



EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE CIUDAD MADERO

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

MAESTRÍA EN CIENCIAS DE LA INGENIERÍA



"POR MI PATRIA Y POR MI BIEN"

TESIS

"ESTIMACIÓN DE LA TEMPERATURA USANDO ANÁLISIS MULTIVARIABLE"

Que para obtener el Grado de
Maestro en Ciencias de la Ingeniería

Presenta
Ing. Allan Mizael Jiménez Fernández
G15070895
CVU 1324429

Director de Tesis
Dr. Juan Javier González Barbosa
CVU 202134

Co-director de Tesis
Dr. Juan Frausto Solís

Cd. Madero, Tamaulipas

Diciembre 2025



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, **10/Diciembre/2025**

Oficio No. U11.129/25

ASUNTO: Autorización de Impresión

C. ALLAN MIZAE L JIMÉNEZ FERNÁNDEZ
No. DE CONTROL G15070895
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 la Maestría en Ciencias de la Ingeniería, el cual está integrado por los siguientes catedráticos:

PRESIDENTE :	DR.	JUAN JAVIER GÓNZALEZ BARBOSA
SECRETARIA:	DRA.	ERIKA ALARCON RUIZ
VOCAL :	DR.	JUAN FRAUSTO SOLIS
SUPLENTE:	DRA.	GUADALUPE CASTILLA VALDEZ

Se acordó autorizar la impresión de su tesis titulada:

"ESTIMACIÓN DE LA TEMPERATURA USANDO ANÁLISIS MULTIVARIABLE"

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.

ATENTAMENTE

Excelencia en Educación Tecnológica.
"Por Mi Patria y Por Mi Bien" ®


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



ccp: Archivo
RCG/NRV/jar



2025
Año de
La Mujer
Indígena

Av. 1° de Mayo y Sor Juana I. de la Cruz S/N Col. Los Mangos
C.P. 89440 Cd. Madero, Tam. Tel. (833) 3574820 ext. 3110
e-mail: depi.cdadero@tecnm.mx www.tecnm.mx
www.cdadero.tecnm.mx



AGRADECIMIENTOS

Le agradezco a Dios por permitirme realizar este proyecto y por concluir mis estudios de maestría, durante todo el trayecto Dios me ha dado fortaleza y me ha mostrado su amor aun en los momentos de dificultad.

A mi esposa Salma por todo su apoyo, comprensión y amor. A mis papás Virginia y Mario por la educación y valores que me han dado. A mi amiga Paola por su amistad y sus consejos.

A los maestros y doctores que fueron un gran apoyo durante mis estudios en especial al doctor Juan Javier González Barbosa. A mis compañeros de la maestría Roberto, Stephanie, Alma y Carlos.

Al Instituto Tecnológico de Ciudad Madero que me abrieron las puertas permitiendo el desarrollo de esta investigación.

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (Secihti) por el apoyo económico otorgado a través de la beca

Resumen

El cambio climático es un fenómeno que afecta al planeta Tierra y a los seres vivos, debido a esto, se presentan diferentes tipos de retos para enfrentar su mitigación. Uno de esos desafíos es predecir la temperatura con precisión y así contribuir a acuerdos como el de París, el cual habla sobre un tratado internacional para reducir el calentamiento global.

Para realizar dicha predicción de la temperatura se busca aplicar mediante una selección cuidadosa un conjunto de métodos de aprendizaje automático como las redes neuronales, regresión lineal múltiple o máquina de soporte vectorial, entre otros. Estos métodos se aplicarán a varias series de tiempo climatológicas de 4 ciudades de México (Tampico, Tijuana, Monterrey y Guadalajara), dichas series de tiempo tienen los datos del comportamiento de los últimos 5 años de variables independientes como la humedad, el punto de rocío, la velocidad del viento, la presión, etc., estas variables afectan a la temperatura que será tomada como la variable dependiente a pronosticar.

Como resultado obtendremos un modelo que nos permita predecir la temperatura de acuerdo con un análisis multivariable y así poder contribuir a la mitigación del cambio climático y apoyar a las diversas áreas que utilizan la temperatura día tras día, también se buscará apoyar con esta investigación a trabajos futuros.

Abstract

Climate change is a phenomenon that affects the planet Earth and living beings, due to this, there are different types of challenges to face its mitigation. One of these challenges is to predict the temperature accurately and thus contribute to agreements such as the Paris Agreement, which talks about an international treaty to reduce global warming.

To make this temperature prediction, a set of machine learning methods such as neural networks, multiple linear regression or support vector machine, among others, are applied through careful selection. These methods will be applied to several climatological time series of 4 cities in Mexico (Tampico, Tijuana, Monterrey and Guadalajara). These time series have data on the behavior of the last 5 years of independent variables such as humidity, dew point, wind speed, pressure, etc. These variables affect the temperature that will be taken as the dependent variable to be predicted.

As a result, we will obtain a model that allows us to predict the temperature according to a multivariate and thus be able to contribute to the mitigation of climate change and support the various areas that use temperature day after day. This research will also seek to support future work.

Índice general

Contenido

Resumen	IV
Abstract.....	V
Índice Tablas	2
Índice de Figuras	3
1 Introducción	4
1.1 Planteamiento del problema.....	5
1.2 Objetivo General.....	5
1.3 Objetivos Específicos	5
1.4 Justificación del estudio.....	6
2 Marco Teórico	7
3 Estado del arte.....	19
4 Metodología.....	25
5 Resultados y conclusiones	36

Índice Tablas

Tabla 1 Estado del arte	24
Tabla 2 Variables utilizadas en el grupo 1 de las series de tiempo climatológicas	26
Tabla 3 Variables utilizadas en el grupo 2 de las series de tiempo climatológicas.....	27
Tabla 4 Descripción de los valores porcentuales de cada fase en la experimentación.....	28
Tabla 5 Hiperparámetros en la red neuronal	31
Tabla 6 Resultados predicción con RLM.	36
Tabla 7 Rendimiento de precisión MAPE del modelo LSTM en las ciudades estudiadas. ...	39
Tabla 8 Rendimiento de precisión MAPE del modelo SVR en las ciudades estudiadas.	41
Tabla 9 Rendimiento (MAPE) del modelo SARIMAX en las ciudades estudiadas.	42
Tabla 10 Análisis comparativo entre las tres técnicas de pronóstico de la temperatura.	43
Tabla 11 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Guadalajara.....	44
Tabla 12 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Tampico.....	45
Tabla 13 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Tijuana.....	45
Tabla 14 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Monterrey.	46

Índice de Figuras

Figura 1. Cambio climático. [1]	7
Figura 2. Acuerdo de París.	8
Figura 3. Temperatura	9
Figura 4. Humedad. [8]	10
Figura 5. Diagrama de un modelo multivariable en la predicción de la temperatura.	11
Figura 6. La Inteligencia Artificial y el cambio climático.	12
Figura 7. Machine learning.....	13
Figura 8 Redes neuronales.....	15
Figura 9 Ejemplo gráfico comparativo usando ARIMA. [23]	16
Figura 10 Ejemplo de una predicción usando un modelo SARIMAX. [26]	17
Figura 11 . Modelo de redes neuronales para predicción de la temperatura [29].....	20
Figura 12. Correlaciones entre las variables meteorológicas vistas en un mapa de calor. [30]	20
Figura 13. Metodología por realizar en el proyecto.	25
Figura 14 Fórmula para obtener datos normalizados.	27
Figura 15. Punto de rocío y temperatura promedio	37
Figura 16. Humedad y temperatura promedio.....	37
Figura 17. Óxido nítrico y temperatura promedio.....	38
Figura 18. Desempeño de los datos reales (azul) y de los datos pronosticados (amarillo) en la ciudad con mejor desempeño (Guadalajara) usando la red neuronal LSTM.	40
Figura 19. Desempeño de los datos reales (azul) y de los datos pronosticados (verde) en la ciudad con mejor desempeño (Guadalajara) usando SVR.	41
Figura 20 Desempeño de los datos reales (rojo) y de los datos pronosticados (verde) en la ciudad con mejor desempeño (Tampico) usando SARIMAX.....	43

1 Introducción

El cambio climático es un fenómeno que afecta al planeta Tierra y a los seres humanos en muchos ámbitos y que provoca cambios de temperatura extremos, estos cambios tan significativos iniciaron desde que los seres humanos lograron industrializar los distintos procesos que hoy en día movilizan al mundo. Debido a esto, el 4 de noviembre del 2016 entro en vigor un tratado internacional sobre el cambio climático llamado el acuerdo de París donde se busca limitar el calentamiento mundial.

La predicción de la temperatura y de las variables que afectan el cambio climático es de suma importancia ya que son utilizadas en la vida del ser humano, tiene aplicación en el desarrollo de áreas como agricultura, transporte, energía, turismo, etc. Así como también permite generar programas de mitigación para enfrentar el problema del cambio climático y tener una idea clara de lo que serán las temperaturas en un futuro.

La base de datos que se analizarán para el proyecto es un conjunto de series de tiempo climatológicas de 4 ciudades de México (Tijuana, Monterrey, Guadalajara y Tampico), donde podemos encontrar el comportamiento de las variables como la temperatura, la humedad, el punto de rocío velocidad del viento, presión, monóxido de carbono (CO), partículas menores a 10 micras y óxido nítrico (NO). Los datos describen el comportamiento diario de estas variables durante los últimos 5 años en cada una de las ciudades. En México existe poco estudio sobre la predicción de la temperatura usando múltiples variables, así como también pocos datos disponibles sobre las variables que afectan a la temperatura, sin embargo, el análisis a realizar con este conjunto de datos será de gran relevancia para trabajos futuros.

En este proyecto se utilizará el análisis de múltiples variables, esto nos ayudará a saber cuáles son las variables independientes de mayor relevancia y como es su relación con la variable dependiente, es decir, la temperatura.

Para poder obtener una predicción precisa se analizarán los distintos métodos de pronóstico, técnicas de machine learning, métodos estadísticos, métodos matemáticos etc. A partir de este análisis se escogerán las mejores técnicas, aplicando parámetros de error, buscando siempre el porcentaje de error mínimo para poder contribuir con una predicción precisa de la temperatura.

1.1 Planteamiento del problema

El cambio climático es uno de los mayores desafíos globales de nuestro tiempo ya que los impactos de este fenómeno están afectando alrededor del mundo, desde eventos climáticos extremos hasta cambios en los ecosistemas. La falta de acción y comprensión del cambio climático puede llevar a consecuencias catastróficas para las generaciones presentes y futuras. [1]

La predicción de la temperatura es afectada por diversas variables es por eso que la falta de inclusión del análisis multivariable en la predicción de la temperatura puede generar predicciones menos confiables y pasar por alto datos relevantes. [2]

1.2 Objetivo General

Pronosticar la variable temperatura en cuatro ciudades de México que cuenten con suficiente información sobre variables de cambio climático

1.3 Objetivos Específicos

- Seleccionar los casos de estudio correspondiente a ciudades con suficiente información sobre variables de cambio climático.
- Evaluar métodos de pronóstico exitosos en la literatura.
- Seleccionar la técnica de machine learning para realizar pronóstico.

1.4 Justificación del estudio

El proyecto a desarrollar es fundamental para conocer cuál será la tendencia de la temperatura en 4 ciudades de México, la investigación de la predicción de la temperatura haciendo un análisis multivariable con ayuda de las técnicas de aprendizaje automático es algo que se ha desarrollado en distintas partes del mundo como en China o la India, pero en México este tipo de estudios son muy poco visto ya que se cuenta con muy pocos datos y además las investigaciones realizadas hacen referencia a un análisis univariable.

