



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**Evaluación de algoritmos de aprendizaje
automático en la localización de zonas de
recarga de acuíferos en la Zona Metropolitana
de la Ciudad de México y La Sierra, Tabasco**

TESIS

Que para obtener el título de
Ingeniera Geofísica

P R E S E N T A

Saraí López Lelo de Larrea

DIRECTOR DE TESIS

M.C. Manuel Ortiz Osio



Ciudad Universitaria, Cd. Mx., 2026



**PROTESTA UNIVERSITARIA DE INTEGRIDAD Y
HONESTIDAD ACADÉMICA Y PROFESIONAL
(Titulación con trabajo escrito)**



De conformidad con lo dispuesto en los artículos 87, fracción V, del Estatuto General, 68, primer párrafo, del Reglamento General de Estudios Universitarios y 26, fracción I, y 35 del Reglamento General de Exámenes, me comprometo en todo tiempo a honrar a la institución y a cumplir con los principios establecidos en el Código de Ética de la Universidad Nacional Autónoma de México, especialmente con los de integridad y honestidad académica.

De acuerdo con lo anterior, manifiesto que el trabajo escrito titulado EVALUACION DE ALGORITMOS DE APRENDIZAJE AUTOMATICO EN LA LOCALIZACION DE ZONAS DE RECARGA DE ACUIFEROS EN LA ZONA METROPOLITANA DE LA CIUDAD DE MEXICO Y LA SIERRA, TABASCO. que presenté para obtener el título de INGENIERA GEOFÍSICO es original, de mi autoría y lo realicé con el rigor metodológico exigido por mi Entidad Académica, citando las fuentes de ideas, textos, imágenes, gráficos u otro tipo de obras empleadas para su desarrollo.

En consecuencia, acepto que la falta de cumplimiento de las disposiciones reglamentarias y normativas de la Universidad, en particular las ya referidas en el Código de Ética, llevará a la nulidad de los actos de carácter académico administrativo del proceso de titulación.

SARAI LOPEZ LELO DE LARREA
Número de cuenta: 318071840

Agradecimientos

Este trabajo es el resultado de mucho tiempo, esfuerzo, dedicación, aprendizaje y mucho más. Pero lo más importante son las personas que me acompañaron en este camino.

A mi papá, la persona más importante de mi vida. No hay palabras suficientes para expresar lo agradecida que estoy de que seas mi papá, de que compartas tus conocimientos conmigo, tu tiempo, tus historias y mucho más. Este trabajo no habría sido posible sin ti. Gracias a ti conocí la carrera de geofísica, de la cual me enamoré, y gracias a tu apoyo durante toda mi educación es que estoy aquí. Eres el pilar de mi vida que siempre está ahí, desde las desveladas para llevarme a mi departamento cerca de la escuela todas las semanas, el recogerme todos los fines de semana para convivir unas cuantas horas, por apoyarme económicamente en todo momento y darme el privilegio de solo dedicarme a la escuela, por tus palabras de aliento cuando dudaba de mis habilidades y conocimientos, y sobre todo por tu honestidad, por confiar en que puedo dar más y asegurarte que este trabajo y todo lo que hago sea de calidad.

A mi tía Mayus, gracias por guiarme en este camino tan confuso en un inicio, por el tiempo que dedicaste a escucharme, a entender de qué trata este trabajo, a hacerme preguntas que en un inicio no había considerado, pero que enriquecieron el trabajo. Tus conocimientos me dieron la claridad para poder expresar los conceptos, procedimientos y resultados de manera clara. Agradezco los momentos que compartimos trabajando, pero también los momentos de ocio, las clases de coro, las pláticas y, sobre todo, el cuidado y el amor que me has dado durante toda mi vida.

A mi hermano, a pesar de nuestras peleas y desacuerdos, siempre estás para mí, gracias por apoyarme, por tu sinceridad y por estar al pendiente de cómo voy, de mi vida, de mis logros. Agradezco que estés en los buenos momentos y, sobre todo, en los no tan buenos. Aunque nunca lo diga, te amo con todo mi corazón.

A mis amigos de la carrera, Gi, Jessi, Amb, Jaime, Pao, Karencita, Eliud, Lalo, Xim. Gracias por compartir estos cinco años de sus vidas conmigo. Agradezco todos los momentos que vivimos juntos, desde las desveladas para hacer tarea, estudiar o terminar de procesar datos, hasta los momentos más random y divertidos que vivimos, por todas las pijamadas, salidas, por todas las veces que fueron a mi casa y yo a las de ustedes, por todas las pláticas y el apoyo que nos dimos en estos años, y espero que, aunque estemos a kilómetros de distancia, sigamos siendo igual de unidos.

A todos mis profesores, ya que cada uno de ellos fue esencial para mi formación como ingeniera. Pero especialmente a Manu Ortiz, Alejandro Arroyo, Carlos Garza y Orestes.

Gracias por ser parte fundamental en mi formación, no solo como ingeniera, sino también como persona. Manu, agradezco tu paciencia, los conocimientos que me compartiste, el tiempo, la dedicación y confianza que tuviste en mí durante el desarrollo de este trabajo.

Profesores Alejandro y Carlos, estoy sumamente agradecida por los conocimientos que me transmitieron, por compartirme su manera de ver el mundo y de formular preguntas que promueven la reflexión y la búsqueda de más conocimiento; pero en especial, agradezco la confianza que tienen en mí, no solo como estudiante e ingeniera, sino como persona.

Orestes, prácticamente todo lo que aprendí de salir campo fue de ti, mis primeras prácticas de campo fueron con usted, y estoy sumamente agradecida de los conocimientos que me compartió, de su paciencia a la hora de explicar, de la pasión que tiene por enseñarnos y asegurarse de que estemos preparados para el mundo laboral. En este trabajo empleo los conocimientos de SIG que me compartió, por lo que usted fue parte fundamental de este logro.

A mis sinodales Alvis e Iza, agradezco el tiempo que se tomaron para revisar mi trabajo, hacerme comentarios para mejorarlo y ser parte de esta etapa que concluye mis estudios de licenciatura.

A mis hijos Botis, Zeus y Clio. A pesar de ser de otra especie, siempre estuvieron para mí. Gracias por acompañarme durante horas en lo que trabajaba, por su cariño y empatía, en especial cuando estaba estresada. Su presencia es más que suficiente para aclarar mi mente y seguir adelante.

Finalmente, a mi mamá, sin ti, yo no existiría. Gracias por darme las habilidades, la educación, la confianza y muchísimo más. Gracias porque sé que siempre estás pendiente de mí, cuidándome y dándome todo tu amor. Agradezco los años que estuviste conmigo y en algún momento, nos volveremos a reunir. Te amo.

Dedicatoria

A mi papá, por su apoyo y amor incondicional.

Este trabajo es tan mío como suyo.

A mi mamá, por educarme, hacerme la mujer que soy y siempre cuidarme.

Índice

Resumen	7
Abstract.....	7
1. Introducción	8
2. Justificación.....	12
3. Planteamiento del problema	12
4. Antecedentes	15
5. Objetivo general	18
5.1. Objetivos específicos	18
6. Marco teórico.....	19
6.1. Hidrogeología y acuíferos.....	19
6.1.1. Parámetros de los acuíferos	21
6.2. Situación del agua subterránea en el mundo y en México	21
6.2.1. En el mundo.....	21
6.2.2. En México	22
6.3. Legislación de recursos hídricos en México	23
7. Análisis exploratorio de datos (EDA)	23
8. Codificación de variables	26
9. Aprendizaje computacional	27
9.1. Bosques aleatorios (Random Forest)	28
9.2. Lógica difusa (Fuzzy Logic (FL))	33
10. Evaluación de modelos de aprendizaje computacional	35
11. Áreas de estudio	37
11.1. Acuífero de la Zona Metropolitana de la Ciudad de México.....	37
11.2. Acuífero de La Sierra, Tabasco	40
12. Materiales y métodos	42
12.1. Creación de las bases de datos.....	42
12.2. Procesamiento de datos.....	43
13. Resultados y análisis de resultados.....	46
14. Conclusiones	59

15.	Recomendaciones y perspectiva a futuro	60
16.	Anexos	62
16.1.	Anexo 1	62
	Legislación de recursos hídricos en México	62
16.2.	Anexo 2	64
	Pasos para la creación de la base de datos	64
16.3.	Anexo 3	66
16.3.1.	Índice de los archivos	66
16.3.2.	Tipo de codificación de las variables	74
16.3.2.1.	Acuífero de la Zona Metropolitana de la Ciudad de México	74
16.3.2.2.	Acuífero de La Sierra, Tabasco	81
16.4.	Anexo 4	86
17.	Referencias	87

Índice de figuras

Figura 1 Mapa que muestra la distribución global de artículos sobre la calidad de las aguas subterráneas (a) y la cantidad de gestiones que utilizan modelos de IA híbrida (b). Los valores porcentuales se basan en el número de artículos publicados por país y en los artículos de IA híbrida relacionados (Zaresefat & Derakhshani, 2023).....	17
Figura 2 Ejemplos de gráficas: a) Diagrama de dispersión (Disfruta las matemáticas, 2025), b) Gráfica de barras (Disfruta las matemáticas, 2025), c) Histograma (Plan de Mejora, 2022), d) Matriz de correlación (Madriral, 2023), e) Matriz de coeficientes de correlación de Spearman (Ortega Martínez et al., 2025), f) Matriz de confusión (Barrios, 2019).....	26
Figura 3 Representación visual de los bosques aleatorios con sus parámetros principales (generada con ChatGPT)	31
Figura 4 Mapa de localización del acuífero de la Zona Metropolitana de la Ciudad de México	38
Figura 5 Mapa con los espesores de la zona no drenada del acuífero de la Zona Metropolitana de la Ciudad de México (Comisión Nacional del Agua (CONAGUA), 2006)	39
Figura 6 Mapa de localización del acuífero de La Sierra, Tabasco	40
Figura 7 Diagrama de flujo de la creación de los modelos de Random Forest (RF) y de Random Forest híbrido con Fuzzy Logic (RF+FL).....	44
Figura 8 Representación visual resumida de los bosques aleatorios y los bosques aleatorios con lógica difusa empleados en el trabajo (generada en chatGPT).	44
Figura 9 Matriz de confusión del modelo de Random Forest	48
Figura 10 Matriz de confusión del modelo de Random Forest con Fuzzy Logic	49
Figura 11 Mapa de comparación de las zonas de recarga del acuífero de la Zona Metropolitana de la Ciudad de México definidas por el INEGI con las obtenidas con el modelo de Random Forest..	57
Figura 12 Mapa de comparación de las zonas de recarga del acuífero de la Zona Metropolitana de la Ciudad de México definidas por el INEGI con las obtenidas con el modelo de Random Forest híbrido con Fuzzy Logic	57
Figura 13 Mapa de comparación de las zonas de recarga del acuífero de La Sierra, Tabasco obtenidas con el modelo de Random Forest con las del modelo de Random Forest con Fuzzy Logic	59
Figura 14 Leyes enfocadas en los recursos hídricos de México (Comisión Nacional del Agua (CONAGUA), 2024 c).....	63
Figura 15 Principales instituciones, entidades y dependencias que trabajan en coordinación con la CONAGUA (Comisión Nacional del Agua (CONAGUA), 2024)	64
Figura 16 Mapa con la malla generada en QGis con una separación de 660 m entre nodos, y algunas de las variables utilizadas en la base de datos del acuífero de La Sierra, Tabasco.	65

