.
[25] C. Chatfield, «The Holt-Winters Forecasting Procedure», *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 27, n.º 3, pp. 264-279, 1978, doi: 10.2307/2347162.
[26] P. R. Winters, «Forecasting Sales by Exponentially Weighted Moving Averages», *Manag. Sci.*, vol. 6, n.º 3, pp. 324-342, abr. 1960, doi: 10.1287/mnsc.6.3.324.
[27] S. Makridakis, S. C. Wheelwright, y R. J. Hyndman, *Forecasting methods and applications*. John wiley & sons, 2008.
[28] D. J. Hand, «Principles of Data Mining», *Drug Saf.*, vol. 30, n.º 7, pp. 621-622, jul. 2007, doi: 10.2165/00002018-200730070-00010.
[29] V. Mirjalili y S. Raschka, *Python Machine Learning*. Marcombo, 2020.
[30] M. J. Zaki y C.-T. Ho, *Large-Scale Parallel Data Mining*. Springer Science & Business Media, 2000.
[31] R. Bekkerman, M. Bilenko, y J. Langford, *Scaling Up Machine Learning: Parallel and Distributed Approaches*. Cambridge University Press, 2012.
[32] A. K. Sangaiah, *Deep Learning and Parallel Computing Environment for Bioengineering Systems*. Academic Press, 2019.
[33] D. Alexander, *Neural Networks: History and Applications*. Nova Science Publishers, Incorporated, 2020.
[34] D. Luebke, «CUDA: Scalable parallel programming for high-performance scientific computing», en *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, may 2008, pp. 836-838. doi: 10.1109/ISBI.2008.4541126.
[35] R. R. Schaller, «Moore's law: past, present and future», *IEEE Spectr.*, vol. 34, n.º 6, pp. 52-59, jun. 1997, doi: 10.1109/6.591665.
[36] R. Ansorge, *Programming in Parallel with CUDA*. Cambridge University Press, 2022.
[37] R. J. Hyndman y G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

[38] «A Comparative Analysis of the Classical and Machine learning Forecasting Methods for the Mexican Stock Exchange. | EBSCOhost». Accedido: 27 de enero de 2026. [En línea]. Disponible en: https://openurl.ebsco.com/EPDB%3Agcd%3A8%3A35685816/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A182289545&crl=c&link_origin=scholar.google.com

[39] J. Purata-Aldaz, J. Frausto-Solís, G. Castilla-Valdez, J. González-Barbosa, y J. P. Sánchez Hernández, «MASIP: A Methodology for Assets Selection in Investment Portfolios», *Math. Comput. Appl.*, vol. 30, n.º 2, p. 34, 2025.

[40] E. Estrada-Patiño, G. Castilla-Valdez, J. Frausto-Solis, J. González-Barbosa, y J. P. Sánchez-Hernández, «A Novel Approach for Temperature Forecasting in Climate Change Using Ensemble Decomposition of Time Series», *Int. J. Comput. Intell. Syst.*, vol. 17, n.º 1, p. 253, sep. 2024, doi: 10.1007/s44196-024-00667-6.

[41] R. B. Cleveland, W. S. Cleveland, J. E. McRae, y I. Terpenning, «STL: A seasonal-trend decomposition», *J Stat*, vol. 6, n.º 1, pp. 3-73, 1990.

[42] M. WEST, «Time series decomposition», *Biometrika*, vol. 84, n.º 2, pp. 489-494, jun. 1997, doi: 10.1093/biomet/84.2.489.

[43] A. Dokumentov y R. J. Hyndman, «STR: Seasonal-trend decomposition using regression», *ArXiv Prepr. ArXiv200905894*, 2020.

[44] R. J. Hyndman y G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2021.