2 Marco Teórico

Cambio climático

El cambio climático es uno de los mayores desafíos que enfrenta la humanidad en el siglo XXI. Este fenómeno, caracterizado por el aumento gradual de la temperatura promedio de la Tierra, tiene múltiples causas y repercusiones que afectan tanto al medio ambiente como a las sociedades humanas. [3]



Figura 1. Cambio climático. [1]

El calentamiento global es causado principalmente por la liberación de gases de efecto invernadero (GEI) a la atmósfera como resultado de actividades humanas. La quema de combustibles fósiles, la deforestación, la agricultura intensiva y otros procesos industriales son las principales fuentes de emisiones de GEI, como el dióxido de carbono (CO_2), el metano (CH_4) y el óxido nitroso (N_2O). [4]

Los efectos del cambio climático se manifiestan en una serie de fenómenos que incluyen el aumento del nivel del mar, el derretimiento de los glaciares y los casquetes polares, cambios en los patrones de precipitación, y el incremento en la frecuencia e intensidad de eventos climáticos extremos como tormentas, inundaciones, sequías y olas de calor. Estos impactos tienen consecuencias devastadoras para los ecosistemas naturales, la biodiversidad y las comunidades humanas en todo el mundo.

Para abordar el cambio climático de manera efectiva, se requiere una combinación de

Capítulo 2. Marco teórico

estrategias de mitigación y adaptación. La mitigación implica la reducción de las emisiones de GEI a través de la transición hacia fuentes de energía renovable, la eficiencia energética, la reforestación y otras medidas destinadas a limitar el calentamiento global. Por otro lado, la adaptación implica ajustes en los sistemas naturales y humanos para hacer frente a los impactos ya inevitables del cambio climático, como la construcción de infraestructuras resilientes, la gestión sostenible de recursos naturales y la implementación de políticas de planificación urbana.

El Acuerdo de París, adoptado en diciembre de 2015 durante la 21ª Conferencia de las Partes (COP21) de la Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC) tiene como objetivo central mantener el aumento de la temperatura global muy por debajo de 2°C respecto a los niveles preindustriales, y esforzarse por limitar ese aumento a 1.5°C, reconociendo la necesidad de reducir los riesgos y los impactos del cambio climático. Además, el Acuerdo busca fortalecer la capacidad de adaptación de los países a los efectos adversos del cambio climático y fomentar la resiliencia de los ecosistemas y las comunidades. [5]



Figura 2. Acuerdo de París. [5]

Temperatura

La temperatura es una de las variables básicas del tiempo y clima. Cuando preguntamos cómo está el tiempo afuera, casi siempre decimos algo sobre la temperatura, como hace frío o hace calor. De nuestra experiencia diaria, sabemos que la temperatura varía en diferentes escalas de tiempo en un mismo lugar, en periodos estacionales, diarios, horarios, etc., y varía también en el espacio.

En meteorología, la temperatura se registra en las estaciones meteorológicas, de las que existen miles en todo el mundo. En estas estaciones se miden, por ejemplo, datos de temperatura a determinadas horas fijas, valores de temperaturas máximas y mínimas o se toman registros continuos en el tiempo, llamados termogramas.. [6]

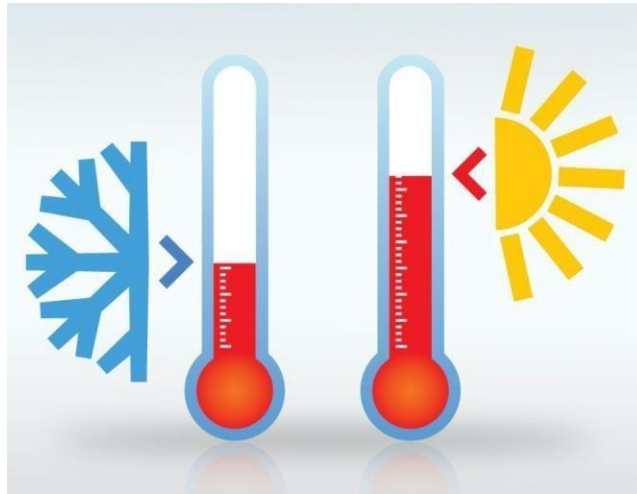


Figura 3. Temperatura. [6]

La temperatura es una variable física fundamental e indispensable en la ciencia, la ingeniería y los sistemas biológicos, ya que mide el grado de calor de un cuerpo o sistema, relacionado directamente con la energía cinética promedio de las partículas que componen la materia. [7] Su importancia radica en su profunda influencia en una amplia gama de fenómenos y procesos críticos.

La temperatura es central en la dinámica de la atmósfera y el clima, afectando la formación de nubes, precipitaciones y la biodiversidad. [7] El calentamiento global, impulsado por el aumento de gases de efecto invernadero, es un problema crítico asociado al incremento de la temperatura media del planeta.

Algunas de las variables climatológicas que afectan a la temperatura son: la humedad, el punto de rocío, velocidad del viento, entre otras. [6]

Humedad

Es un componente natural de la atmósfera y procede de la cantidad de vapor de agua existente en el aire, la humedad constituye una parte integral del ciclo hidrológico, puesto que el vapor

Capítulo 2. Marco teórico

se genera de manera continua como resultado de la evaporación y se elimina tras la condensación. Cuando la temperatura sube, el aire adquiere la capacidad de contener más vapor de agua, lo cual quiere decir que, cuanto más cálido sea el clima, mayores pueden llegar a ser los niveles de humedad. [8]

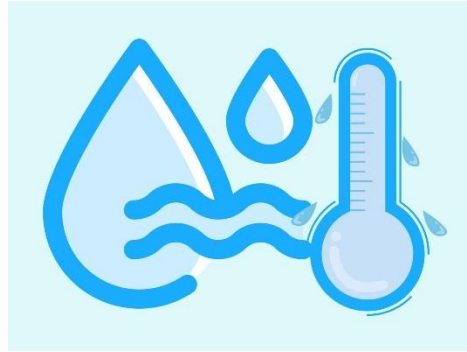


Figura 4. Humedad. [8]

Punto de rocío

Es la temperatura a la que empieza a condensar el vapor de agua contenido en el aire, produciendo rocío o niebla. Es cuando el vapor de agua presente alcanza su valor máximo o su condición de presión de saturación de vapor de agua con respecto al equilibrio con una superficie plana de agua. Aunque el punto de rocío sea expresado como una temperatura, esta correlacionado con la cantidad de vapor de agua en el aire, y por lo tanto no es dependiente en la temperatura ambiente. [8]

Velocidad del viento

El viento es el movimiento natural del aire atmosférico. En meteorología esta palabra se refiere, en general, a un movimiento de conjunto del aire cerca de la superficie terrestre o en altitud. El viento puede ser considerado como un vector con magnitud y dirección. La dirección del viento es la de su procedencia. [8]

Análisis multivariable

El análisis multivariable es una técnica que se utiliza para examinar y comprender las

Capítulo 2. Marco teórico

relaciones entre múltiples variables simultáneamente. A diferencia del análisis univariable, que se centra en una sola variable, el análisis multivariable considera la interacción entre dos o más variables. [9]

En este tipo de análisis, se exploran las relaciones entre múltiples variables independientes (predictoras) y una variable dependiente (resultado o respuesta). El objetivo principal es identificar patrones, tendencias y asociaciones complejas que no pueden ser descubiertas mediante el análisis de una sola variable. [9]

El análisis multivariable es fundamental en el pronóstico de la temperatura al modelar la compleja interdependencia entre variables climatológicas como la humedad, el punto de rocío y la velocidad del viento. Estos modelos permiten capturar la dinámica atmosférica de manera más precisa que los modelos univariados. [10]

Multivariate Climatological Forecasting Model

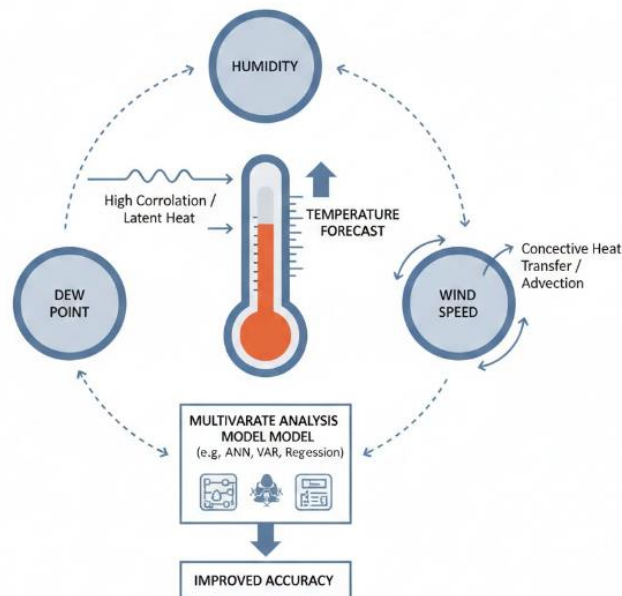


Figura 5. Diagrama de un modelo multivariable en la predicción de la temperatura.

Los modelos multivariables pueden estimar con mayor precisión la temperatura al incorporar estas variables (humedad y punto de rocío), ya que el calor latente liberado o absorbido por los cambios de fase del agua (evaporación/condensación) influye significativamente en la temperatura. [11]

Al incluir la velocidad del viento como variable predictora, los modelos multivariables

pueden capturar mejor estos efectos de transporte, mejorando el pronóstico a corto y mediano plazo. Por ejemplo, se ha documentado que los resultados de correlación son superiores al 80% al ajustar la información de velocidad del viento y temperatura ambiente en un análisis de componentes principales. [11]

Inteligencia artificial y su relación con el cambio climático

La inteligencia artificial es un campo amplio que busca crear sistemas que puedan realizar tareas que simulan la inteligencia humana (razonamiento, percepción, resolución de problemas, entre otros). [12]

Un subconjunto de la inteligencia artificial es el aprendizaje automático que se especializa en algoritmos que permite a las máquinas aprender patrones y tomar decisiones a partir de datos históricos en algún campo de estudio. Existen diversos tipos de técnicas de aprendizaje automático, pero no todas son estrictamente predictivas, para este proyecto nos enfocamos exclusivamente en las de aprendizaje supervisado.

Las técnicas de aprendizaje supervisado se entrenan con datos etiquetados (respuestas conocidas) para predecir el valor de una variable de interés. Esta área del aprendizaje automático es el corazón de la inteligencia artificial predictiva como por ejemplo el pronóstico de la temperatura. [12]

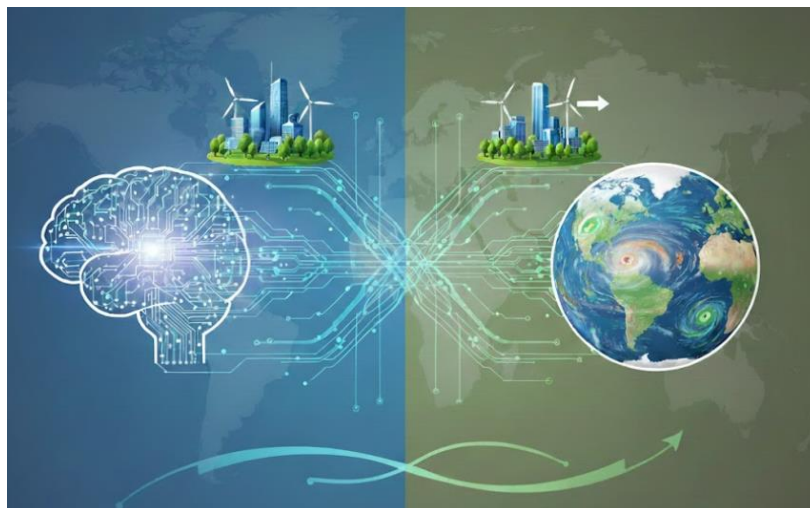


Figura 6. La Inteligencia Artificial y el cambio climático.