Índice de gráficas

Gráfica 1 Volumen total de agua concesionado para uso consuntivo a escala nacional en 2023 (Comisión Nacional del Agua (CONAGUA))	13
Gráfica 2 Fuente del volumen total de agua concesionado para uso consuntivo en 2023 (Comisión Nacional del Agua (CONAGUA)).....	13
Gráfica 3 Importancia de las variables en el modelo de Random Forest	50
Gráfica 4 Incremento del error tras la permutación en el modelo de Random Forest.....	50
Gráfica 5 Importancia de las variables en el modelo de Random Forest con Fuzzy Logic	51
Gráfica 6 Incremento del error tras la permutación en el modelo de Random Forest con Fuzzy Logic	51
Gráfica 7 Evaluación de la importancia de las variables en cada ejecución del modelo de Random Forest	52
Gráfica 8 Evaluación del incremento del error tras la permutación de las variables en cada ejecución del modelo de Random Forest.....	52
Gráfica 9 Evaluación de la importancia de las variables en cada ejecución del modelo de Random Forest con Fuzzy Logic	53
Gráfica 10 Evaluación del incremento del error tras la permutación de las variables en cada ejecución del modelo de Random Forest con Fuzzy Logic	53

Índice de tablas

Tabla 1 Mejores hiperparámetros para el modelo de Random Forest.....	47
Tabla 2 Mejores hiperparámetros para el modelo de Random Forest con Fuzzy Logic.....	47
Tabla 3 Valores de los parámetros de evaluación del modelo de Random Forest	47
Tabla 4 Valores de los parámetros de evaluación del modelo de Random Forest con Fuzzy Logic	48
Tabla 5 Nombre e información de las 10 variables más importantes de los modelos de Random Forest y Random Forest híbrido con Fuzzy Logic	55

Resumen

La importancia de los acuíferos radica en que proveen alrededor del 40% del agua que se utiliza en México para uso consuntivo (CONAGUA). Su estudio es complejo, debido a la falta de monitoreo de sus condiciones y a la falta de información actualizada. La identificación de zonas de recarga de acuíferos constituye una herramienta fundamental para la gestión sostenible del agua subterránea. Para ello, es necesario desarrollar metodologías que permitan localizarlas, para contribuir con su protección, conservación y si es necesario, su restauración.

El objetivo de esta investigación es comparar el desempeño de dos modelos de inteligencia artificial: Random Forest (RF) y un modelo híbrido de Random Forest con Lógica Difusa (RF+FL)) para identificar las zonas de recarga mediante el entrenamiento del modelo correspondiente, usando como entrada una base de datos de un acuífero con identificación previa de sus zonas de recarga. Para posteriormente evaluar la generalización del modelo, aplicándolo a ubicaciones sin conocimiento previo de zonas de recarga. Para ello, se utilizaron variables geoespaciales geológicas, hidrológicas y de uso de suelo; con ellas se generaron bases de datos georreferenciadas en QGis.

Los resultados muestran que ambos modelos obtuvieron un desempeño satisfactorio. En particular, el modelo híbrido de Random Forest con Fuzzy Logic obtuvo mejores indicadores de clasificación y una distribución espacial más consistente con la dinámica hidrogeológica observada del acuífero de la Zona Metropolitana de la Ciudad de México.

Abstract

Aquifers are crucial, as they provide about 40% of the water used in Mexico consumption. However, they are poorly understood due to their complexity, the lack of monitoring of their conditions, and the lack of up-to-date information. Therefore, identifying aquifer recharge zones is a fundamental tool for the sustainable management of groundwater, especially in regions facing high levels of water stress, such as Mexico, where a large portion of the country's aquifers are overexploited. It is necessary to develop methodologies that enable the identification of priority areas such as recharge zones for their protection, conservation, and, if necessary, restoration.

The objective of this research is to compare the performance of two artificial intelligence models: Random Forest (RF) and a hybrid Random Forest–fuzzy logic model (RF+FL), to identify recharge zones by training the corresponding model, using as input data a database of an aquifer in which the recharge zones have already been identified. Subsequently, the model's generalization ability was

evaluated by applying it to locations with no prior knowledge of recharge zones. To do this, geospatial variables related to geology, hydrology, and land use were used; these were used to generate georeferenced databases in QGIS.

The results show that both models performed satisfactorily; particularly, the hybrid Random Forest model with Fuzzy Logic yielded better classification metrics and a spatial distribution more consistent with the observed hydrogeological dynamics of the Mexico City Metropolitan Area aquifer.

1. Introducción

El agua subterránea constituye uno de los recursos más importantes para el ser humano, debido a que a nivel mundial representa aproximadamente el 99% del agua dulce líquida disponible (UN-Water Members and Partners, 2022), y representa alrededor del 40% del volumen de agua concesionado para el uso consuntivo en México (Comisión Nacional del Agua (CONAGUA), 2026). En los últimos años han aumentado los problemas de sobreexplotación y deterioro de la calidad de los acuíferos en el mundo, siendo el crecimiento poblacional, el incremento en las actividades productivas y el cambio climático, algunos de los factores que influyen en este proceso.

México es uno de los países que presenta mayor estrés hídrico en el mundo (Kuzma, Saccoccia, & Chertock, 2023), por lo que, en cualquier momento, si se presenta una época de sequía, la población se vería afectada, ya que no sólo se tendría escasez de agua para uso doméstico, sino que afectaría otros sectores, como la agricultura, ganadería, la industria, entre otros. Actualmente nos encontramos en esta situación, en parte, debido al monitoreo insuficiente de los recursos hídricos, a una gestión inadecuada y a una distribución desigual del recurso, que comprometen su disponibilidad para las generaciones futuras.

Debido a ello, es imperativo comprender los procesos que regulan el funcionamiento de los acuíferos, especialmente su recarga natural, ya que en las zonas de recarga es donde el agua de la precipitación o de fuentes superficiales se infiltra en el subsuelo y repone las reservas de agua subterránea. Identificar estas zonas es esencial para el diseño de estrategias de protección de la cantidad y calidad del recurso. Además, es útil para su restauración y manejo sustentable. Debido a la complejidad de los sistemas hidrogeológicos, la delimitación de zonas de recarga continúa siendo un desafío ineludible.

Aunado a esto, la falta de información dificulta la comprensión de estos sistemas. En el caso de las áreas de estudio incluidas en este trabajo, esta falta de información es evidente a la hora de generar las bases de datos. En específico, estos acuíferos tienen muy poca información geoquímica, la cual incluye datos sobre su

temperatura, pH, valores de distintos elementos químicos como K, Ca, Cl, Na, sólidos, entre otros, además de mediciones piezométricas y del nivel estático y dinámico del agua en los pozos. Los niveles antes mencionados son de suma utilidad para el cálculo del espesor de la zona drenada, el cual no se puede obtener para las condiciones actuales de los acuíferos debido a la falta de información que se tiene de ellos. Para el acuífero de la Zona Metropolitana de la Ciudad de México, en 2026 no se cuenta con ninguna medición piezométrica y para el acuífero de La Sierra solo se cuenta con una medición. De acuerdo con la información de las mediciones piezométricas desde 1996 hasta 2026 de la gerencia de aguas subterráneas, en el acuífero de la Zona Metropolitana de la Ciudad de México se tienen 125 pozos con piezómetros, sin embargo, ninguno presenta información continua en los años de ese periodo, el año que tiene la mayor cantidad de mediciones es el 2005 con 78 mediciones y los años que tienen la menor cantidad son 2020 y 2024 con 13 mediciones cada uno. En el caso del acuífero de La Sierra, Tabasco se tiene menos información, ya que, a pesar de abarcar una mayor área, cuenta con menos pozos con piezómetros, en total son 59 y ninguno de ellos tiene información continua en el periodo de 1996 a 2026; el año en que se tuvo la mayor cantidad de mediciones fue 2011 con 50 mediciones en todos los pozos, el 2026 fue el año en el que se tuvieron menos mediciones (1) y en ese periodo de tiempo se tienen 11 años sin información (Gerencia de aguas subterráneas, 2026).

En el análisis de los procesos hidrogeológicos generalmente se utilizaban métodos tradicionales; como metodologías multicriterio basadas en sistemas de información geográfica (SIG) tales como: el proceso analítico jerárquico (AHP, por sus siglas en inglés). No obstante, estos métodos a pesar de ser útiles presentan algunas desventajas: como su dificultad para representar variables complejas y no lineales, además de dificultades con la subjetividad de la asignación de pesos a las variables. Por esto surge la implementación de técnicas de aprendizaje computacional, ya que éstas subsanan dichas deficiencias y demuestran una mejora significativa en el modelado hidrogeológico.

Los algoritmos de aprendizaje computacional tienen la capacidad de procesar grandes volúmenes de información geoespacial, identificar patrones complejos y generar modelos predictivos con altos niveles de exactitud, motivo por el cual existen cada vez más ejemplos de su aplicación en hidrogeología. Entre estos algoritmos, *Random Forest (RF)* (árboles aleatorios) destaca por su robustez, su capacidad para trabajar con datos heterogéneos y su resistencia al sobreajuste.

A pesar de sus ventajas, los modelos clásicos de aprendizaje computacional presentan limitaciones para incorporar conocimiento experto y manejar explícitamente la incertidumbre asociada a los fenómenos naturales. Por ello, han surgido los modelos híbridos de inteligencia artificial, los cuales combinan técnicas

de aprendizaje automático con herramientas de razonamiento simbólico, como la lógica difusa, la cual es útil para otorgarle valores de pertenencia a los datos y dar resultados más realistas. Más aún, si se cuenta con el conocimiento de expertos, se pueden definir los rangos de pertenencia de los datos de cada variable, con el fin de obtener modelos más exactos.