[45] Q. Wen, J. Gao, X. Song, L. Sun, H. Xu, y S. Zhu, «RobustSTL: A robust seasonal-trend decomposition algorithm for long time series», en *Proceedings of the AAAI conference on artificial intelligence*, 2019, pp. 5409-5416. Accedido: 5 de noviembre de 2024. [En línea]. Disponible en: https://aaai.org/ojs/index.php/AAAI/article/view/4480

[46] S. A. Alasadi y W. S. Bhaya, «Review of data preprocessing techniques in data mining», *J. Eng. Appl. Sci.*, vol. 12, n.º 16, pp. 4102-4107, 2017.

[47] F. Kamiran y T. Calders, «Data preprocessing techniques for classification without discrimination», *Knowl. Inf. Syst.*, vol. 33, n.º 1, pp. 1-33, oct. 2012, doi: 10.1007/s10115-011-0463-8.

[48] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, y O. Tabona, «A survey on missing data in machine learning», *J. Big Data*, vol. 8, n.º 1, p. 140, oct. 2021, doi: 10.1186/s40537-021-00516-9.

[49] L. O. Joel, W. Doorsamy, y B. S. Paul, «A Review of Missing Data Handling Techniques for Machine Learning», *Int. J. Innov. Technol. Interdiscip. Sci.*, vol. 5, n.º 3, Art. n.º 3, sep. 2022, doi: 10.15157/IJITIS.2022.5.3.971-1005.

[50] K. Seu, M.-S. Kang, y H. Lee, «An intelligent missing data imputation techniques: A review», *JOIV Int. J. Inform. Vis.*, vol. 6, n.º 1-2, pp. 278-283, 2022.

[51] E. Steiner, *Matemáticas para las ciencias aplicadas*. Reverte, 2005.

[52] A. Ahuja, L. Al-Zogbi, y A. Krieger, «Application of noise-reduction techniques to machine learning algorithms for breast cancer tumor identification», *Comput. Biol. Med.*, vol. 135, p. 104576, ago. 2021, doi: 10.1016/j.compbiomed.2021.104576.

[53] S. Gupta y A. Gupta, «Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review», *Procedia Comput. Sci.*, vol. 161, pp. 466-474, ene. 2019, doi: 10.1016/j.procs.2019.11.146.