La Inteligencia Artificial (IA), el Aprendizaje Automático (Machine Learning - ML) y el Aprendizaje Supervisado se relacionan con el cambio climático como herramientas poderosas para la mitigación (reducir emisiones) y la adaptación (gestionar los impactos) climática, principalmente a través del análisis avanzado de datos, la predicción y la optimización de sistemas. [13]

Técnicas de aprendizaje automático para predicción

Las técnicas de aprendizaje automático son algoritmos que se entrenan con datos históricos para identificar patrones y relaciones en los datos, y luego utilizan estos patrones para hacer predicciones sobre datos futuros o desconocidos [14].

Estos modelos procesan enormes volúmenes de datos satelitales, de sensores terrestres e información histórica para mejorar la precisión de los modelos climáticos globales y regionales. Esto permite una mejor simulación de escenarios futuros de emisiones y sus impactos. [15]

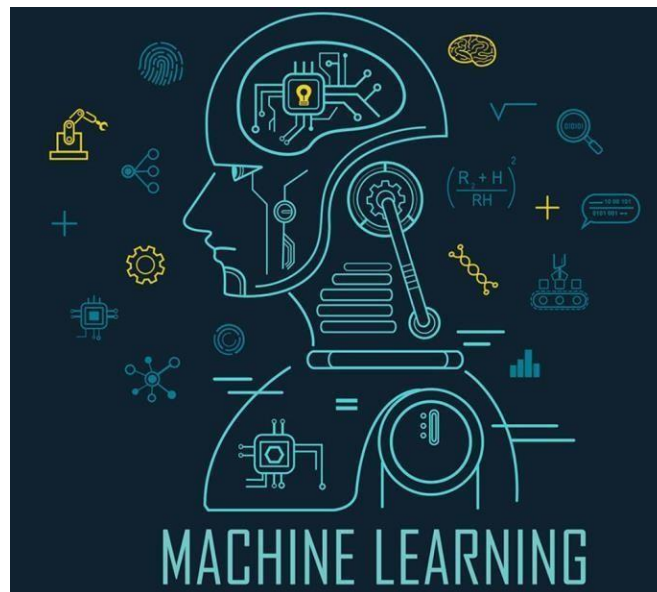


Figura 7. Machine learning [14]

Algunas de las técnicas de aprendizaje automático son las siguientes:

Regresión lineal: Este método se utiliza para predecir el valor de una variable continua en función de una o más variables independientes. Es un algoritmo estadístico que modela la relación entre una variable dependiente (y) y una o más variables independientes (x) al ajustar una línea recta a los datos observados. Su objetivo es predecir un valor de salida continuo. Por ejemplo, predecir el precio de una casa en función de su tamaño, ubicación y número de habitaciones u otro ejemplo sería el valor de la temperatura en función de la humedad, punto de rocío y velocidad del viento [14].

Árboles de decisión: Los árboles de decisión son modelos de aprendizaje supervisado que dividen repetidamente el conjunto de datos en subconjuntos más pequeños, basándose en características específicas, con el objetivo de predecir el valor de la variable objetivo. Son útiles para problemas de clasificación y regresión [16].

Máquinas de vectores de soporte (SVM): Las SVM son algoritmos de aprendizaje supervisado utilizados tanto para problemas de clasificación como de regresión. Se utilizan para encontrar el hiperplano que mejor separa las clases en el espacio de características, permitiendo hacer predicciones para nuevos datos [17].

Redes neuronales: Las redes neuronales son modelos de aprendizaje profundo que están compuestos por múltiples capas de unidades de procesamiento interconectadas, llamadas neuronas. Son muy utilizadas para problemas de predicción en una amplia gama de campos, como la visión por computadora, el procesamiento del lenguaje natural y la predicción de series temporales [17].

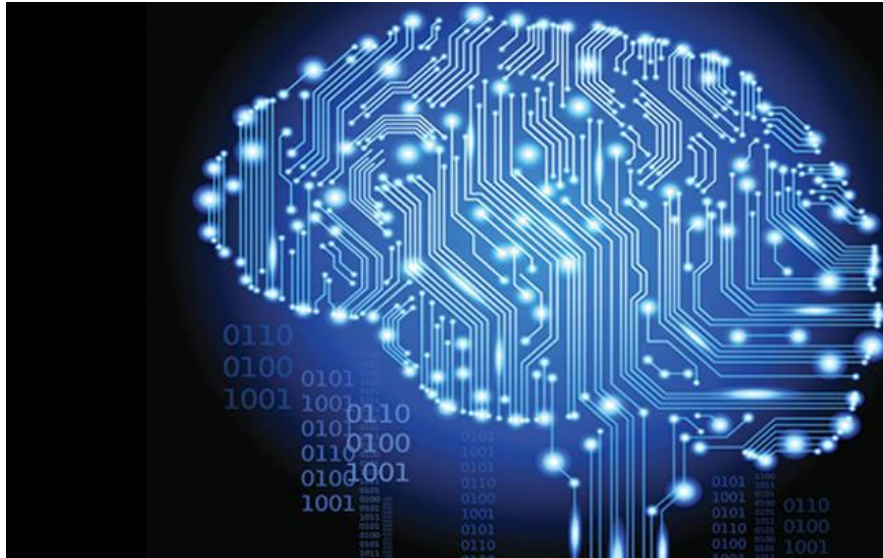


Figura 8 Redes neuronales. [18]

Existen diferentes tipos de redes neuronales, en esta investigación nos enfocaremos en las Redes Neuronales de Memoria a Largo Corto Plazo (LSTM) debido a que son una herramienta fundamental en el Aprendizaje Automático, especialmente para datos secuenciales como series temporales climáticas.

Las Redes Neuronales de Memoria a Largo Corto Plazo (Long Short-Term Memory - LSTM) son un tipo especializado de Red Neuronal Recurrente (RNN) diseñadas para modelar dependencias a largo plazo en datos secuenciales. A diferencia de las RNN estándar, que sufren del problema de gradiente evanescente (donde la información pasada se pierde rápidamente a medida que la secuencia se hace más larga), las unidades LSTM introducen una estructura de celda de memoria que les permite retener u olvidar información de forma selectiva durante períodos de tiempo prolongados [19]. Esta capacidad es crucial para tareas como el pronóstico climático, el procesamiento del lenguaje natural o el análisis de series temporales, donde el contexto de eventos ocurridos mucho antes es vital para predecir el siguiente paso.

Las redes neuronales LSTM son útiles para pronosticar variables climáticas clave, como la temperatura, la precipitación o la velocidad del viento, utilizando datos históricos multivariados y también son capaces de aprender patrones en los datos que preceden a

Capítulo 2. Marco teórico
eventos extremos (sequías, inundaciones, olas de calor), permitiendo sistemas de alerta temprana más precisos. [20]

Modelos estadísticos de series temporales

ARIMA

El modelo ARIMA (Auto Regressive Integrated Moving Average) es un enfoque estadístico diseñado para comprender y predecir series temporales no estacionarias. Una serie de tiempo es no estacionaria si sus propiedades estadísticas (como la media y la varianza) cambian con el tiempo. [22]

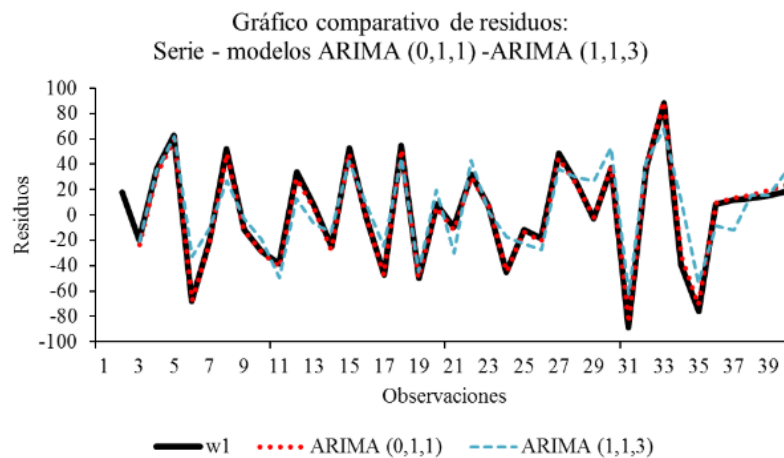


Figura 9 Ejemplo gráfico comparativo usando ARIMA. [23]

El modelo ARIMA se especifica como (p, d, q) combina tres componentes clave:

- Autorregresivo (AR - AutoRegressive): Indica que el valor actual de la serie se relaciona linealmente con sus propios valores anteriores. Se denota por el parámetro p.
- Integrado (I - Integrated): Representa el número de diferencias (diferenciación) necesarias para hacer que la serie temporal sea estacionaria. Se denota por el parámetro d.

- Medias Móviles (MA - Moving Average): Indica que el error del modelo actual es una combinación lineal de los errores de los pasos de tiempo anteriores. Se denota por el parámetro q . [22]

SARIMAX

El modelo SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous Regressors) es una extensión del modelo ARIMA que aborda dos limitaciones principales de su predecesor:

- Estacionalidad (S - Seasonal): Muchas series temporales relacionadas con el clima (como la temperatura, las precipitaciones o la concentración de CO₂) exhiben un patrón que se repite a intervalos de tiempo regulares (ej. anual, mensual). SARIMAX añade componentes estacionales (AR, I, MA) que se denotan como (P, D, Q)_s, donde s es la longitud del período estacional [24].
- Variables Exógenas (X - Exogenous): Permite incluir la influencia de otras series temporales externas o regresores exógenos en la predicción. Por ejemplo, al pronosticar la temperatura media de una región (Y), se podría incluir una variable exógena (X) como la concentración de aerosoles volcánicos, que se sabe que influye en el clima [25].

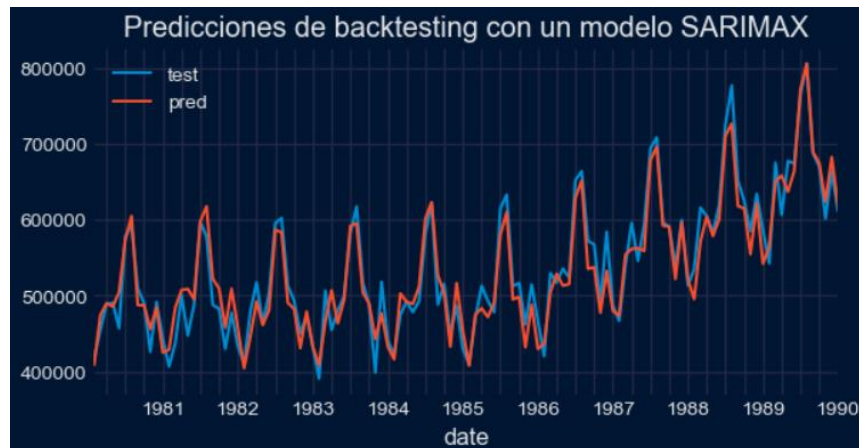


Figura 10 Ejemplo de una predicción usando un modelo SARIMAX. [26]

SARIMAX es una herramienta valiosa para el pronóstico climático, ya que los fenómenos climáticos suelen presentar tanto tendencias a largo plazo (no estacionariedad) como patrones estacionales bien definidos, y a menudo se requiere incluir datos de influencias externas (ej. índices de oscilaciones oceánicas, regulaciones de emisiones) para mejorar la precisión [27].