La integración de *Random Forest* con *Fuzzy Logic* (FL)(lógica difusa) permite aprovechar simultáneamente la capacidad predictiva de los algoritmos de aprendizaje supervisado y la flexibilidad de la lógica difusa para representar procesos complejos e inciertos, lo que puede mejorar la calidad de los resultados. Actualmente existen algunos trabajos en los que aplican modelos híbridos para el análisis de procesos hidrogeológicos, sin embargo, es un campo al que le falta crecer mucho y desarrollar trabajos en lugares con distintas características geológicas, climatológicas y sociales. Específicamente en México, prácticamente no hay trabajos que implementen estos métodos. En este trabajo se compara un algoritmo sencillo de *Machine Learning* (ML), con uno híbrido, con el fin de evaluar el potencial de estas herramientas en distintos contextos geológicos a nivel nacional y contribuir con el desarrollo de metodologías innovadoras para la gestión del agua subterránea en el país.

Específicamente, este trabajo tiene como objetivo comparar el desempeño del método de árboles aleatorios con el método híbrido de árboles aleatorios combinado con lógica difusa, para identificar las zonas de recarga en los acuíferos. En particular en el acuífero de la Zona Metropolitana de la Ciudad de México para el entrenamiento y validación del modelo y el acuífero de La Sierra en Tabasco para evaluar la generalización del modelo en distintas zonas geológicas y en zonas que cuentan con menor disponibilidad de información geológica e hidrológica.

El acuífero de la Zona Metropolitana de la Ciudad de México se encuentra en el sur poniente de la Cuenca del Valle de México, la cual es una cuenca endorreica, que es una región geográfica en donde el agua de lluvia y los ríos que fluyen en ella no llegan al mar ni a otras cuencas, sino que se infiltra al subsuelo, se acumula en lagos y lagunas y se evapora (Portillo, 2024). Este acuífero es una cuenca endorreica de carácter lacustre, que está circundada por montañas y algunos volcanes y tiene un conjunto de lagos someros como el de Xochimilco, Chalco y el Lago de México. La litología de este acuífero está constituida principalmente por intercalaciones de productos volcánicos como lavas, tobas y cenizas con materiales granulares. Dichos materiales están cubiertos por espesores variables de arcilla y arenas finas, los cuales sirven como confinante o semiconfinante, por lo que el acuífero se clasifica como semiconfinado. El acuífero es recargado por infiltraciones de la precipitación (recarga vertical) principalmente en las sierras circundantes, de las cuales, la Sierra de las Cruces es una de las zonas más importantes de recarga,

ya que está constituida por lavas y tobas altamente fracturadas con alta porosidad, lo que facilita la infiltración del agua; por recarga de flujo horizontal subterráneo proveniente de las zonas montañosas contiguas y por recarga incidental compuesta por la infiltración de los excedentes del riego agrícola, del agua residual de las descargas urbanas y de las fugas en las redes de distribución del agua potable. No existen salidas naturales del acuífero, la única salida es por la extracción por bombeo, además, debido a que los niveles del acuífero se encuentran a 20 m de profundidad y a que la extensión de la pavimentación cubre una gran área del acuífero, no se consideran salidas por evapotranspiración. Actualmente el acuífero de la Zona Metropolitana de la Ciudad de México se encuentra sobreexplotado y la calidad de su agua rebasa las concentraciones de la Norma. Esto puede deberse al drenaje de la ciudad y a la infiltración de aguas residuales (Comisión Nacional del agua (CONAGUA), 2023 a).

En cambio, el acuífero de La Sierra se encuentra ubicado en el centro-sur del estado de Tabasco y parte del norte de Chiapas. Los materiales que afloran en el área son de tipo arcillo-arenoso, areniscas, lutitas y calizas, predominando los materiales granulares provenientes de la erosión de la Sierra de Chiapas, los cuales presentan buena permeabilidad. Las principales formaciones geológicas que se encuentran en el área son las siguientes: La formación Tres Puentes formada por sedimentos constituidos por arcillas homogéneas que presentan algunos horizontes arenosos de estratificación laminar y cuerpos de arena de cuarzo, arcilla, zonas carbonosas, turba, yeso y lignito; la Formación Belem está conformada por sedimentos marinos y lacustres de aproximadamente 250 metros de espesor, conformados por arena, capas de arcilla y restos de fósiles con yeso y la Formación Chicontepec. En la zona de estudio se encuentran alternancias de capas clásticas (lutitas-areniscas) y de carbonatos con arcillas (calizas-lutitas) y sobre ellos se encuentran depósitos de ambiente litoral conformados por areniscas, lutitas, limolitas y conglomerados. Además, se encuentran grandes espesores de conglomerados masivos que contienen entre sus clastos fragmentos de rocas ígneas. Administrativamente este acuífero se encuentra en la zona decretada de Veda del Río Grijalva, publicada en el Diario Oficial de la Federación el 25 de octubre de 1957. Esto significa que no se autorizan aprovechamientos de agua adicionales a los establecidos legalmente, con el fin de proteger al acuífero de la sobreexplotación, preservar su calidad y garantizar su disponibilidad futura (Comisión Nacional del Agua (CONAGUA), 2014). El acuífero se clasifica como zona de disponibilidad 4, lo que significa que tiene disponibilidad. El acuífero es libre y en algunos lugares opera como un acuífero semiconfinado, debido a que se tiene un paquete arcillo-arenoso que sobreyace al paquete de arenas. Su flujo subterráneo va del sur a norte. El acuífero recibe recarga vertical por infiltración en la parte sur del acuífero y en Teapa (INEGI, 2001) y recarga horizontal proveniente del piedemonte de la sierra norte de Chiapas.

Las salidas subterráneas por flujo subterráneo se presentan en el norte, hacia el acuífero Samaría-Cunduacán y Macuspana y vertical por evaporación (Comisión Nacional del agua (CONAGUA), 2023 b).

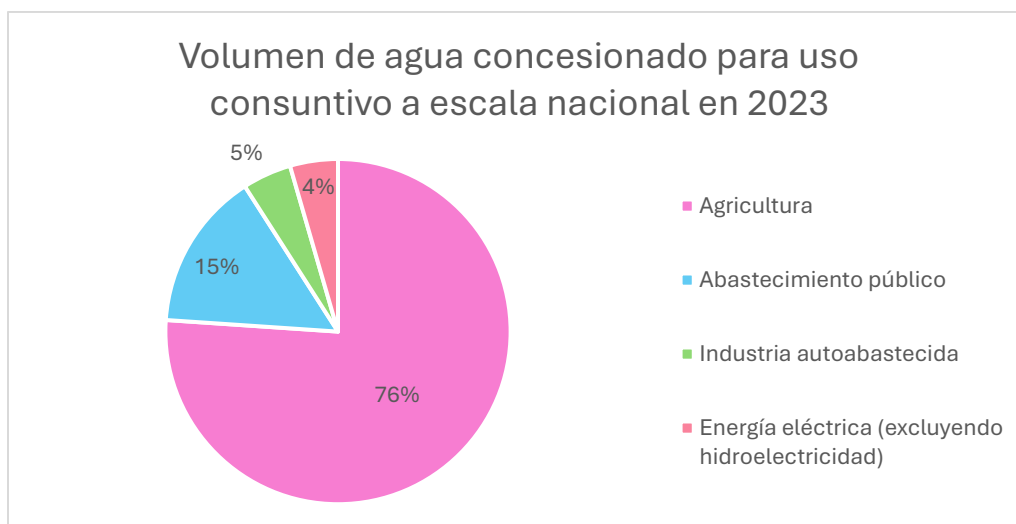
Todo ello con el fin de desarrollar una metodología que permita identificar de la manera más eficiente y realista las zonas prioritarias, como las zonas de recarga, para asegurar su debida protección, restauración y conservación y por ende preservar la cantidad y calidad de los recursos hídricos subterráneos de México (Comisión Nacional del agua (CONAGUA), 2023 b).

2. Justificación

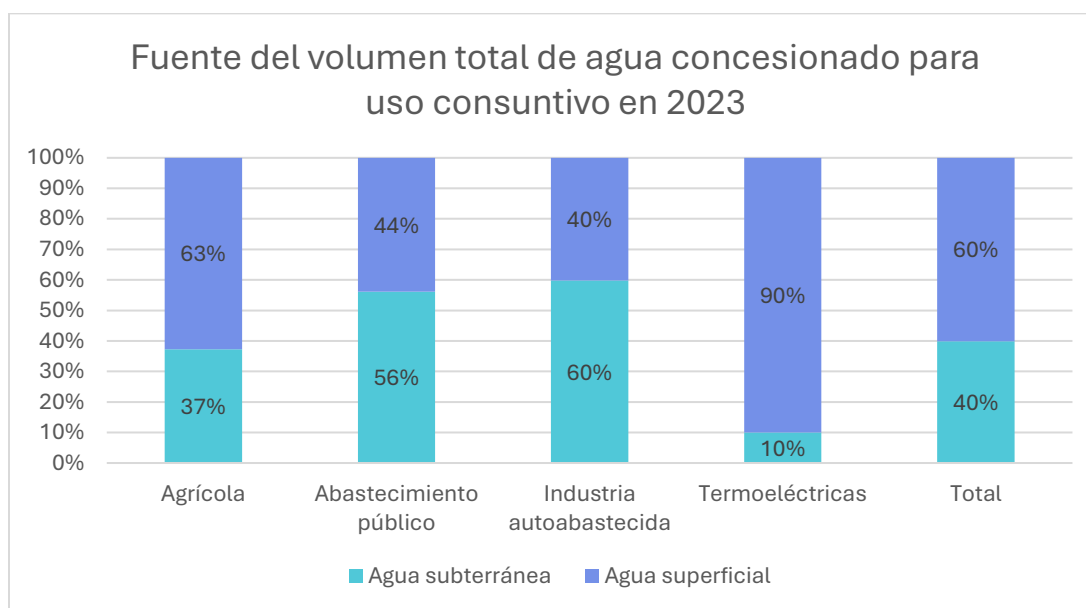
México se encuentra entre los 30 países con mayor estrés hídrico en el mundo, es decir, la proporción de extracción de agua es de un 40 al 80% mayor que la de recarga. Dicho problema es sumamente grave, ya que puede afectar a la economía y a diversas actividades como la agricultura y la ganadería, lo que impacta la seguridad alimentaria del país y la salud de la población. De acuerdo con el Instituto Mexicano de Tecnología del Agua, en 2022 el agua subterránea abasteció alrededor del 70% de las necesidades del sector público-urbano, por lo que es de suma importancia cuidar esta fuente de agua. En este trabajo se plantea comparar modelos de aprendizaje computacional para la determinación de zonas de recarga de un acuífero bien estudiado (acuífero del Valle de México), con uno menos estudiado (acuífero de La Sierra, Tabasco). Mediante la utilización de sistemas de información geográfica y algoritmos de aprendizaje computacional (*Machine Learning* (ML)), para encontrar un proceso que sea más sencillo, accesible, rápido, económico y apegado a la realidad.