[54] A. Kamran, L. Guozhu, A. F. Rafique, y Q. Zeeshan, «\pm3-Sigma based design optimization of 3D Finocyl grain», *Aerosp. Sci. Technol.*, vol. 26, n.º 1, pp. 29-37, 2013.
[55] S. Seo, «A review and comparison of methods for detecting outliers in univariate data sets», PhD Thesis, University of Pittsburgh, 2006.
[56] P. K. Sadasivan y D. N. Dutt, «SVD based technique for noise reduction in electroencephalographic signals», *Signal Process.*, vol. 55, n.º 2, pp. 179-189, dic. 1996, doi: 10.1016/S0165-1684(96)00129-6.
[57] K.-C. Lee, J.-S. Ou, y M.-C. Fang, «Application of SVD noise-reduction technique to PCA based radar target recognition», *Prog. Electromagn. Res.*, vol. 81, pp. 447-459, 2008.
[58] H. Wang, X. Wang, Y. Cheng, y X. Tuo, «Research on noise reduction method of RDTS using D-SVD», *Opt. Fiber Technol.*, vol. 48, pp. 151-158, 2019.
[59] C. H. Sebastian, B. S. Agustín, y G. G. Ismael, *Manual de Álgebra lineal*. Universidad del Norte, 2018.
[60] J. Wodecki, A. Michalak, y P. Stefaniak, «Review of smoothing methods for enhancement of noisy data from heavy-duty LHD mining machines», *E3S Web Conf.*, vol. 29, p. 00011, 2018, doi: 10.1051/e3sconf/20182900011.
[61] S. G. K. Patro y K. K. Sahu, «Normalization: A Preprocessing Stage», 19 de marzo de 2015, *arXiv*: arXiv:1503.06462. Accedido: 5 de noviembre de 2024. [En línea]. Disponible en: http://arxiv.org/abs/1503.06462
[62] P. J. M. Ali, R. H. Faraj, E. Koya, P. J. M. Ali, y R. H. Faraj, «Data normalization and standardization: a technical report», *Mach Learn Tech Rep*, vol. 1, n.º 1, pp. 1-6, 2014.
[63] A. Malhi y R. X. Gao, «PCA-based feature selection scheme for machine defect classification», *IEEE Trans. Instrum. Meas.*, vol. 53, n.º 6, pp. 1517-1525, 2004.
[64] L. Breiman, «Random forests», *Mach. Learn.*, vol. 45, n.º 1, pp. 5-32, 2001.
[65] S. Makridakis y M. Hibon, «ARMA Models and the Box-Jenkins Methodology», *J. Forecast.*, vol. 16, n.º 3, pp. 147-163, may 1997, doi: 10.1002/(SICI)1099-131X(199705)16:3<147::AID-FOR652>3.0.CO;2-X.
[66] B. María-Esther García-Morales, M. Lucila Morales-Rodríguez, A. Morales-Rodríguez, y R. Salas-Cabrera, «Towards a methodology to integrate knowing-being with emotions, speech acts and natural language processing in conversational agents.», *Int. J. Comb. Optim. Probl. Inform.*, vol. 16, n.º 3, 2025, Accedido: 27 de enero de 2026. [En línea]. Disponible en: https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=20071558&AN=188045371&h=P0dygaL6gFGgXOYIjbzDt0ZwBAyG%2BbLqGp7uVz7AIi%2Fdyz1eDdyg99G2ZjJ1cvgsv2OX%2BfhHWMdn4omBIH7hIQ%3D%3D&crl=c
[67] H. Esmaily, M. Tayefi, H. Doosti, M. Ghayour-Mobarhan, H. Nezami, y A. Amirabadizadeh, «A comparison between decision tree and random forest in determining the risk factors associated with type 2 diabetes», *J. Res. Health Sci.*, vol. 18, n.º 2, p. 412, 2018.
[68] J. Ali, R. Khan, N. Ahmad, y I. Maqsood, «Random Forests and Decision Trees», 2012.
[69] T. R. Prajwala, «A comparative study on decision tree and random forest using R tool», *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, n.º 1, pp. 196-199, 2015.
[70] E. Estrada-Patiño, G. Castilla-Valdez, J. Frausto-Solis, y J. D. Terán Villanueva, «Enfoque de aprendizaje híbrido evolutivo para redes neuronales en la clasificación de