3 Estado del arte

A través del tiempo, el estudio de la predicción de la temperatura ha avanzado a pasos agigantados en búsqueda de obtener mejores resultados y un menor error en su pronóstico debido a la importancia que esta predicción tiene en las diferentes áreas donde se utiliza esta información diariamente, a continuación, repasaremos los estudios más destacados en forma cronológica.

Malik et al., 2014 hicieron un estudio en la India donde utilizaron redes neuronales para predicción climatológica. Utilizaron el algoritmo “Back Propagation Neural Network” y afirman que las redes neuronales artificiales son mejor herramienta que los métodos tradicionales y numéricos, además de proponer el algoritmo antes mencionado como el más importante para entrenar una red neuronal para el pronóstico del tiempo [28].

T. Anjali et al., 2019 investigaron en la universidad de Calicut en la India donde realizan una predicción de la temperatura usando tres técnicas de machine learning las cuales fueron regresión lineal múltiple, redes neuronales artificiales y máquina de soporte vectorial. En este estudio se utilizaron datos de las variables que afectan a la temperatura como la humedad, la presión, la velocidad del viento, la dirección del viento, la lluvia, el punto de rocío, entre otros. Los resultados que obtuvieron fueron analizados mediante algunas técnicas para medir el error donde se encontró que la técnica más confiable es la de regresión lineal múltiple con un promedio de error de 1.0782, después la técnica de máquina de soporte vectorial con un promedio de error de 1.1371 y por último la técnica menos confiable la de redes neuronales con un promedio de error de 1.2958 [29].

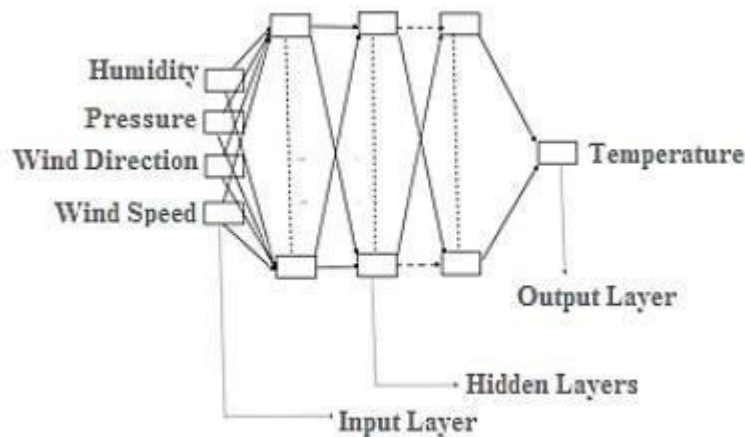


Figura 11 . Modelo de redes neuronales para predicción de la temperatura [29].

Debneil SR., 2020 trabajó en un proyecto en Stony Brook University en los Estados Unidos donde exploró la aplicación del aprendizaje profundo en la predicción de la temperatura del aire. Utilizó las siguientes técnicas: Multi-Layer Perceptron (MLP), Long Short Term Memory Network (LSTM) y una combinación de Convolutional Neural Network (CNN) y LSTM. Las variables utilizadas fueron el promedio de la velocidad del viento, la lluvia, la caída de nieve, profundidad de la nieve, el promedio de la temperatura, la temperatura máxima y la temperatura mínima [30].

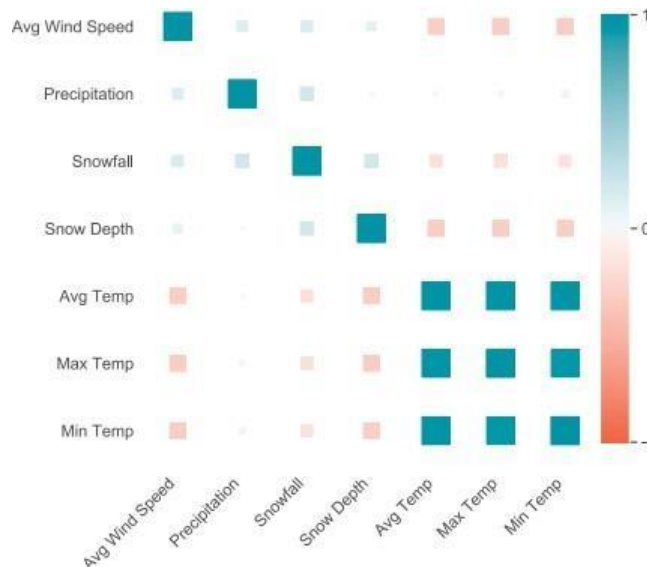


Figura 12. Correlaciones entre las variables meteorológicas vistas en un mapa de calor. [30]

I. Indan et al., 2021 analizaron en la Universitas Dipa Makassar de Indonesia cual es la mejor técnica de machine learning para la obtención del pronóstico del clima. Compararon los algoritmos de “Backpropagation” y “Bayesian Regularization”, el resultado que obtuvieron fue que el algoritmo “Bayesian Regularization” es mejor que las otras técnicas de aprendizaje automático [31].

M. Megalai et al., 2023 desarrollaron un trabajo en el SRM Institute of Science and Technology de la India para obtener un modelo de predicción de la temperatura utilizando la técnica de aprendizaje automático llamada “random forest”, para proporcionar monitoreo del clima en tiempo real y lograr cifras climáticas a través de modelos valiosos para que las personas conozcan las condiciones climáticas constantemente [32].

E. Nketiah et al., 2023 estudiaron en la Taiyuan University of Technology de China una mejora en la precisión del pronóstico de la temperatura basado en aprendizaje profundo utilizando las variables de la temperatura atmosférica, temperatura del punto de rocío, humedad relativa, presión del aire y la velocidad del viento en cinco ciudades de China [33].

Autor	Título de la Publicación	Técnicas	Variables Utilizadas	Resultados
Malik et al. (2014)	“An effective weather forecasting using neural network”	Redes neuronales (algoritmo “Back Propagation Neural Network”)	Temperatura, presión, humedad y dirección del viento.	Concluyeron que las RNA son mejor herramienta que los métodos tradicionales y numéricos.
T. Anjali et al. (2019)	"Temperature Prediction using Machine Learning Approaches"	Regresión lineal múltiple, redes neuronales artificiales y máquina de soporte vectorial.	Humedad, presión, velocidad, dirección del viento, lluvia, punto de rocío, entre otras.	Obtuvieron que la técnica más confiable es la de regresión lineal múltiple.

Capítulo 3. Estado del arte

Debnail SR. (2020)	“The Air Temperature at a Weather Station Using Deep Neural Networks”	Multi-Layer Perceptron (MLP), Long Short Term Memory Network (LSTM) y una combinación de Convolutional Neural Network (CNN) y LSTM.	Promedio de la velocidad del viento, la lluvia, la caída de nieve, profundidad de la nieve, el promedio de la temperatura y las temperaturas máxima y mínima.	La combinación de Convolutional Neural Network (CNN) y LSTM obtuvo mejor rendimiento para predecir la temperatura del aire.
I. Indan et al. (2021)	. “Performance Analysis of Weather Forecasting using Machine Learning Algorithms”	Machine learning (“Backpropagation” y “Bayesian Regularization”).	Lluvia, velocidad del viento y humedad.	El algoritmo “Bayesian Regularization” es mejor que las otras técnicas de aprendizaje automático.
M. Megalai et al. (2023)	“Tenki Weather Prediction model”	Redes Neuronales. Random forest	Temperatura	Desarrollaron “Weather APIs” una aplicación móvil que permite acceder al pronóstico climatológico y una base de datos histórica.

E. Nketiah et al. (2023)	“Recurrent neural network modeling of multivariate time series and its application in temperature forecasting.”	Redes neuronales. Aprendizaje profundo.	Temperatura atmosférica, temperatura del punto de rocío, humedad relativa, presión del aire y la velocidad del viento.	El rendimiento de los modelos de aprendizaje profundo obtuvieron mejor desempeño que los modelos de referencia.
-----------------------------	---	--	--	---

Tabla 1 Estado del arte.

4 Metodología

Para realizar el proyecto de investigación de tomará en cuenta la siguiente metodología:

1. Hacer revisión bibliográfica.
2. Obtener la base de datos necesaria para realizar el proyecto.
3. Seleccionar las mejores variables independientes para predecir la temperatura.
4. Elegir las mejores técnicas de predicción y machine learning, así como elegir la mejor combinación de ellas.
5. Aplicar las técnicas seleccionadas para obtener resultados de la predicción de la temperatura en las 4 ciudades de México.
6. Realizar pruebas de parámetros de error para minimizar lo más posible el error en la predicción.

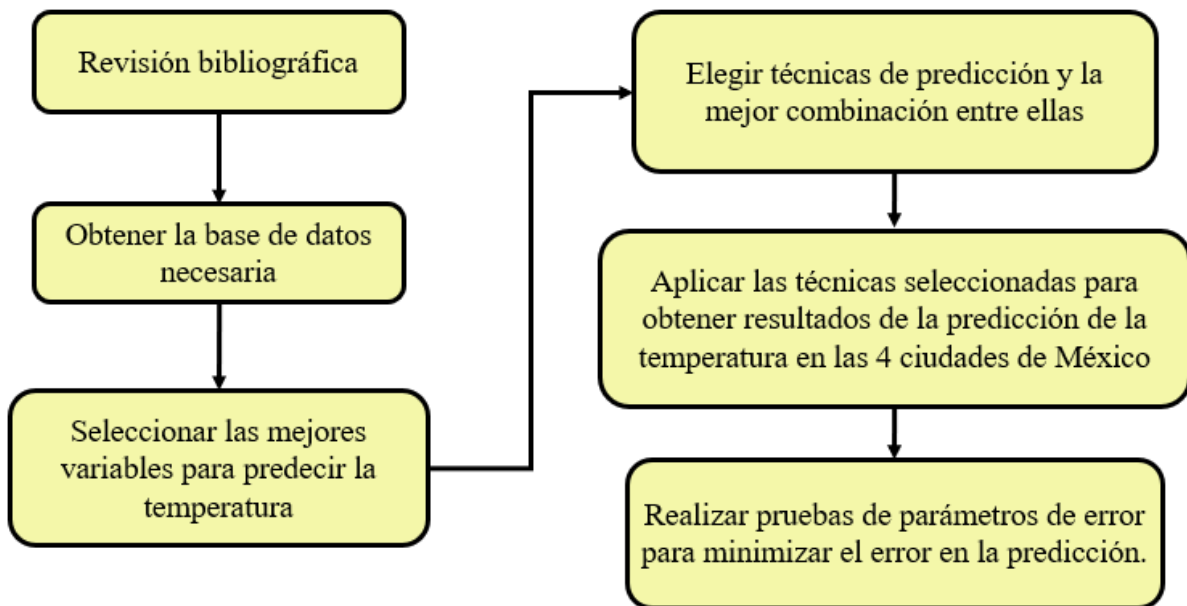


Figura 13. Metodología por realizar en el proyecto.

Series de tiempo

La base de datos utilizada en esta investigación son series de tiempo con datos climatológicos obtenidas en los sitios de internet weather underground y sinaica. Dichas series de tiempo se dividen en 2 grupos, en el primer grupo las series de tiempo describen el comportamiento de 14 variables climatológicas durante el periodo de enero del 2012 a diciembre del 2022 en las ciudades de Guadalajara, Monterrey, Tampico y Tijuana.

En el segundo grupo de series de tiempo también describen el comportamiento de variables climatológicas, pero además se agregaron variables de tipo antropogénicas; dando un total de 17 variables. Los datos de las series de tiempo de las variables del segundo grupo de series de tiempo fueron obtenidos durante el periodo de enero del 2017 a diciembre del 2021 en las ciudades de Guadalajara, Monterrey y Tijuana, no se tomó en cuenta la ciudad de Tampico debido a que no cuenta con información de las variables antropogénicas.