3. Planteamiento del problema

En México, el volumen total de agua concesionado en 2023 fue de 267,350 hm³. De ese volumen, el 66.1% fue para uso no consuntivo (generación de energía hidroeléctrica y conservación ecológica) y el 33.9% fue para usos consuntivos. En la gráfica 1 se muestra la proporción del volumen de agua concesionado para uso consuntivo por rubro. Mientras que, en la gráfica 2 se muestra la proporción de agua subterránea y superficial que se le concesionó a cada rubro de uso consuntivo en 2023 (Comisión Nacional del Agua (CONAGUA), 2026).



Gráfica 1 Volumen total de agua concesionado para uso consuntivo a escala nacional en 2023 (Comisión Nacional del Agua (CONAGUA))



Gráfica 2 Fuente del volumen total de agua concesionado para uso consuntivo en 2023 (Comisión Nacional del Agua (CONAGUA))

El agua subterránea, a pesar de ser un recurso sumamente importante, es poco comprendida, en parte debido a la falta de información y de monitoreo constante. Actualmente, es indispensable abordar este problema lo más pronto posible debido al alto estrés hídrico que hay en México, al aumento poblacional y a la contaminación de los recursos hídricos.

El acuífero de la Zona Metropolitana de la Ciudad de México es de suma importancia, ya que abastece a la Ciudad de México y la zona conurbada. Este acuífero es de tipo semiconfinado y es recargado por infiltraciones de la

precipitación principalmente en las sierras circundantes, por recarga de flujo horizontal subterráneo proveniente de las zonas montañosas contiguas y por recarga incidental. No tiene salidas naturales, la única salida es por la extracción por bombeo. Este acuífero se encuentra sobreexplotado, por lo que es una necesidad comprender su funcionamiento y generar un plan efectivo para su administración (Comisión Nacional del agua (CONAGUA), 2023 a). En cambio, el acuífero de La Sierra en Tabasco se encuentra con disponibilidad, ya que se encuentra en la zona de veda del Río Grijalva. Este acuífero es libre y en algunos lugares opera como un acuífero semiconfinado. Recibe recarga vertical por infiltración en la parte sur del acuífero y en Teapa (INEGI, 2001) y recarga horizontal proveniente del Piedemonte de la sierra norte de Chiapas. Las salidas subterráneas por flujo subterráneo se presentan en el norte, hacia el acuífero Samaría-Cunduacán y Macuspana y vertical por evaporación (Comisión Nacional del agua (CONAGUA), 2023 b).

La falta de información y de monitoreo se hace evidente en los acuíferos estudiados en este trabajo, ya que ninguno cuenta con mediciones piezométricas constantes, ni suficientes estudios geoquímicos del agua, ni información actualizada sobre el nivel estático y dinámico del agua. En el caso del acuífero de la Zona Metropolitana de México no se cuenta con datos de evapotranspiración ni de pruebas de bombeo. En el caso del acuífero de la Sierra, a pesar de su gran extensión, cuenta con menos información comparada con el primero. El INEGI no proporciona información específica sobre las zonas de recarga y de extracción, como las que se tienen en el primer acuífero. Además, se han desarrollado menos investigaciones en esta zona.

En 2025, el INEGI publicó las “Estadísticas a propósito del día mundial del agua” en donde se menciona que en 2023 el costo asociado al agotamiento del agua subterránea y la degradación del agua superficial ascendió a \$102,029.4 millones de pesos, que equivalen a 0.32% del PIB de ese mismo año. En específico, el agotamiento del agua subterránea tuvo un costo imputado de \$35,751.8 millones de pesos, equivalente al 0.11% del PIB de la economía total del país. También, se menciona que el costo para el tratamiento de descargas de aguas residuales que degradan la calidad de los cuerpos de agua ascendió a \$66,277.6 millones de pesos, equivalente a 0.21% del PIB (Instituto Nacional de Estadística y Geografía (INEGI), 2025).

Estos datos evidencian la necesidad imperativa que existe en México de investigar mejor el funcionamiento de los acuíferos, integrando todos los sistemas que interactúan con ellos, con el propósito de proteger, gestionar y en la medida de lo posible, restaurarlos. Para lograrlo, es necesario generar modelos geoespaciales de los acuíferos.

4. Antecedentes

La Revolución Mexicana que inició en 1910 y concluyó con la promulgación de la Constitución Política de los Estados Unidos Mexicanos en 1917 y con el reparto de tierras y aguas especialmente a los que participaron como soldados en los distintos ejércitos revolucionarios. En 1926 comenzó un gran impulso a la agricultura, industrialización y urbanización promoviendo el desarrollo de infraestructura con la construcción de grandes embalses para irrigación y control de avenidas creándose la Comisión Nacional de Irrigación y se aprobó el Código Forestal que es un antecedente indirecto de protección de áreas de recarga. (Secretaría de agricultura y fomento, 1926). En 1937 se fundó la Comisión Federal de Electricidad para organizar y administrar un sistema nacional de generación eléctrica, transmisión y distribución de energía eléctrica (Comisión Federal de Electricidad (CFE), s.f.). El 1 de diciembre 1946 se fundó la Secretaría de Recursos Hidráulicos (SRH), que en ese momento fue el único ministerio del agua en todo el continente americano lo que indicó la importancia del agua como recurso estratégico. (Instituto Mexicano de Tecnología del Agua, 2020). Además, se continuó y aumentó un fuerte impulso a la construcción de grandes embalses. El 16 de enero de 1989 se creó la Comisión Nacional del Agua como órgano administrativo desconcentrado de la Secretaría de Agricultura y Recursos Hidráulicos (Instituto Mexicano de Tecnología del Agua, 2020) (Comisión Nacional del Agua (CONAGUA), 2009).

En los últimos 16 años ha cambiado la política hídrica de México debido al creciente déficit hídrico que presenta el país. Los principales cambios han sido en los siguientes rubros: la administración de acuíferos y sobreexplotación mediante la delimitación estricta de extracciones contenida en la NORMA Oficial Mexicana NOM-011-CONAGUA-2015; Establecimiento de vedas por tiempo indefinido en acuíferos sobreexplotados; Promoción de la protección de la infiltración a través de la recarga gestionada de Acuíferos (MAR); la creación de un catálogo de delitos hídricos; gestión, manejo y Protección de Cuencas; revisión de concesiones de CONAGUA para combatir la sobreexplotación y regular el uso agrícola; controles al acaparamiento; normativa más estricta para mejorar la calidad del agua y la conservación de los acuíferos y de las áreas de recarga; entre otras medidas (Presidencia de la República, 2026).

Históricamente la delimitación de zonas potenciales de recarga de agua subterránea ha dependido de metodologías cualitativas y de superposición de mapas basados en el juicio de expertos. En la literatura reciente, el Proceso Analítico Jerárquico (AHP) combinado con Sistemas de Información Geográfica (SIG) y teledetección ha sido el estándar para abordar la gestión de acuíferos.

Un ejemplo de esto es el artículo de Mejía Pineda y Murillo Álvarez (2025), en el que determinaron las zonas potenciales de recarga en un acuífero crítico para el abastecimiento urbano que enfrentaba un escenario latente de sobreexplotación. Utilizaron análisis multicriterio espacial (AHP) integrando seis variables geoambientales: geología, densidad de lineamientos, pendiente topográfica, densidad de drenaje, tipo de suelo y uso del suelo. El AHP tiene algunas limitaciones, como la alta dependencia del criterio de expertos para asignar pesos a cada variable, además de la dificultad para capturar dinámicas no lineales complejas (Mejía Pineda & Murillo Álvarez, 2025).

En los últimos años, se ha observado un notable incremento en el uso de algoritmos de aprendizaje supervisado para abordar los límites que tienen las metodologías tradicionales en la evaluación de sistemas hidrológicos complejos. Existen diversos trabajos que utilizan y comparan distintos métodos de aprendizaje computacional aplicados a la hidrología, especialmente utilizando algoritmos como bosques aleatorios, máquinas de soporte vectorial, redes neuronales artificiales, algoritmos genéticos, entre otros. En su revisión integral, Xu y Liang (2021) analizan cómo algoritmos basados en ensambles de árboles de decisión, particularmente Random Forest, superan las limitaciones de los enfoques estadísticos tradicionales al procesar de forma masiva datos de teledetección y variables hidrológicas complejas sin asumir supuestos rígidos de normalidad o linealidad (Xu & Liang, 2021).

De igual manera, Fang (2024) evalúa cinco métodos de selección de variables basados en algoritmos de Random Forest aplicados a la modelación de caudales, y demostró que las estrategias de selección hacia adelante acopladas con la flexibilidad de RF no solo reducen de forma drástica la redundancia de variables, sino que incrementan significativamente la precisión general del modelo en comparación con correlaciones estadísticas lineales tradicionales (Fang, y otros, 2024).

Además de los trabajos antes mencionados, existen diversos artículos que integran métodos de aprendizaje computacional en la hidrología; algunos de ellos están mencionados en la revisión de Xu y Liang (2021), en donde se menciona que la IA es ampliamente usada en el modelado hidrológico, gracias a su rentabilidad, velocidad y precisión en el análisis de procesos no lineales. Algunos de los usos de IA en hidrología son para la predicción de precipitación, para la predicción de inundaciones, para el modelado del escurrimiento pluvial, para la predicción del estado de los ríos, la estimación de evapotranspiración, la simulación del balance de agua, predicción de la calidad del agua, entre otros (Xu & Liang, 2021).

Los métodos de aprendizaje computacional clásicos, a pesar de producir muy buenos resultados y de su capacidad de trabajar con información no lineal, tienen

limitaciones: como la dificultad de abordar modelos con escasez de datos o cuando se ignoran las leyes físicas y el conocimiento experto del dominio. Por ello se han comenzado a integrar modelos simbólicos, como la lógica difusa, con modelos de aprendizaje computacional, resultando en modelos de inteligencia artificial (IA) híbrida. La cual puede ayudar a resolver problemas como la escasez de datos, la interpretabilidad física del modelo y la incertidumbre. El artículo de Zaresefat y Derakhshani (2023) describe a detalle acerca del manejo de agua subterránea utilizando modelos de IA híbrida. En este artículo se menciona que los usos principales que se le da a esta herramienta en hidrogeología son para evaluar el potencial del agua subterránea, estimar su recarga, administrar y predecir los niveles de agua subterránea, simular el flujo del agua subterránea, para identificar sus fuentes de contaminación, para gestionar y planificar los recursos de agua subterránea, predecir sus condiciones futuras, entre otros usos. A pesar de que ha aumentado la cantidad de modelos utilizando IA híbrida en la hidrogeología en los últimos años, es necesario que se realicen más estudios en diferentes ubicaciones geográficas para evaluar la eficacia de los modelos propuestos.