casos médicos», *Program. Matemática Softw.*, vol. 9, n.º 3, Art. n.º 3, dic. 2017, doi: 10.30973/progmat/2017.9.3/8.
[71] S. Hochreiter y J. Schmidhuber, «Long Short-Term Memory», *Neural Comput.*, vol. 9, n.º 8, pp. 1735-1780, nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[72] Y. Yu, X. Si, C. Hu, y J. Zhang, «A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures», *Neural Comput.*, vol. 31, n.º 7, pp. 1235-1270, jul. 2019, doi: 10.1162/neco_a_01199.
[73] C. Dumitru y V. Maria, «Advantages and Disadvantages of Using Neural Networks for Predictions.», *Ovidius Univ. Ann. Ser. Econ. Sci.*, vol. 13, n.º 1, 2013.
[74] J. Bobadilla, *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. Ediciones de la U, 2021.
[75] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, y Y. Zhang, «Short-term residential load forecasting based on LSTM recurrent neural network», *IEEE Trans. Smart Grid*, vol. 10, n.º 1, pp. 841-851, 2017.
[76] A. Glassner, *Deep Learning: A Visual Approach*. No Starch Press, 2021.
[77] P. K. Nalajam y R. Varadarajan, «A Hybrid Deep Learning Model for Layer-Wise Melt Pool Temperature Forecasting in Wire-Arc Additive Manufacturing Process», *IEEE Access*, vol. 9, pp. 100652-100664, 2021, doi: 10.1109/ACCESS.2021.3097177.
[78] R. M. Balabin y E. I. Lomakina, «Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data», *Analyst*, vol. 136, n.º 8, pp. 1703-1712, mar. 2011, doi: 10.1039/C0AN00387E.
[79] A. Chakrabarti y J. K. Ghosh, «AIC, BIC and recent advances in model selection», *Philos. Stat.*, pp. 583-605, 2011.
[80] J. Frausto-Solis, L. Rodriguez-Moya, J. González-Barbosa, G. Castilla-Valdez, y M. Ponce-Flores, «FCTA: A Forecasting Combined Methodology with a Threshold Accepting Approach», *Math. Probl. Eng.*, vol. 2022, p. e6206037, mar. 2022, doi: 10.1155/2022/6206037.
[81] E. Estrada-Patiño, G. Castilla-Valdez, J. Frausto-Solis, J. J. Gonzalez-Barbosa, y J. P. Sánchez-Hernández, «HELI: An Ensemble Forecasting Approach for Temperature Prediction in the Context of Climate Change», *Comput. Sist.*, vol. 28, n.º 3, Art. n.º 3, sep. 2024, doi: 10.13053/cys-28-3-5175.
[82] C. D. Sutton, «Classification and regression trees, bagging, and boosting», *Handb. Stat.*, vol. 24, pp. 303-329, 2005.
[83] J. H. Holland, «Genetic Algorithms», *Sci. Am.*, vol. 267, n.º 1, pp. 66-73, 1992.
[84] «The Generalized PSO: A New Door to PSO Evolution - Fernández Martínez - 2008 - Journal of Artificial Evolution and Applications - Wiley Online Library». Accedido: 27 de enero de 2026. [En línea]. Disponible en: https://onlinelibrary.wiley.com/doi/full/10.1155/2008/861275
[85] «Servicio Meteorológico Nacional». Accedido: 5 de octubre de 2022. [En línea]. Disponible en: https://smn.conagua.gob.mx/es/
[86] I. N. de E. y Geografía (INEGI), «Mapas. Climatológicos». Accedido: 5 de octubre de 2022. [En línea]. Disponible en: https://www.inegi.org.mx/temas/climatologia/
[87] C. D. S. Ecosystem, «Ecosistema de espacio de datos de Copernicus | Los ojos de Europa en la Tierra». Accedido: 27 de enero de 2026. [En línea]. Disponible en: https://dataspace.copernicus.eu/

[88] L. Lagunes Cuevas, «VARIABLES DEL CAMBIO CLIMATICO EN MEXICO EN BASE A ANALISIS ESTADISTICOS Y MODELOS MATEMATICOS», abr. 2024, Accedido: 27 de enero de 2026. [En línea]. Disponible en: https://rinacional.tecnm.mx/jspui/handle/TecNM/7857

[89] J. Frausto Solís, E. Estrada-Patiño, M. Ponce Flores, J. P. Sánchez-Hernández, G. Castilla-Valdez, y J. González-Barbosa, «TAE Predict: An Ensemble Methodology for Multivariate Time Series Forecasting of Climate Variables in the Context of Climate Change», *Math. Comput. Appl.*, vol. 30, n.º 3, p. 46, jun. 2025, doi: 10.3390/mca30030046.

[90] P. Aghelpour, B. Mohammadi, y S. M. Biazar, «Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA», *Theor. Appl. Climatol.*, vol. 138, n.º 3, pp. 1471-1480, nov. 2019, doi: 10.1007/s00704-019-02905-w.

[91] M. A. Zaytar y C. El Amrani, «Sequence to sequence weather forecasting with long short-term memory recurrent neural networks», *Int. J. Comput. Appl.*, vol. 143, n.º 11, pp. 7-11, 2016.