Grupo 1
Variables
Temperatura máxima
Temperatura mínima
Temperatura promedio
Punto de rocío máximo
Punto de rocío mínimo
Punto de rocío promedio
Humedad máxima
Humedad promedio
Velocidad del viento máxima
Velocidad del viento mínima
Velocidad del viento promedio
Presión máxima
Presión mínima
Presión promedio

Tabla 2 Variables utilizadas en el grupo 1 de las series de tiempo climatológicas

Grupo 2
Variables
Temperatura máxima
Temperatura mínima
Temperatura promedio
Punto de rocío máximo
Punto de rocío mínimo
Punto de rocío promedio
Humedad máxima
Humedad promedio
Velocidad del viento máxima
Velocidad del viento mínima
Velocidad del viento promedio
Presión máxima
Presión mínima
Presión promedio
Monóxido de carbono (CO)
Óxido nítrico (NO)
Partículas menores a 10 micras

Tabla 3 Variables utilizadas en el grupo 2 de las series de tiempo climatológicas.

Todos los datos de las variables antes mencionadas fueron normalizados para facilitar el análisis e interpretación de la base de datos. El método de normalización utilizado es el de normalización Min-Max; mediante este método obtenemos un valor dentro del rango de 0 a 1 y su fórmula se describe como:

$$T_{promedio_normalizada} = \frac{T_{promedio_estimada} - T_{min_total}}{T_{max_total} - T_{min_total}}$$

Figura 14 Fórmula para obtener datos normalizados.

Donde:

- $T_{promedio_normalizada}$ es la temperatura promedio normalizada.
- $T_{promedio_estimada}$ es la temperatura promedio obtenida de los valores de temperatura máxima y temperatura mínima de la base de datos.
- T_{min_total} es la temperatura mínima absoluta observada en todo el conjunto de datos.
- T_{max_total} es la temperatura máxima absoluta observada en todo el conjunto de datos.

Después de organizar y normalizar los datos, necesitamos dividir la base de datos para la fase de experimentación, dichas series de tiempo fueron divididas en 3 etapas: entrenamiento, validación y prueba. Los porcentajes fueron distribuidos de la siguiente manera:

Entrenamiento	80%
Validación 1	5%
Validación 2	5%
Prueba	10%

Tabla 4 Descripción de los valores porcentuales de cada fase en la experimentación.

Selección de variables

Para realizar la estimación de la temperatura se necesita hacer una selección de variables para identificar las variables que más afectan a la temperatura, por lo que en este paso de la metodología se utilizó la técnica de selección de variables llamada información de ganancia. Con esta técnica podemos decir que las variables con más información de ganancia son consideradas las más relevantes para la predicción.

Técnicas de pronóstico

Dentro de la metodología se describe la importancia de la elección de las mejores técnicas de pronóstico para las series de tiempo climatológicas. Para la experimentación de este proyecto de investigación se utilizaron diversas técnicas de predicción, tales como regresión lineal múltiple, redes neuronales, SVR y SARIMAX. El uso de una técnica convencional como regresión lineal múltiple es muy buena para comenzar la experimentación, continuamos con la segunda experimentación con una red neuronal de tipo recurrente llamada LSTM (long short term memory), la siguiente experimentación se desarrolla con la técnica SVR y por último una cuarta experimentación utilizamos ARIMA como técnica para la predicción de la temperatura.

Primera experimentación

En la primera experimentación se realiza una estimación de la temperatura utilizando las series temporales de las variables climatológicas del grupo 2 (tabla 3) en las ciudades de Guadalajara, Monterrey y Tijuana. Se aplicó la técnica llamada **Regresión lineal múltiple (RLM)**, en este modelo se busca explicar la relación que existe entre una variable dependiente (Y) en nuestro caso la temperatura y varias variables independientes (X).

El modelo que se utilizamos tiene la siguiente expresión:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \epsilon$$

Donde:

- Y: Es el valor **estimado o predicho** de la variable dependiente (la Temperatura).
- $X_1, X_2, \dots, X_k$: Son las **variables independientes** o predictoras utilizadas (las variables climatológicas del grupo 2).
- β_0 : Es el **intercepto** (ordenada al origen), el valor de Y cuando todas las variables X_i son cero.
- $\beta_1, \beta_2, \dots, \beta_k$: Son los **coeficientes de regresión** que representan el cambio esperado en Y por un cambio de una unidad en la variable X_i , manteniendo todas las demás variables constantes.
- ϵ : Es el **término de error aleatorio** o residuo, que captura la variación en Y que no explicada por el modelo.

Segunda experimentación

La segunda fase experimental se centra en el desarrollo de un modelo de **Aprendizaje Profundo** basado en redes **Long Short-Term Memory (LSTM)** para la predicción multivariable de la **Temperatura (T)**. Este enfoque se aplica a las series temporales de las cuatro ciudades seleccionadas (Monterrey, Guadalajara, Tampico y Tijuana) utilizando un conjunto refinado de variables.

El objetivo central de esta metodología es establecer un proceso de evaluación robusta que

mitigue el sesgo a la naturaleza estocástica de la inicialización de los pesos de las Redes Neuronales y las divisiones aleatorias de la validación cruzada. Para ello, se implementó un ciclo de **30 repeticiones independientes** del proceso completo de sintonización y entrenamiento. Las variables utilizadas en esta experimentación son las series de tiempo climatológicas del grupo 1 en la tabla 2.

Partición Temporal y Escalamiento

Los datos de cada ciudad fueron segmentados estrictamente por tiempo en tres conjuntos:

- **Entrenamiento (80%):** Utilizado para el ajuste del modelo y la validación cruzada.
- **Validación (10%):** Utilizado para el entrenamiento final del modelo, una vez seleccionados los hiperparámetros.
- **Prueba (10%):** Mantenido en reserva para la evaluación final del rendimiento.

Posteriormente, se aplicó un escalamiento al rango $[0, 1]$ mediante el transformador MinMaxScaler. Este paso es crucial para la estabilidad y convergencia de la red LSTM, asegurando que las funciones de activación operen dentro de un rango óptimo. El ajuste (fit) de los escaladores se realizó únicamente con los datos de entrenamiento para evitar la fuga de información (data leakage).

Arquitectura y Optimización de la Red Neuronal

Arquitectura LSTM

El modelo implementa una arquitectura secuencial con dos capas LSTM apiladas, lo que permite a la red aprender dependencias temporales de diversos niveles de abstracción:

- **Capa LSTM 1:** Utiliza la opción `return_sequences=True` para asegurar que la secuencia completa de la primera capa sirva como entrada para la siguiente capa LSTM.
- **Dropout:** Se aplicó una capa de Dropout con una tasa de 0.2 después de cada capa LSTM para regularizar el modelo y prevenir el sobreajuste.
- **Capas Densas:** La salida de la última LSTM se conecta a una capa densa intermedia con activación `relu`, seguida de una capa de salida con una única neurona (predicción T).

Sintonización de Hiperparámetros con Grid Search

La optimización de la red se realizó mediante un **Grid Search**, aprovechando el wrapper KerasRegressor de Scikit-learn para integrar el modelo Keras en la estructura de Cross-Validation. La búsqueda en rejilla se enfocó en parámetros clave de la arquitectura y el proceso de aprendizaje:

Tipo de Parámetro	Hiperparámetros (Ejemplos de la Rejilla)
Arquitectura	Unidades LSTM (64, 128), Unidades capa densa (32, 64)
Optimización	Tasa de Aprendizaje de Adam (LR: 0.01, 0.001, 0.0001)
Ajuste	Épocas (epochs)

Tabla 5 Hiperparámetros en la red neuronal

Evaluación Estocástica: Para introducir la aleatoriedad necesaria en la sintonización, se utilizó la clase KFold con shuffle = True y una semilla controlada (random_state) basada en el índice de la repetición externa. Esto garantiza que las divisiones de la validación cruzada y la inicialización de los pesos de la red son diferentes en cada una de las 30 repeticiones.

3. Evaluación Robusta y Resultados Finales

3.1. Proceso de Repetición y Entrenamiento Final

El bucle de **30 repeticiones** siguió el siguiente protocolo en cada iteración:

1. **Asignación de Semilla:** Se establece una semilla aleatoria basada en el índice de la repetición para controlar la inicialización de pesos y la CV.
2. **Grid Search:** Se ejecuta el Grid Search para encontrar el mejor conjunto de hiperparámetros para esa semilla.
3. **Entrenamiento Final:** El modelo con los hiperparámetros óptimos es reentrenado con el conjunto Xcombined (Entrenamiento + Validación) durante las epochs óptimas.
4. **Predicción:** Se realiza el pronóstico sobre el conjunto de prueba (Xlstm_test).

Tercera experimentación

En la tercera experimentación se fundamenta en un modelo de Regresión de Vectores de Soporte (Support Vector Regression, **SVR**) bajo un esquema de análisis de series de tiempo multivariable. Para garantizar la solidez y la estimación imparcial del rendimiento del modelo, se diseñó un protocolo riguroso que incluye el preprocesamiento de los datos, la transformación de la serie de tiempo a un formato supervisado mediante ventanas deslizantes, la optimización exhaustiva de hiperparámetros (HP) con ayuda de la técnica de **Grid Search** y una evaluación estadística repetitiva.

Las variables utilizadas en esta experimentación son las series de tiempo climatológicas del grupo 1 en la tabla 2 y la división de los datos en esta experimentación esta descrita en la tabla 4.

Dada la sensibilidad de los algoritmos basados en la distancia (como SVR) a la escala de las variables, todas las características de entrada (X) y la variable objetivo (Y) fueron escaladas linealmente al rango [0, 1] mediante el método MinMaxScaler. Es fundamental destacar que la función fit del escalador se aplicó exclusivamente al conjunto de entrenamiento para evitar la fuga de información hacia los conjuntos de validación y prueba.

La optimización de los hiperparámetros del SVR se realizó mediante la técnica de **Grid Search** sobre el conjunto de entrenamiento, combinada con una validación cruzada (CV=3). La rejilla de búsqueda abarcó los siguientes parámetros clave:

- **Parámetro de Penalización:** {0.1, 1, 10} (error de entrenamiento y la regularización).
- **Coeficiente del Kernel:** {'scale'}, 0.01, 0.1} (Define la influencia de cada muestra de entrenamiento).
- **Parámetro Epsilon:** {0.01, 0.1, 0.2} (Define el margen de tolerancia de la banda de regresión, dentro del cual no se aplica penalización).

Una vez identificada la mejor combinación de hiperparámetros para cada repetición del Grid Search, el modelo SVR fue reentrenado con estos parámetros óptimos. El entrenamiento final se llevó a cabo utilizando el conjunto combinado de Entrenamiento y Validación para maximizar la cantidad de datos disponibles para el ajuste del modelo antes de la evaluación final. Esta experimentación se repite en cada una de las ciudades para obtener los resultados de

un pronóstico de la temperatura con un análisis multivariable utilizando SVR en las ciudades de Monterrey, Guadalajara, Tampico y Tijuana.

Cuarta experimentación

En la última experimentación se realiza la predicción de la temperatura mediante un modelo **SARIMAX (Seasonal AutoRegressive Integrated Moving Average with exogenous variables)**, una extensión robusta de los modelos ARIMA que incorpora efectos estacionales y la influencia de variables exógenas. Esta metodología se considera superior a los modelos univariados ARIMA en contextos de análisis multivariable.

El análisis se fundamenta en la **serie endógena** (Y), que es la temperatura a pronosticar, y un conjunto de **variables exógenas** (X) que comprenden la humedad, velocidad del viento y el punto de rocío.