Un claro ejemplo de un país con poco desarrollo en el uso de modelos híbridos de IA es México, de acuerdo con Zaresefat y Derakhshani (2023), la cantidad de artículos que utilizan modelos híbridos en México es prácticamente nula. Como se puede observar en la figura 1, en cambio, los países que tienen la mayor cantidad de modelos de agua subterránea usando IA híbrida son China, Estados Unidos, India, Irán, Turquía, Italia, Brasil, Canadá y Rusia (Zaresefat & Derakhshani, 2023).

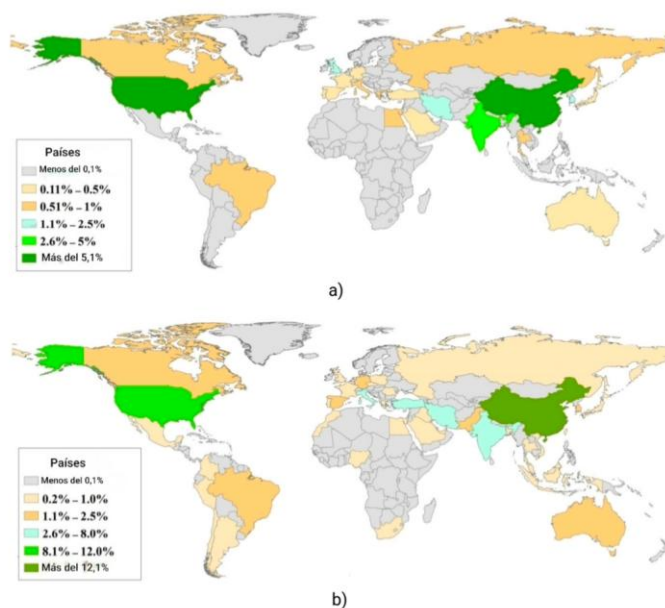


Figura 1 Mapa que muestra la distribución global de artículos sobre la calidad de las aguas subterráneas (a) y la cantidad de gestiones que utilizan modelos de IA híbrida (b). Los valores porcentuales se basan en el número de artículos publicados por país y en los artículos de IA híbrida relacionados (Zaresefat & Derakhshani, 2023).

Aunado a lo anterior, en México no hay artículos que implementen inteligencia artificial híbrida para la localización de zonas de recarga en los acuíferos. Sin embargo, sí hay artículos que implementan métodos de aprendizaje computacional, como el de Salgado Albiter *et al.* (2026) en el que obtuvieron como resultado que el método de *Random Forest* tuvo mejor desempeño que otros modelos de aprendizaje computacional y que los métodos tradicionales para generar mapas de los ecosistemas dependientes del agua subterránea (Salgado Albiter *et al.*, 2026).

Respecto a la localización de zonas de recarga con métodos tradicionales en México, podemos citar el artículo de González-Ortigoza *et al.* (2023), en el que se estima la distribución espacial y temporal de la recarga potencial de agua subterránea a escala regional en la Cuenca de México para el periodo de 2000-2021. En dicho trabajo, los autores implementaron el modelo de balance de agua en el suelo (SWB) a escala diaria y evaluaron cuatro escenarios que combinaron datos de estaciones locales con información satelital (como CHIRPS y Daymet) para subsanar la escasez de estaciones climatológicas en zonas altas. Identificaron de manera consistente que la Sierra de las Cruces y la Sierra de Chichinautzin operan como las principales zonas de recarga debido a su alta precipitación, elevación y litología volcánica fracturada, mientras que, en las zonas bajas del valle, la recarga es prácticamente despreciable (González-Ortigoza, Hernández-Espriú, & Arciniega-Esparza, 2023).

5. Objetivo general

Evaluar la eficacia de dos métodos de aprendizaje computacional, en específico Random Forest (RF) y Random Forest híbrido con Fuzzy Logic (RF+FL), mediante el entrenamiento con bases de datos geoespaciales, con el fin de ubicar zonas potenciales de recarga de dos acuíferos, uno bien estudiado, con sus respectivas zonas de recarga previamente definidas por el INEGI (acuífero de la Zona Metropolitana de la Ciudad de México) y uno poco estudiado, sin sus zonas de recarga definidas (acuífero de La Sierra, Tabasco).

5.1. Objetivos específicos

- Recopilar datos geoespaciales de los acuíferos de la zona metropolitana de la Ciudad de México y de La Sierra, Tabasco, y generar una base de datos para cada uno que contenga la litología, pendiente, edafología, uso de suelo, tipo de vegetación, cuerpos de agua, datos de pozos, estudios químicos del agua, entre otras variables.
- Codificar las bases de datos para convertir la información alfanumérica en numérica, para que pueda ser utilizada en los modelos de aprendizaje automático.

- Implementar algoritmos de *Random Forest* y de *Random Forest con Fuzzy Logic* para entrenar los modelos usando la base de datos del acuífero de la zona metropolitana de la Ciudad de México, con el fin de delimitar las zonas con potencial de recarga.
- Evaluar los modelos entrenados a partir de parámetros como pendiente topográfica, porosidad, permeabilidad, uso de suelo, tipo de vegetación, entre otros y comparar las zonas de recarga de cada modelo con las reportadas por el INEGI y la CONAGUA.
- Aplicar los modelos entrenados a la base de datos del acuífero de La Sierra en Tabasco para generar un mapa de sus posibles zonas de recarga e identificar el método más eficiente, para que se pueda aplicar a zonas menos investigadas.
- Contribuir con el desarrollo de metodologías innovadoras para la gestión del agua subterránea en el país.

6. Marco teórico

6.1. Hidrogeología y acuíferos

La hidrogeología estudia el almacenamiento, circulación y distribución de las aguas terrestres en las zonas saturadas y no saturadas de las formaciones geológicas, tomando en cuenta sus *propiedades físicas y químicas, sus interacciones con el medio físico y biológico* y sus reacciones a la acción del hombre (Water Science School, 2019).

Una parte de los rasgos hidrológicos es el agua subterránea, situada en el subsuelo por debajo del nivel freático. Se encuentra en la zona saturada, que es en donde todos los poros de las rocas contienen agua. El agua subterránea es parte del ciclo del agua e interactúa y está interconectada con el agua superficial.

De acuerdo con el Servicio geológico de los Estados Unidos (USGS), los acuíferos son formaciones geológicas que contienen suficiente agua para proveer cantidades usables a pozos y manantiales. De hecho, es la principal reserva de agua dulce y representa el 97% del agua dulce físicamente disponible en los continentes, de acuerdo con una publicación de la Gaceta UNAM.

Existen tres tipos principales de acuíferos, definidos por su geometría y su relación con la topografía y la geología del subsuelo, que son:

- Acuífero libre: Son aquellos en los que el nivel de agua se encuentra por debajo del techo de la formación permeable. Su formación se extiende prácticamente hasta la superficie terrestre, por lo que el acuífero está expuesto a la presión atmosférica. Liberan agua por desaturación, es decir,

el agua que ceden es la procedente del drenaje de sus poros (Ordoñez Gálvez, 2011).

- **Acuífero confinado:** Están cubiertos por una capa impermeable confinante. Su nivel de agua está por encima del techo de la formación acuífera. Están aislados de la comunicación de presión con las formaciones geológicas suprayacentes o subyacentes por una o más capas o unidades confinantes; y suelen estar sometidos a una presión considerable, que puede derivarse de la recarga en elevaciones más altas o del peso de la roca y el suelo suprayacentes (Tipos de acuíferos, 2017). El agua que ceden viene de la expansión del agua y de la descompresión de la estructura permeable vertical, cuando se produce la depresión en el acuífero (Ordoñez Gálvez, 2011).
- **Acuíferos semiconfinados:** Pueden considerarse un caso particular de los acuíferos confinados, en los que el muro, el techo o ambos no son totalmente impermeables, sino que permiten la circulación vertical del agua (Ordoñez Gálvez, 2011).

El agua subterránea está generalmente más protegida contra la contaminación, comparada con el agua superficial, pero una vez contaminada, es mucho más difícil de remediar. Mientras más superficiales estén los acuíferos, más vulnerables son (Ordoñez Gálvez, 2011).

La recarga de los acuíferos es el proceso mediante el cual el agua proveniente principalmente de la precipitación se infiltra a través del suelo y la zona no saturada hasta incorporarse a la zona saturada del subsuelo. Sin embargo, no toda el agua infiltrada llega a convertirse en recarga efectiva, ya que una parte puede perderse por evapotranspiración, permanecer almacenada temporalmente en la zona no saturada o desplazarse lateralmente antes de alcanzar el acuífero. La magnitud de la recarga depende de diversos factores, como la permeabilidad y porosidad de los materiales geológicos, la pendiente del terreno, el espesor de la zona no saturada, las condiciones climáticas y el uso del suelo (Martínez Guerra).

La disminución de la recarga ocurre cuando estos factores limitan la infiltración y el tránsito del agua hacia el acuífero. Entre las principales causas se encuentran la presencia de materiales de baja conductividad hidráulica, la expansión de superficies impermeables por la urbanización, la deforestación, el cambio de uso del suelo y la extracción intensiva de agua subterránea, que puede generar un desequilibrio entre la recarga y la descarga del acuífero. En consecuencia, aunque exista infiltración, la cantidad de agua que alcanza la zona saturada puede ser insuficiente para compensar las pérdidas naturales y las extracciones, ocasionando un balance hídrico negativo y una disminución progresiva del almacenamiento del agua subterránea (Martínez Guerra).

6.1.1. Parámetros de los acuíferos

Los acuíferos se definen y caracterizan por parámetros hidrogeológicos fundamentales que determinan su capacidad para almacenar agua, permeabilidad y permitir su extracción de forma viable, entre los más importantes están:

Litología (Porosidad, permeabilidad y fracturas)

La porosidad y permeabilidad de la roca son parámetros de suma importancia en la recarga de los acuíferos, en específico, qué tan rápido se mueve el agua en el subsuelo, ya que mientras más permeable sea el material y con mayor porosidad, va a permitir que el agua se mueva más rápidamente, lo que significa que el agua va a infiltrar a una tasa mayor (Water Science School, 2019).

Uso de suelo y vegetación

Es importante considerar el uso que se le da al suelo para identificar zonas en las que se ha contaminado el agua, ya que el impacto de las prácticas de uso del suelo puede tardar décadas en manifestarse en el agua subterránea. Es importante considerar el lapso que transcurre entre la aplicación de pesticidas y fertilizantes en el suelo y su llegada a los pozos, en la toma de decisiones para el manejo y la protección del agua subterránea. Dependiendo de la profundidad del agua subterránea, al disminuir la aplicación de productos químicos en la tierra, su calidad va a mejorar, mientras más somera se encuentre, más rápido va a ser este proceso, en cambio si el agua subterránea es profunda, podría tardar décadas (Water Science School, 2018 b).