Los datos se dividen estrictamente de forma **cronológica** en conjuntos de **Entrenamiento (80%)** y **Prueba (20%)**. La partición asegura que la sintonización del modelo se realice únicamente con información histórica, reservando el último segmento para una evaluación imparcial del rendimiento fuera de la muestra. A diferencia de las Redes Neuronales, los modelos SARIMAX no requieren normalización de datos; no obstante, se verifica que tanto Y como X sean de tipo numérico flotante.

La identificación de la estructura del modelo SARIMAX (p, d, q)(P, D, Q, m) se realiza de manera eficiente utilizando el algoritmo **auto_arima** de la librería pmdarima. Este algoritmo implementa una búsqueda de pasos (stepwise) para explorar el espacio de hiperparámetros de manera exhaustiva pero controlada.

Variables Exógenas: El conjunto Xtrain se incluye explícitamente en la función de búsqueda para determinar el modelo SARIMAX multivariado óptimo.

Parámetros de Búsqueda: Se define un rango de órdenes no estacionales (hasta pmax=5 y qmax=5 y estacionales (hasta Pmax=1 y Qmax=1).

Frecuencia Estacional (\$m\$): Se establece un valor de m=7, asumiendo un patrón de estacionalidad semanal en las variables climáticas.

Criterio de Información: El modelo se selecciona minimizando el **Criterio de Información de Akaike (AIC)**, una métrica que penaliza la complejidad del modelo (número de parámetros) a la vez que recompensa su capacidad de ajuste.

El modelo final se construye utilizando los órdenes (p, d, q)(P, D, Q, m) determinados por el proceso auto_arima.

Esta experimentación se repite en cada una de las ciudades para obtener los resultados de un pronóstico de la temperatura con un análisis multivariable utilizando SARIMAX en las ciudades de Monterrey, Guadalajara, Tampico y Tijuana.

Métricas de evaluación

RMSE

En la primera experimentación se mide el desempeño del modelo con **Raíz del Error Cuadrático Medio (RMSE}**). El RMSE es una métrica de escala dependiente que mide la desviación promedio entre los valores pronosticados y los valores reales.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

Donde:

- **n:** Es el **número total de observaciones** en el conjunto de prueba.
- **t:** Es el índice de cada observación, de 1 a n.
- **Y_t :** Es el **valor real** observado de la variable dependiente (Temperatura) en el instante de tiempo t.
- **$\hat{Y}_t$:** Es el **valor pronosticado** por el modelo para la variable dependiente en el instante de tiempo t.
- **$(Y_t - \hat{Y}_t)^2$:** Elevar el error al cuadrado asegura que las diferencias negativas y positivas no se anulen, y que los errores mayores tengan un peso significativamente superior en

la suma final.

- $\sqrt{}$: Aplicar la raíz cuadrada al final devuelve el error a las unidades originales de la variable dependiente.

Un valor bajo de RMSE no solo indica una buena capacidad predictiva, sino también que el modelo no produjo desviaciones extremadamente grandes en sus pronósticos.

MAPE

En la segunda, tercera y cuarta experimentación se mide el desempeño de los modelos mediante el **Error Porcentual Absoluto Medio (MAPE)** que proporciona una medida del error del pronóstico directamente interpretable en términos de porcentaje de desviación.

La métrica MAPE calcula la media de los errores porcentuales absolutos de cada observación, proporcionando una medida del error de predicción en términos relativos.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100$$

Donde:

- **n** : Es el número total de observaciones en el conjunto de prueba.
- **t** : Es el índice de cada observación, de \$1\$ a \$n\$.
- **Y_t** : Es el valor real observado de la variable dependiente (Temperatura) en el instante de tiempo \$t\$.
- **$\hat{Y}_t$** : Es el valor pronosticado por el modelo para la variable dependiente en el instante de tiempo \$t\$.
- **$|\dots|$** : Denota el valor absoluto, asegurando que los errores de sobreestimación y subestimación se sumen de forma positiva.
- **$\times 100$** : Se utiliza para expresar el resultado como un porcentaje.

El resultado del MAPE es el error promedio de su pronóstico, expresado como un porcentaje del valor real. Por ejemplo, un MAPE del 5% indica que el pronóstico promedio del modelo se desvía en un 5% de los valores reales.

5 Resultados y conclusiones

El presente capítulo expone los resultados obtenidos de las cuatro experimentaciones realizadas, cuyo objetivo principal fue evaluar la capacidad predictiva de diferentes técnicas de predicción de modelado multivariable (Regresión Lineal Múltiple, Aprendizaje Profundo, Máquinas de Vectores de Soporte y Econometría de Series de Tiempo) para la temperatura en las ciudades seleccionadas. La evaluación se centra en la comparación de las métricas MAPE y RMSE.

Primera experimentación (RLM)

Los resultados se muestran a continuación:

Regresión lineal		
Variable dependiente	Variables independientes	RMSE (Error cuadrático medio)
Temperatura promedio	- Punto de rocío - Humedad	2.0277
Temperatura promedio	- Punto de rocío - Oxido Nitrico (NO)	5.2335

Tabla 6 Resultados predicción con RLM.

En la tabla 6 se pueden ver los resultados obtenidos en base a la raíz del error cuadrático medio (RMSE) de la predicción de la temperatura promedio, donde notablemente es mejor el desempeño del modelo con las variables punto de rocío (PR) y humedad (H).

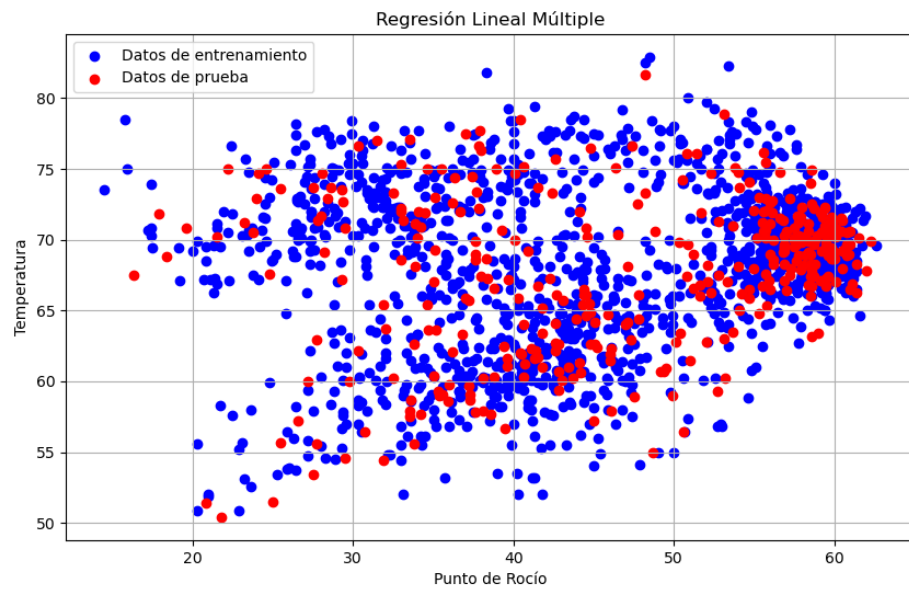


Figura 15. Punto de rocío y temperatura promedio.

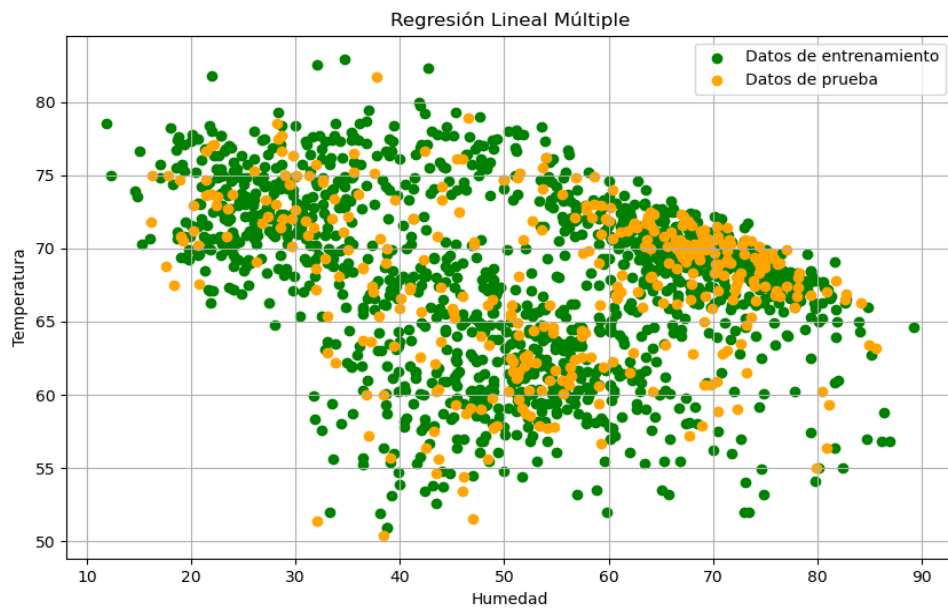


Figura 16. Humedad y temperatura promedio.

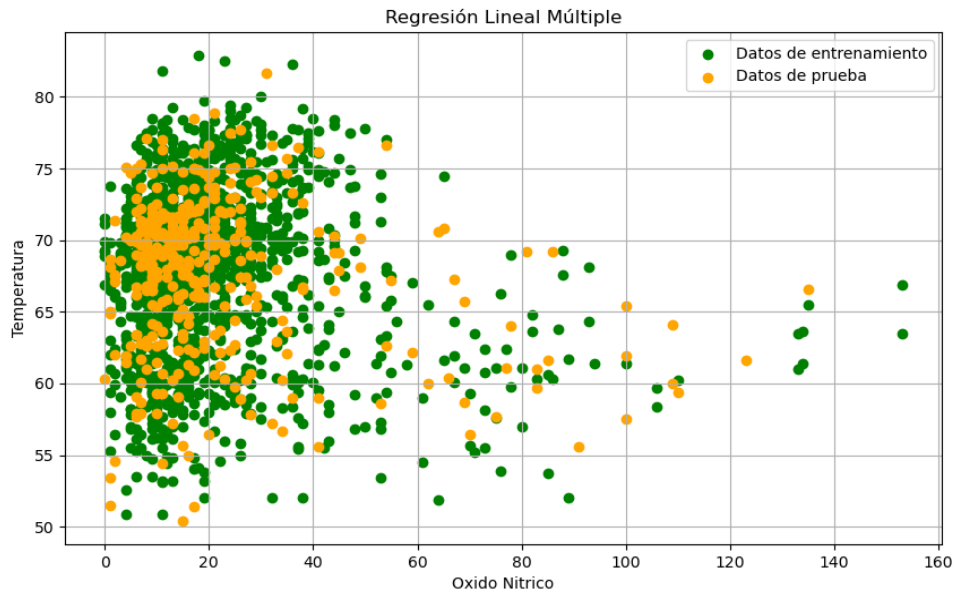


Figura 17. Óxido nítrico y temperatura promedio.

En las figuras 15, 16 y 17 se muestra la relación entre las variables dependientes con la variable independiente (temperatura promedio). Se puede observar como la tendencia de las figuras 15 y 16 son parecidas, es por eso que al realizar una predicción con un modelo donde se incluye las variables dependientes de dichas figuras (punto de rocío en la figura 15 y humedad en la figura 16) se obtiene una estimación de la temperatura con menor error a comparación de la predicción obtenida de las variables de las figuras 15 y 17 (punto de rocío pero ahora en combinación del óxido nítrico) donde no se ve alguna relación en el comportamiento de las variables dependientes con respecto de la variable independiente.

Las siguientes experimentaciones detallan los resultados obtenidos de cada técnica de pronóstico implementada (LSTM, SVR y SARIMAX) para las mismas variables independientes antes de realizar el análisis comparativo por ciudad. La evaluación se centra en la métrica Error Porcentual Absoluto Medio (MAPE), la cual proporciona una medida del error de predicción en términos relativos, facilitando una comparación de la precisión entre los distintos contextos geográficos.

Segunda experimentación (LSTM)

La experimentación basada en redes **Long Short-Term Memory (LSTM)** se diseñó para aprovechar su capacidad de capturar dependencias temporales de largo plazo y modelar las relaciones no lineales dentro del conjunto de variables multivariantes. El MAPE es el resultado promedio de las 30 ejecuciones robustas por ciudad.

Ciudad	MAPE (%)
Guadalajara	2.29
Tampico	3.15
Tijuana	3.50
Monterrey	5.35
Promedio	3.57

Tabla 7 Rendimiento de precisión MAPE del modelo LSTM en las ciudades estudiadas.

El MAPE promedio para el modelo LSTM fue de **3.57**, indicando una precisión general competente. El mejor desempeño de la red neuronal recurrente se observó en **Guadalajara** (2.29), sugiriendo que las dinámicas temporales y multivariantes en esa ciudad fueron óptimamente mapeadas por la complejidad y profundidad del modelo. En contraste, el error más alto se registró en **Monterrey** (5.35%), lo que puede implicar que las características de alta volatilidad o los patrones lineales dominantes en esta serie presentaron un desafío mayor para el modelo LSTM en su intento de extraer información no lineal.

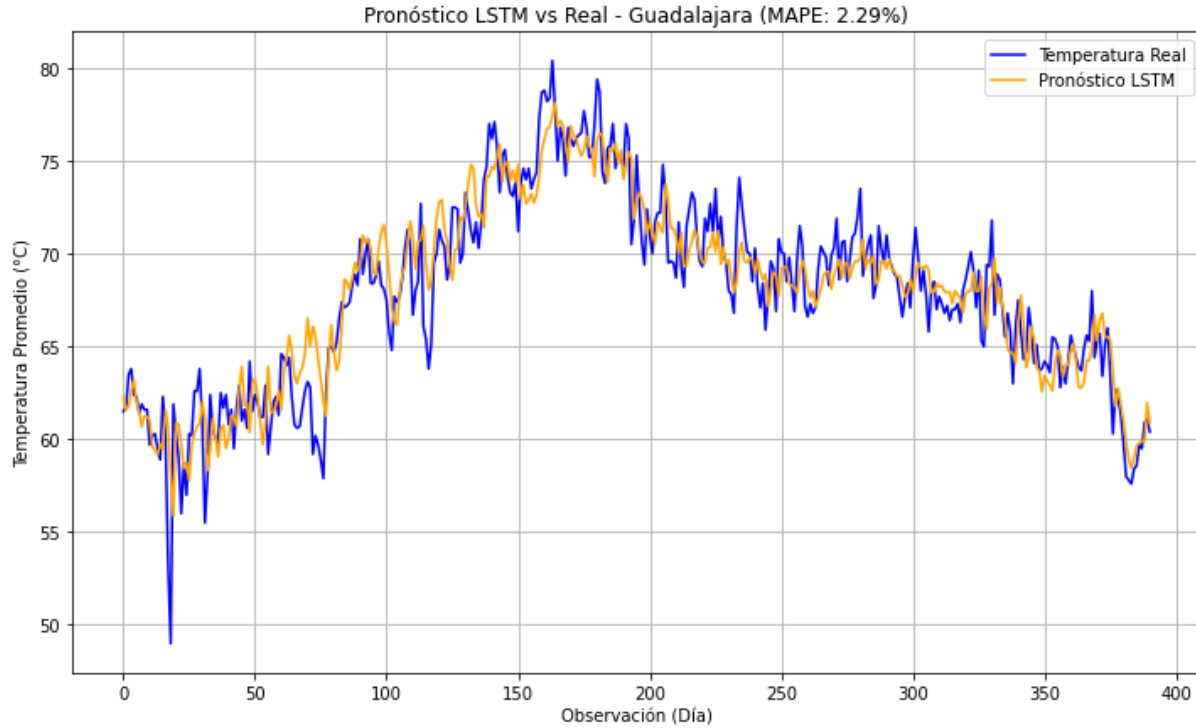


Figura 18. Desempeño de los datos reales (azul) y de los datos pronosticados (amarillo) en la ciudad con mejor desempeño (Guadalajara) usando la red neuronal LSTM.

Tercera experimentación (SVR)

El modelo de **Regresión de Vectores de Soporte (SVR)**, representando al Machine Learning clásico, utiliza el principio de minimización del riesgo estructural y maximización del margen. Su implementación con el kernel RBF y optimización Grid Search buscó determinar la función de regresión más plana posible dentro de una banda de tolerancia.

Ciudad	MAPE (%)
Guadalajara	2.32
Tampico	3.07
Tijuana	3.49
Monterrey	5.30
Promedio	3.55

Tabla 8 Rendimiento de precisión MAPE del modelo SVR en las ciudades estudiadas.

El SVR alcanzó un MAPE promedio de **3.55%**, mostrando una precisión global marginalmente superior a la obtenida por el LSTM. Es relevante que el SVR exhibió un desempeño casi idéntico al LSTM en los contextos donde la no linealidad fue mejor abordada (Guadalajara, 2.32%; Tijuana, 3.49%), confirmando su eficacia como un poderoso clasificador/regresor no lineal. El error más significativo también ocurrió en **Monterrey** (5.30%), lo que refuerza la hipótesis de que esta serie temporal posee características (posiblemente estacionales o de tendencia) que son difíciles de modelar con precisión por los algoritmos enfocados en el margen o la complejidad no lineal.

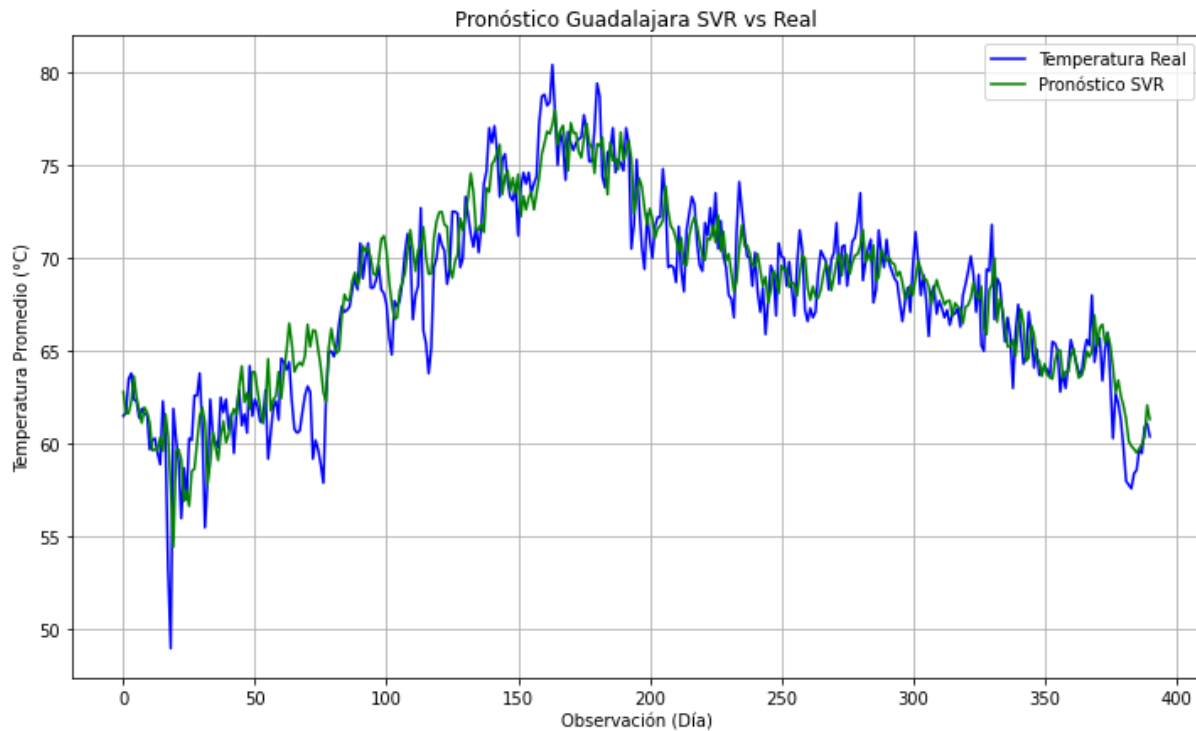


Figura 19. Desempeño de los datos reales (azul) y de los datos pronosticados (verde) en la ciudad con mejor desempeño (Guadalajara) usando SVR.

Cuarta experimentación (SARIMAX)

La modelización con **SARIMAX** representó el enfoque estadístico clásico de series de tiempo. Su función es descomponer la serie endógena en componentes autorregresivos (AR), de media

móvil (MA), y sus contrapartes estacionales (S), además de integrar el efecto de variables exógenas (X).

Ciudad	MAPE (%)
Tampico	1.13
Monterrey	2.74
Guadalajara	3.91
Tijuana	5.47
Promedio	3.31

Tabla 9 Rendimiento (MAPE) del modelo SARIMAX en las ciudades estudiadas.

El SARIMAX logró el **mejor MAPE promedio** de **3.31%**. Este resultado se debe a su desempeño excepcional en dos ciudades: Tampico (1.13%) y Monterrey (2.74%). En estos contextos, donde la estructura subyacente de la serie temporal parece estar dominada por componentes lineales o estacionales fuertes, la capacidad de SARIMAX para aislar y modelar explícitamente estas estructuras resultó en una superioridad de la precisión.

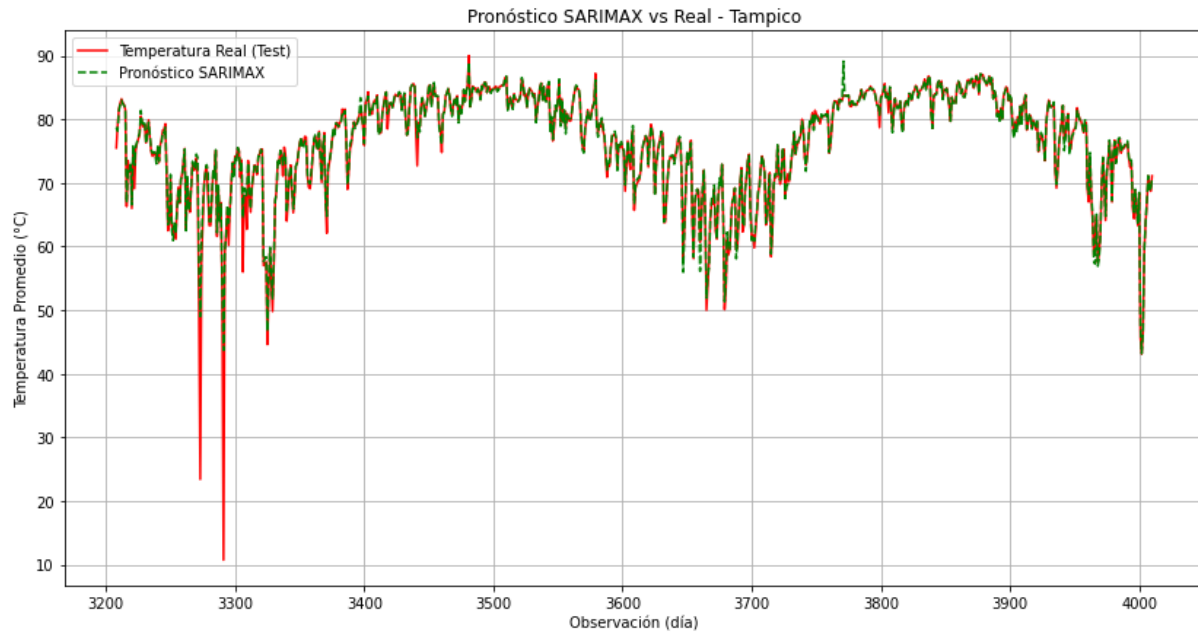


Figura 20 Desempeño de los datos reales (rojo) y de los datos pronosticados (verde) en la ciudad con mejor desempeño (Tampico) usando SARIMAX.