Pendiente topográfica

La pendiente determina la velocidad del agua de lluvia superficial (*runoff*), actuando de forma inversamente proporcional a la recarga. A menor pendiente, el agua fluye lento y tiene más tiempo para infiltrarse y por ende, un alto potencial de recarga de agua subterránea. En cambio, a mayor pendiente, la velocidad aumenta y el agua se convierte rápidamente en escorrentía y el potencial de recarga disminuye drásticamente (Mejía Pineda & Murillo Álvarez, 2025).

6.2. Situación del agua subterránea en el mundo y en México

6.2.1. En el mundo

El reporte del Desarrollo Mundial del Agua de las Naciones Unidas de 2022 señala que el agua subterránea representa aproximadamente el 99% del agua dulce líquida en la Tierra. Esta ofrece oportunidades económicas, sociales, ambientales, y contribuye a la adaptación al cambio climático y a cumplir las metas de desarrollo sustentable (SDGs). El agua subterránea provee aproximadamente el 50% del volumen de agua usado para uso doméstico mundialmente y alrededor del 25% para riego (UN-Water Members and Partners, 2022).

A pesar de ser un recurso sumamente importante, es poco entendido, infravalorado, mal administrado y sobreexplotado. Por lo que es necesario manejar y comprender este recurso adecuadamente (UN-Water Members and Partners, 2022). Para ello hay que realizar actividades de exploración, evaluación y monitoreo constante del recurso, lo cual es costoso y difícil, lo que limita el conocimiento que se tiene sobre los sistemas locales de agua subterránea.

Existe un rápido aumento en la escasez del agua en varias partes del mundo; y ya no se puede ignorar el mal manejo del agua. Parte del aumento en esta escasez es debido al rápido aumento poblacional e industrial: como agricultura, ganadería, producción energética y manufactura. De acuerdo con información de WRI's Aqueduct Water Risk Atlas, se indica que en 2023, alrededor del 50% de la población mundial, que son alrededor de 4 mil millones de personas, vive en condiciones de alto estrés hídrico al menos un mes al año y se estima que para 2050, mil millones de personas adicionales van a vivir en condiciones de estrés hídrico extremadamente altas, lo que significa que usarían más del 80% de la recarga de agua de sus acuíferos y una sequía a corto plazo podría ponerlos en peligro de quedarse sin agua (Kuzma, Saccoccia, & Chertock, 2023).

6.2.2. En México

Alrededor del 70% del agua que se usa en las ciudades proviene de los acuíferos, lo que equivale a alrededor de 80 millones de habitantes en las zonas urbanas y 40 millones de personas que viven en zonas rurales, dependen en su mayor parte de los acuíferos. Además, una tercera parte de la superficie total irrigada depende de ellos (Pérez, 2020).

En el territorio mexicano hay 653 acuíferos, de los cuales 286 están sobreexplotados conforme a la definición que utiliza CONAGUA, esto quiere decir que la extracción de agua excede a la recarga en al menos un 10%. Además, 17 acuíferos presentan problemas de intrusión marina (Pérez, 2020).

Aunado a lo anterior, 32 de los acuíferos del país tienen problemas con la calidad del agua, específicamente el agua subterránea salobre, lo que significa que el agua deja de ser dulce, es decir, se va salinizando en consecuencia del uso agrícola de los suelos. En general los fertilizantes y los pesticidas empleados en la agricultura, junto con contaminantes derivados de desechos industriales, son la principal fuente de contaminación del agua (Pérez, 2020).

Información publicada en agosto de 2023 por WRI's Aqueduct Water Risk Atlas indica que México se encuentra en una situación de alto estrés hídrico, ocupando el lugar 26 de la lista de los países con mayor estrés hídrico en el mundo. Permanecer en estas condiciones de alto estrés hídrico afecta la vida de los ciudadanos, incluyendo los trabajos y la seguridad alimentaria y energética. Esto se debe a que

el agua es esencial para la agricultura y ganadería, en la producción de electricidad, en industrias, minería, y en cumplir con los objetivos climáticos globales (Kuzma, Saccoccia, & Chertock, 2023).

Actualmente, existe un desconocimiento del estado real de los acuíferos del país, ya que no se cuenta con suficiente información, ni con información actualizada, debido a la falta de monitoreos adecuados, lo que dificulta su administración y la implementación de medidas para protegerlos, restaurarlos y modelarlos adecuadamente. Esto se debe a que el sistema de acuíferos es sumamente complejo e involucra diversos sistemas que interactúan entre sí. Algunos de estos sistemas son el sistema natural, que incluye el ciclo del agua, la interacción del agua subterránea con el agua superficial, la interacción con los ecosistemas, entre otros. El sistema social, que abarca las localidades, el uso que le dan al agua, las necesidades de la población, etc. El sistema político y legal, que incluye leyes para proteger y extraer el recurso como: zonas de veda, la distribución de concesiones de agua, la distribución del agua, entre otras; y la parte económica, que abarca el uso de agua para industrias, agricultura, minería y para la población. Además, en los últimos años, también se debe tomar en cuenta la creciente escasez de agua y los cambios climáticos derivados del calentamiento global (Pérez, 2020).

El principal problema al que nos enfrentamos al elaborar análisis del agua subterránea y superficial es que, con frecuencia, contamos con pocos datos y con información y conocimiento limitados. Por ello, existe mucha incertidumbre al realizar nuestros análisis (Pérez, 2020).

6.3. Legislación de recursos hídricos en México

La Comisión Nacional del Agua (Conagua) es el órgano administrativo, normativo, técnico, consultivo y desconcentrado de la Secretaría de Medio Ambiente y Recursos Naturales (Semarnat). Su función es preservar las aguas nacionales y sus bienes públicos para su administración sustentable y garantizar la seguridad hídrica, con la corresponsabilidad de los órdenes de gobierno y de la sociedad en general (Comisión Nacional del Agua (CONAGUA), 2024 c).

Existen diversas normas aplicables al tema del agua conforme a la Ley Federal sobre Metrología y Normalización, las Normas Oficiales Mexicanas (NOM) son regulaciones técnicas de observancia obligatoria, en tanto que las Normas Mexicanas (NMX) son de aplicación voluntaria (Comisión Nacional del Agua (CONAGUA), 2024 c). Para más información puede consultar el anexo 1.

7. Análisis exploratorio de datos (EDA)

El análisis exploratorio de datos es un método que se utiliza para analizar conjuntos de datos y resumir sus características principales. Implica el uso de gráficos y

visualizaciones (JMP statistical discovery, 2026) es útil para determinar la mejor manera de manejar datos, lo que facilita la identificación de patrones, la detección de anomalías, la puesta a prueba de hipótesis o la verificación de superposiciones. Asimismo, nos otorga una mejor comprensión de las variables de un conjunto de datos y la relación entre ellas. También puede ayudar a determinar si las técnicas estadísticas utilizadas para el análisis son adecuadas (IBM, 2020).

Otro uso del análisis EDA es para garantizar que los resultados producidos sean válidos y aplicables a los resultados y objetivos deseados. Además, puede ayudar a responder preguntas sobre desviaciones estándar, variables categóricas e intervalos de confianza. Una vez que el EDA se completa y se extraen los hallazgos, sus características se pueden utilizar para análisis o modelado de datos más sofisticados, como el aprendizaje computacional (IBM, 2020).

El EDA es un proceso de investigación en el que se aplican resúmenes, estadísticas y herramientas gráficas, como visualizaciones multivariantes, modelos predictivos como la regresión lineal, técnicas de agrupamiento y reducción de dimensiones, K-medias, entre otros (IBM, 2020).

Hay cuatro tipos principales de EDA: Univariante no gráfico y gráfico, multivariante no gráfico y gráfico (IBM, 2020).

Las gráficas para el EDA utilizadas en este trabajo fueron las gráficas de dispersión (matriz de dispersión), histogramas, gráficas de barras, la matriz de correlación, la matriz de confusión y la matriz de índices de Spearman. A continuación, se presenta un breve resumen de cada gráfico (IBM, 2020).

Diagrama de dispersión

Muestra la relación lineal entre dos variables continuas. Esta relación puede ser: creciente, decreciente, sin correlación, correlación curvilínea; además, podemos identificar valores atípicos con este tipo de gráfico. En la matriz de dispersión se muestra la relación entre múltiples variables (JMP statistical discovery a, 2026 b).

Gráfica de barras

Refleja los conteos de frecuencia de valores de los distintos niveles de una variable categórica o nominal, ayuda a entender los niveles de sus variables y se puede usar para detectar errores y para representar estadísticas como porcentajes (JMP statistical discovery, 2026 a).

Histograma

Se usa para comprobar valores extremos o atípicos y ayudar a comprender la distribución de sus datos, los cuales son continuos. Las barras representan la

cantidad de valores que se dan en un rango específico del eje horizontal (JMP statistical discovery b, 2026 c).

Matriz de correlación

Es una tabla que muestra los coeficientes de correlación que miden la fuerza y la dirección de las relaciones entre pares de variables. El coeficiente de correlación varía de -1 a +1, donde -1 significa una correlación negativa perfecta, +1 significa una correlación positiva perfecta y 0 significa que no existe correlación alguna entre las variables (Al Masud, 2026).

Coefficiente de correlación de Spearman

Es una medida no paramétrica de la correlación entre dos variables. Calcula la dependencia estadística de la jerarquía entre dos variables (Parra, 2026). Utiliza los intervalos de los datos en lugar de los datos en sí; además, no se requiere una distribución normal de los datos. El coeficiente de correlación de Spearman varía entre -1 y 1. Si tenemos un coeficiente entre -1 y 0, existe una correlación negativa. Si el coeficiente está entre 0 y 1, hay una correlación positiva, y si es 0, no se tiene correlación. Su ecuación es la siguiente: (numiqo, 2026)

$$r_s = 1 - \frac{6 * \sum d_i^2}{n * (n^2 - 1)}$$

Donde:

- r_s es el coeficiente de correlación de Spearman
- n es el número de casos
- d es la diferencia de rangos entre las variables

Matriz de confusión

Es una herramienta que evalúa el desempeño de un modelo de clasificación supervisada. Su objetivo es mostrar cómo se comporta un algoritmo al comparar las predicciones realizadas contra los valores reales del conjunto de datos. La matriz es una tabla en la que las filas representan las clases reales, mientras que las columnas representan las clases predichas por el modelo. La matriz permite identificar cuántas predicciones fueron correctas y qué tipo de errores comete el modelo (Barrios, 2019).