Por otro lado, el SARIMAX registró los errores más altos en Tijuana (5.47%) y Guadalajara (3.91%), sugiriendo que la alta variabilidad o la intensa no linealidad de estas series no pudo ser capturada eficientemente por su marco paramétrico, resultando en una considerable subestimación de la complejidad predictiva necesaria.

Análisis comparativo

Técnica	MAPE Promedio(%)	Conclusión Preliminar
SARIMAX	3.31	Más precisa a nivel global por sus resultados sobresalientes en Tampico y Monterrey.
SVR	3.55	Alta consistencia, ligeramente mejor que LSTM en el promedio.
LSTM	3.57	Fuerte capacidad no lineal, aunque fue penalizado por el alto error en Monterrey.

Tabla 10 Análisis comparativo entre las tres técnicas de pronóstico de la temperatura.

Los resultados preliminares indican que el modelo **SARIMAX** es el más preciso en el promedio global. Sin embargo, la marcada dispersión de los errores (excelente en Tampico/Monterrey vs. deficiente en Tijuana/Guadalajara) subraya la necesidad de un análisis más profundo en los siguientes resultados para determinar el modelo óptimo para cada ciudad individualmente, concluyendo que no existe una solución única para todos los climas.

Resultados por ciudad

Esta sección tiene como objetivo interpretar los resultados de MAPE mediante un análisis por ciudad. La presentación de los resultados en cada contexto geográfico, ordenados de mejor a peor desempeño, permite una identificación precisa del modelo óptimo para cada estructura de serie temporal, sustentando la conclusión de que la selección de la técnica de pronóstico es condicional a la naturaleza climática.

Guadalajara, Jalisco

Modelo	MAPE (%)
LSTM	2.29
SVR	2.32
SARIMAX	3.91

Tabla 11 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Guadalajara.

El pronóstico de la temperatura en Guadalajara evidencia una clara ventaja de los modelos no lineales. La red LSTM fue marginalmente el modelo más preciso con un MAPE de 2.29%, seguida inmediatamente por el SVR (2.32%). Esta cercanía en el rendimiento sugiere que ambos enfoques de Machine Learning fueron altamente efectivos para capturar las dinámicas multivariantes y no lineales. El modelo SARIMAX se ubicó en último lugar con un error de 3.91%. Esta diferencia significativa de 1.62 puntos porcentuales respecto al modelo ganador confirma la existencia de dependencias no lineales relevantes en la serie temporal de Guadalajara que el marco econométrico no logró modelar adecuadamente.

Tampico, Tamaulipas

Modelo	MAPE (%)
SARIMAX	1.13
SVR	3.07
LSTM	3.15

Tabla 12 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Tampico.

Tampico presentó el resultado más ilustrativo y atípico de la investigación. El modelo SARIMAX demostró una superioridad contundente con un MAPE de apenas 1.13%, superando ampliamente a SVR (3.07%) y LSTM (3.15%). Este fenómeno es crucial, pues indica que la estructura de la serie temporal de Tampico es predominantemente lineal y estacional (posiblemente por la fuerte influencia del Golfo de México). El SARIMAX capturó esta estructura subyacente de manera más eficiente que los modelos no lineales, los cuales, a pesar de su complejidad, no pudieron traducir las dinámicas temporales con la misma precisión.

Tijuana, Baja California

Modelo	MAPE (%)
SVR	3.49
LSTM	3.50
SARIMAX	5.47

Tabla 13 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Tijuana.

En Tijuana, los modelos no lineales fueron nuevamente los dominantes. El SVR fue el de mejor desempeño con un MAPE de 3.49%, con una diferencia mínima respecto al LSTM (3.50%). Este patrón establece que la complejidad de la serie de Tijuana requiere el uso de técnicas de modelado no lineal. El SARIMAX registró el MAPE más alto (5.47%), una desviación que subraya su limitación para manejar la alta variabilidad y las posibles interacciones no lineales que caracterizan este entorno climático.

Monterrey, Nuevo León

Modelo	MAPE (%)
SARIMAX	2.74
SVR	5.30
LSTM	5.35

Tabla 14 Resultados del rendimiento (MAPE) de cada técnica de pronóstico en la ciudad de Monterrey.

Monterrey, que generó los mayores errores de pronóstico para el Machine Learning, demostró ser otro dominio del enfoque estadístico. El SARIMAX alcanzó la mejor precisión con un MAPE de 2.74%. Los modelos SVR (5.30%) y LSTM (5.35%) mostraron una gran dificultad, registrando los errores más altos de toda la experimentación para sus respectivas técnicas. Este resultado implica que la serie de Monterrey está regida por patrones fuertemente estructurados (tendencias y estacionalidad) que el SARIMAX logra aislar con efectividad, mientras que los modelos no lineales se ven afectados por la incapacidad de la arquitectura para priorizar la señal lineal subyacente.

Conclusión General del Desempeño por Ciudad

El análisis desagregado por ciudad permite concluir que la superioridad del modelo es estrictamente condicional a las propiedades de la serie temporal en estudio:

- Dominio No Lineal (Guadalajara y Tijuana): Los modelos de Aprendizaje Profundo/Vector de Soporte (LSTM y SVR) son la opción óptima en ciudades con alta no linealidad y variabilidad, demostrando una mayor capacidad para mapear relaciones complejas.

- Dominio Estructural/Lineal (Tampico y Monterrey): El modelo SARIMAX es el enfoque superior en ciudades donde la serie está dominada por fuertes estructuras estacionales o ciclos predecibles. En estos casos, el SARIMAX ofrece una solución más precisa

Esta evidencia rechaza la noción de un "mejor modelo" universal para la predicción climática y sugiere que una metodología robusta debe incluir la selección del modelo en función del diagnóstico de la estructura de la serie temporal para garantizar la máxima precisión del pronóstico.

Referencias

1. Ledesma, M. (2011). Principios de meteorología y climatología. Editorial Parainfo.
2. Julio Meneses, Introducción al análisis multivariante, Universidad Obertat de Catalunya, 2019.
3. IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.
4. IPCC (Panel Intergubernamental sobre Cambio Climático). (2018). Informe Especial sobre el Calentamiento Global de 1.5 °C. Ginebra, Suiza: IPCC
5. Acuerdo de París dentro del marco de la Convención Marco de las Naciones Unidas sobre el Cambio Climático Fecha: 12 de diciembre de 2015 Organización: Convención Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC)
6. Mederos, L. (2018). Meteorología: Un libro para entender los fundamentos de la meteorología. Editorial Tutor.
7. L. Neira y E. Pérez, "TEMPERATURA Y CALOR. CONCEPTOS BÁSICOS EN LOS TEXTOS DE FÍSICA EN LA EDUCACIÓN MEDIA GENERAL," Revista Arjé, vol. 19, no. 1, pp. 327–339, 2020.
8. Ledesma, M. (2011). Principios de meteorología y climatología. Editorial Parainfo.
9. Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics (7th ed.). Pearson Education.
10. A. D. García et al., "Univariate and Multivariate Time Series Modeling using a Harmonic Decomposition Methodology," IEEE Latin America Transactions, vol. 20, no. 3, pp. 385–393, Mar. 2022. doi: 10.1109/TLA.2022.9734107.
11. J. A. Ruiz-Gascón, J. A. Olivas-Varela, and C. A. Gascón-López, "Análisis estadístico de variables climatológicas en la ciudad de Cúcuta," Ciencia y Tecnología de la Ingeniería, vol. 18, no. 1, pp. 1-13, 2018.
12. M. U. K. S. M. Masum, S. Hasan, M. A. Hasan, and M. Z. M. Masud, "Artificial Intelligence, Machine Learning and Deep Learning: A Brief Comparison," in 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), Khulna, Bangladesh, 2021, pp. 1-4.

13. M. Materia et al., "Artificial Intelligence in Climate Change Modeling and Prediction," *IEEE Access*, vol. 12, pp. 45000-45015, 2024.
14. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
15. A. Salcedo-Sanz et al., "Deep learning techniques for regional climate downscaling: A review and perspective," *Bull. Amer. Meteor. Soc.*, vol. 105, no. 5, pp. E1126-E1145, 2024.
16. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
17. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
18. Grapheverywhere, "Redes Neuronales - Conoce que son y cuales son sus características," Grapheverywhere. [En línea]. Disponible en: <https://www.grapheverywhere.com/redes-neuronales/>.
19. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
20. X. Zhang et al., "Long short-term memory neural networks for climate change prediction: A review," in *Proc. 2023 IEEE Int. Conf. Energy Internet (ICEI)*, 2023, pp. 1–6.
21. D. T. B. de Oliveira, A. L. N. da Silva, R. A. F. de Oliveira, and L. H. P. de Macedo, "A comprehensive review of long short-term memory (LSTM) neural networks: Applications and challenges," *IEEE Access*, vol. 12, pp. 11200-11225, 2024.
22. ARIMA and LSTM-Based Hybrid Model for Environmental Forecasting: A Comparative Analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2022.
23. G. Amaris, H. Ávila, y T. Guerrero, "Aplicación de modelo ARIMA para el análisis de series de volúmenes anuales en el río Magdalena," *Tecnura*, vol. 21, no. 52, pp. 88-101, 2017. doi: 10.14483/udistrital.jour.tecnura.2017.2.a07
24. "Forecasting of climate change using Seasonal Autoregressive Integrated Moving Average (SARIMA) model," in *Proc. 2021 Int. Conf. on Artificial Intelligence and Smart Systems (ICAIS)*, 2021, pp. 1297–1302.
25. Forecasting of Air Quality Parameters using SARIMAX Model," in *Proc. 2020 IEEE Int. Conf. on Power Electronics, Smart Grid and Renewable Energy (PESGRE)*, 2020, pp. 1–6.

26. J. A. Rodrigo y J. E. Ortiz, "Modelos ARIMA y SARIMAX con Python," *cienciadedatos.net*, Sep. 2023. [En línea]. Disponible en: <https://cienciadedatos.net/documentos/py51-modelos-arima-sarimax-python>.
27. "Time Series Forecasting of Temperature Using SARIMA and Neural Network Models," *IEEE Trans. Climate Change*, vol. 5, no. 3, pp. 250–261, 2018.
28. Malik, Pooja, Saranjeet Singh, and Binni Arora. "An effective weather forecasting using neural network." *International Journal of Emerging Engineering Research and Technology* 2, no. 2 (2014).
29. T. Anjali, K. Chandini, K. Anoop and V. L. Lajish, "Temperature Prediction using Machine Learning Approaches," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019.
30. Debneil SR. Forecasting. "The Air Temperature at a Weather Station Using Deep Neural Networks". *Procedia Comput Sci*. 2020 Jan 1.
31. Intan I, Rismayani , Ghani SAD, Nurdin , T CKA. "Performance Analysis of Weather Forecasting using Machine Learning Algorithms". *J Pekommas*. 2021.
32. Manimegalai, C. T., & Muthu, K. (2023). "Tenki Weather Prediction model". *Research Square (Research Square)*.
33. Nketiah, E. A., Li, C., Yingchuan, J., & Aram, S. A. (2023). "Recurrent neural network modeling of multivariate time series and its application in temperature forecasting."