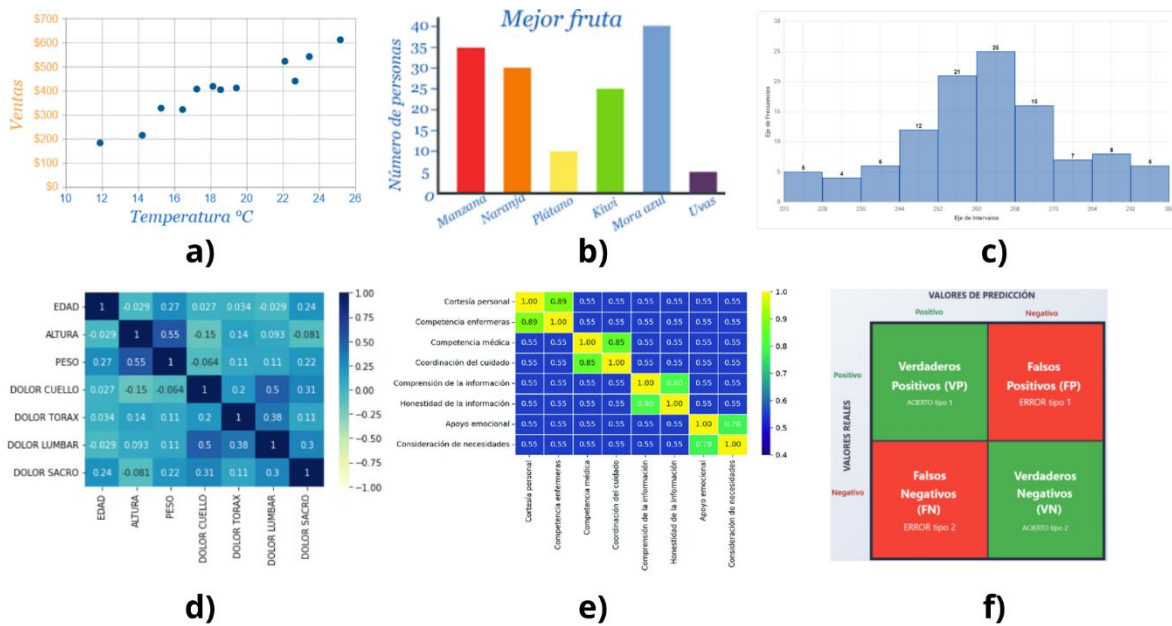


Figura 2 Ejemplos de gráficas: a) Diagrama de dispersión (Disfruta las matemáticas, 2025), b) Gráfica de barras (Disfruta las matemáticas, 2025), c) Histograma (Plan de Mejora, 2022), d) Matriz de correlación (Madrigal, 2023), e) Matriz de coeficientes de correlación de Spearman (Ortega Martínez et al., 2025), f) Matriz de confusión (Barrios, 2019).

8. Codificación de variables

Para realizar un modelo de *Machine Learning* o algún tipo de análisis estadístico, es necesario que los datos estén representados de forma numérica. En información geohidrológica es muy común encontrar datos que no necesariamente son numéricos y que en su lugar representan categorías; a este tipo de valores se llaman datos categóricos. Existen diversos métodos para obtener la representación numérica de los datos categóricos y debemos escoger el adecuado según las características de los datos (codificandobits, 2024), los cuales se dividen en dos grupos:

- Variables categóricas nominales: Son los datos que no tienen un orden inherente. Algunos ejemplos son los nombres de ciudades, tipos de cuerpos de agua, uso de suelo, entre otros, y los podemos nombrar en cualquier orden y van a seguir siendo lo mismo (codificandobits, 2024).
- Variables categóricas ordinales: Son las variables que tienen un orden inherente, como el tamaño de la vegetación, la porosidad, permeabilidad, calidad del agua, entre otros, donde la jerarquía está definida (codificandobits, 2024).

Para realizar la conversión de datos categóricos a numéricos realizamos un tipo de pre-procesamiento llamado codificación, el cual depende del tipo de variable categórica (nominal u ordinal). Existen diversos métodos de codificación; para este

trabajo se usaron *embeddings* para las variables nominales, codificación ordinal para las ordinales y *StandardScaler* para estandarizar las variables numéricas de la base de datos (codificandobits, 2024). A continuación, se describirá cada método:

- **Codificación ordinal:** Esta codificación asigna valores numéricos a las categorías, tomando en cuenta el orden de éstas. Ejemplo: Bajo = 0, Medio = 1, Alto = 2. Su desventaja es que asume que los intervalos entre las categorías son iguales, lo cual no siempre es cierto (codificandobits, 2024).
- **Embeddings:** Los embeddings convierten palabras en vectores matemáticos. Permiten capturar las relaciones semánticas y contextuales entre las palabras, lo que facilita su procesamiento y análisis por parte de los modelos de aprendizaje automático. En el espacio vectorial, palabras con significados similares van a estar más cerca unas de otras. Por ejemplo, “Agrícola” y “AGRIGOLA” estarán cerca, pero “Agrícola” y “Forestal” estarán más alejados (Huet, 2024) (Espíndola, 2023).
- **StandardScaler:** Estandariza cada variable de entrada por separado restando la media y dividiendo por la desviación estándar para que la distribución tenga una media de cero y una desviación estándar de uno. Además, no tiene un intervalo de delimitación, por lo tanto, incluso si tiene valores atípicos en sus datos, no se verán afectados por la estandarización. La mayoría de los valores van a estar entre -1 y 1, lo que disminuye las distancias relativas entre los datos (Upadhyay, 2020) (Brownlee, 2020).

9. Aprendizaje computacional

El aprendizaje computacional es un conjunto de algoritmos y herramientas que usan las computadoras para hacer y mejorar predicciones y comportamientos basados en datos. Tienen como objetivo detectar patrones en los datos que puedan utilizarse para determinar cómo se abordan los problemas que no son visibles en sistemas complejos (Dashti Latif & Najah Ahmed, 2023).

El aprendizaje computacional se divide en (Dashti Latif & Najah Ahmed, 2023):

- **Aprendizaje supervisado:** Su objetivo es inferir una función o mapeo a partir de datos de entrenamiento etiquetados, los cuales consisten en un vector de entrada de etiquetas (X) y un vector de salida de etiquetas (Y). Cada ejemplo en los datos de entrenamiento se representa por una etiqueta en el vector de salida Y. Una persona generalmente se encarga de supervisar el modelo, llamado supervisor, y proporcionan estas etiquetas para el vector de salida Y. El criterio de un experto es más costoso que el de una máquina, por lo que, para este proceso, también pueden emplearse equipos de cómputo. Sin embargo, las tasas de error en los datos etiquetados por máquinas son

mayores, lo que implica que el criterio humano es preferible, confiable y valioso.

- **Aprendizaje no supervisado:** En este caso no hay supervisores, ni datos de entrenamiento, solo tenemos datos no etiquetados. Su objetivo es encontrar una estructura dentro de los datos. Algunas razones por las que no tenemos datos etiquetados son debido a falta de recursos para pagar a un supervisor humano o a la naturaleza intrínseca de los datos.
- **Aprendizaje semi-supervisado:** Los datos que se utilizan en este tipo de aprendizaje son una combinación de datos clasificados y no clasificados. Esta mezcla se usa para generar un modelo de categorización de datos que sea adecuada. Su objetivo es crear un modelo que pueda predecir mejores clases de datos de prueba futuras que las obtenidas a partir de un modelo creado únicamente con datos etiquetados.

9.1. Bosques aleatorios (Random Forest)

Los bosques aleatorios son un algoritmo de ensamble de aprendizaje supervisado que combina la predicción de varios árboles de decisión aleatorios como base del aprendizaje. Los árboles de decisión son un método de aprendizaje supervisado no paramétrico utilizado para la clasificación y la regresión. El objetivo de los árboles de decisión es generar un modelo, que prediga el valor de una variable objetivo por medio del aprendizaje de reglas de decisión simples inferidas a partir de las características de los datos (scikit learn , 2026a).

Los árboles de decisión son fáciles de entender y de interpretar; también, pueden trabajar con datos numéricos y categóricos, por lo que se requiere poca preparación de los datos; incluso en algunos casos pueden trabajar con valores faltantes. Además, pueden resolver problemas que tienen varias salidas y es posible validar el modelo con pruebas estadísticas para obtener la fiabilidad del modelo (scikit learn , 2026a).

Sin embargo, tienen algunas desventajas, como el sobreajuste (“overfitting”), que se presenta en árboles excesivamente complejos que no generalizan bien los datos, lo cual usualmente ocurre cuando se tienen datos con muchas características. Otra desventaja es que pueden ser inestables, ya que pequeñas variaciones en los datos pueden dar lugar a la generación de un árbol completamente diferente. Aunado a esto, si algunas clases predominan, se generan árboles sesgados. Además, no son adecuados para la extrapolación ni pueden garantizar que se obtenga un árbol de decisión globalmente óptimo (scikit learn , 2026a).

Sin embargo, algunas de las desventajas antes mencionadas se pueden minimizar al modificar la base de datos que se está trabajando. Se puede reducir la dimensionalidad mediante la aplicación del análisis de componentes principales,

para eliminar las variables altamente correlacionadas, las variables vacías o con muy poca información, además juntar variables que cuenten con la misma información. Por otro lado, entrenar varios árboles con muestras del conjunto de entrenamiento seleccionadas aleatoriamente también ayuda a que todos los árboles en conjunto funcionen de forma más eficiente y eviten el sobreajuste, evitando que sean muy complejos. Asimismo, encontrar los mejores hiperparámetros para el modelo, lo que ayuda a que sus predicciones sean mejores.

Para la implementación de los bosques aleatorios se utilizan dos matrices de entrada: la primera matriz X que contiene las muestras del entrenamiento y una matriz Y que tiene la variable objetivo. Una vez ajustado, el modelo se puede utilizar para predecir la clase de las muestras (scikit learn , 2026a).

Los bosques aleatorios son métodos de ensamble, los cuales combinan las predicciones de varias variables construidas con un algoritmo de aprendizaje determinado, en este caso árboles de decisión, con el fin de mejorar la generalización y la robustez en comparación con un solo estimador. Esta predicción del ensamble se obtiene como el promedio de las predicciones de las variables individuales (scikit learn, 2026b).

Dos de los tipos de *ensamble* más utilizados son:

- **Bagging:** Se ajustan múltiples modelos, cada uno con un subconjunto distinto de los datos de entrenamiento. Para predecir, todos los modelos que forman el conjunto participan aportando su predicción. Cada modelo es distinto porque cada uno se entrena con una muestra distinta obtenida mediante bootstrapping, que es el muestreo repetido con reposición. El resultado toma la media de todas las predicciones o la clase más frecuente cuando se tienen variables categóricas (Amat Rodrigo, 2025).
- **Boosting:** Se ajustan secuencialmente múltiples modelos sencillos, llamados *weak learners*, de forma que cada modelo aprende de los errores del anterior. Como valor final, se toma la media de todas las predicciones o la clase más frecuente (Amat Rodrigo, 2025).

Para que los métodos de ensamble tengan mejores resultados que cualquiera de sus modelos individuales, es necesario que los modelos que los forman sean lo más diversos posible. Random Forest utiliza *bagging*, que es el diminutivo de *bootstrap aggregation*, con el fin de reducir la varianza de algunos modelos de aprendizaje estadístico. A veces no se tiene acceso a múltiples muestras, pero se puede simular el proceso recurriendo a *bootstrapping*, generando pseudomuestras con las que ajustar diferentes modelos y después agregarlas (Amat Rodrigo, 2025).

Conforme aumenta la complejidad del modelo, este adquiere una mayor flexibilidad para adaptarse a las observaciones, lo que reduce el sesgo y mejora su capacidad predictiva. Sin embargo, una vez que se tiene cierto nivel de complejidad, se presenta el sobreajuste. Lo ideal es generar el modelo óptimo, que es aquel que encuentra el equilibrio adecuado entre sesgo y varianza (Amat Rodrigo, 2025).

El sesgo se refiere a cuánto se desvían en promedio las predicciones de un modelo con respecto a los valores reales y refleja la capacidad del modelo para captar la verdadera relación entre los predictores y la variable de respuesta. Por otro lado, la varianza hace referencia a cuánto cambia el modelo en función de los datos utilizados en el entrenamiento. Idealmente, un modelo no debería cambiar demasiado ante pequeñas variaciones en los datos de entrenamiento; si esto ocurre, indica que el modelo está memorizando los datos en lugar de aprender la verdadera relación entre los predictores y la variable de respuesta (Amat Rodrigo, 2025).

El método de bosques aleatorios tiene ventajas adicionales sobre los árboles de decisión. Son capaces de seleccionar automáticamente los predictores más relevantes, no requieren que los datos sigan una distribución específica, son menos susceptibles a ser influenciados por valores atípicos (*outliers*), si el valor de una variable no está disponible para una observación, aún se puede realizar una predicción utilizando las observaciones del último nodo alcanzado; además, permiten identificar de forma rápida y eficiente las variables predictoras más importantes, es posible estimar el error de validación con el *Out-of-Bag Error*, y son adecuados para conjuntos de datos con un gran número de observaciones (Amat Rodrigo, 2025).

Durante la construcción de cada árbol del bosque, se toma en cuenta un subconjunto aleatorio de variables. El parámetro *max features* controla el tamaño de este subconjunto y puede incluir todas las variables de entrada o un subconjunto aleatorio de ellas. El propósito de esta aleatoriedad, es reducir la varianza del estimador de bosque, ya que los árboles de decisión individuales suelen presentar una alta varianza y tienden al sobreajuste. Adicionalmente, al introducir esta aleatoriedad, obtenemos árboles con errores de predicción relativamente independientes entre sí. Al promediar esas predicciones, algunos errores pueden anularse y así se logra reducir la varianza al combinar diversos árboles, a veces a costa de un ligero aumento del sesgo; por consiguiente, obtenemos un modelo globalmente mejor (scikit learn, 2026b).

Para este trabajo se usó la implementación *RandomForestClassifier* de *scikit learn*. Los parámetros principales son *n_estimators*, que es el número de árboles que hay en el bosque; mientras más grande sea, el modelo va a ser mejor, pero va a tardar

más tiempo en procesar y los resultados van a dejar de mejorar después de un número crítico de árboles. El otro parámetro principal es `max_features`, que es el tamaño de los subconjuntos aleatorios de variables consideradas al dividir un nodo. Cuanto menor sea este valor, mayor será la reducción de la varianza, pero también mayor será el aumento del sesgo. Al emplear *Bootstrap* en el modelo de RF, el error de generalización puede estimarse a partir de las muestras excluidas o de las muestras *out-of-bag* (scikit learn, 2026b).

A continuación, se presenta un resumen de los parámetros que forman parte del método de bosques aleatorios y en la figura 3 se muestra una representación visual de un bosque aleatorio con sus parámetros principales.

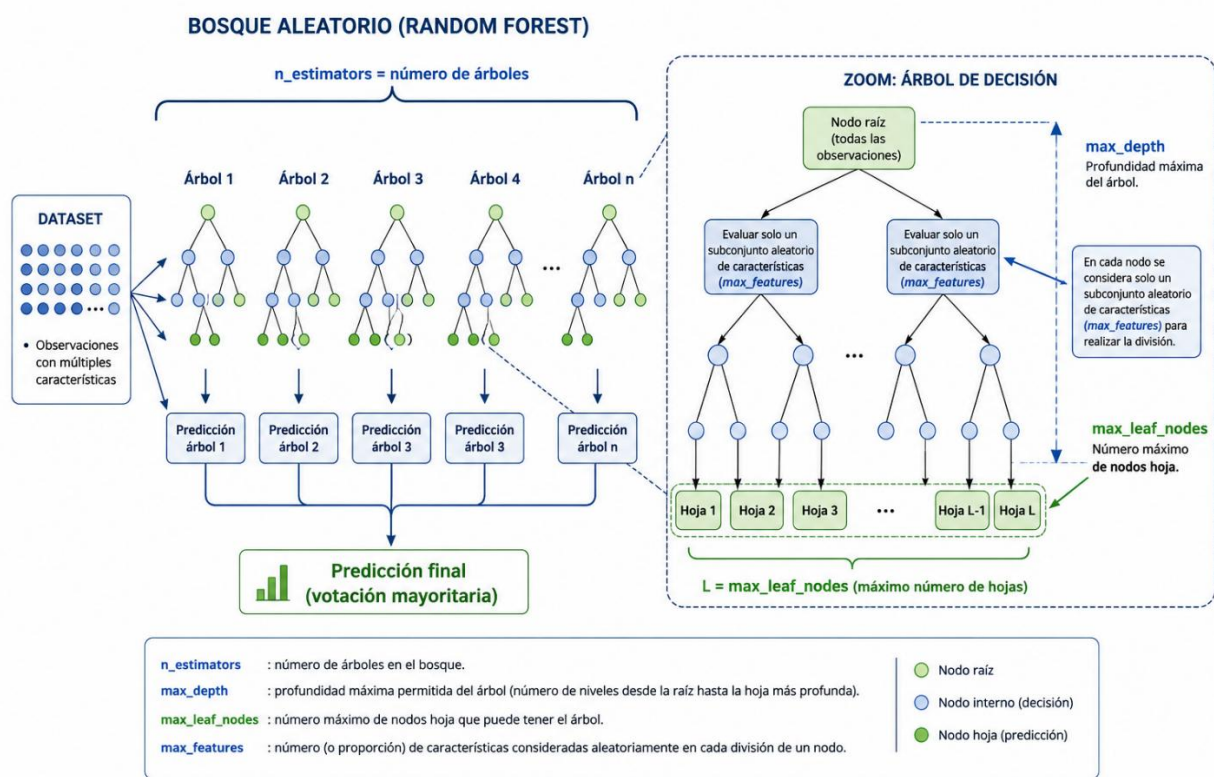


Figura 3 Representación visual de los bosques aleatorios con sus parámetros principales (generada con ChatGPT)

Parámetros:

- `n_estimators`: número de árboles incluidos en el modelo.
- `max_depth`: profundidad máxima que pueden alcanzar los árboles.
- `min_samples_split`: número mínimo de observaciones que debe tener un nodo para que pueda dividirse.

- `min_samples_leaf`: número mínimo de observaciones que debe tener cada uno de los nodos hijos para que se produzca la división.
- `max_leaf_nodes`: número máximo de nodos terminales que pueden tener los árboles.
- `max_features`: número de predictores considerados en cada división. Puede ser: un valor entero, una fracción del total de predictores, la raíz cuadrada del número total de predictores, el log2 del número total de predictores o todos los predictores.
- `oob_score`: si se calcula o no el *out-of-bag* R^2 .
- `n_jobs`: número de núcleos empleados para el entrenamiento.
- `random_state`: semilla para que los resultados sean reproducibles.
- `criterion`: define la función matemática que utiliza el modelo para medir la calidad de cada división en los árboles de decisión. Para problemas de clasificación el valor predeterminado es “gini” que calcula la impureza de Gini y suele ser computacionalmente más rápido. También se puede utilizar “entropy” que calcula la entropía, suele generar árboles ligeramente más profundos y es más estricto dividiendo nodos. (scikit learn, 2026c)

RandomForestClassifier tiene valores por defecto para sus hiperparámetros; sin embargo, no se puede saber de antemano si estos son los más adecuados. La forma de identificarlos es mediante el uso de estrategias de validación, como el *Out-of-Bag error*, que permite obtener una estimación del error de entrenamiento sin recurrir a la validación cruzada, que es computacionalmente costosa. Aunque el análisis individual de los hiperparámetros es útil para entender su impacto en el modelo e identificar intervalos de interés, la búsqueda de hiperparámetros no debe realizarse de forma secuencial, ya que cada hiperparámetro interacciona con los demás; por ello, es preferible recurrir a *grid search* o *random search* para analizar varias combinaciones de hiperparámetros (Amat Rodrigo, 2025).

El uso de *Out-of-Bag Error* tiene dos limitaciones. No es adecuado cuando las observaciones tienen una relación temporal y sus estimaciones son muy optimistas cuando se hace el preprocesado de forma conjunta (Amat Rodrigo, 2025).

Respecto a la evaluación de la importancia de las variables en el modelo, los árboles aleatorios estiman la contribución de cada variable analizando con qué frecuencia y qué tan efectivas son las divisiones que realiza en los árboles de decisión para mejorar la clasificación de los datos. En términos generales, las variables utilizadas en los primeros niveles de los árboles suelen tener mayor importancia, ya que influyen en la clasificación de un mayor número de observaciones y generan una

mayor reducción de la incertidumbre del modelo. Por ello, la importancia de una variable refleja su contribución relativa a la capacidad predictiva del modelo y no necesariamente una relación de causa y efecto con el fenómeno estudiado (scikit learn, 2026b).

En el contexto de este trabajo, si la pendiente aparece repetidamente en los primeros niveles de los árboles para diferenciar zonas de recarga y descarga, mientras que variables como la presencia de cuerpos de agua se utilizan únicamente en niveles más profundos o en pocos árboles, el modelo asignará una mayor importancia a la pendiente. Esto indica que dicha variable aporta más información para identificar las zonas potenciales de recarga, aunque su importancia siempre debe interpretarse en conjunto con el conocimiento hidrogeológico del área de estudio.

Por otra parte, la importancia de las variables por permutación identifica la influencia que tiene cada predictor sobre una determinada métrica de evaluación del modelo, estimada por *out-of-bag error* o validación cruzada. Si el predictor permutado contribuye al modelo, este aumenta su error, ya que se pierde la información que proporciona esa variable. Lo que cuantifica esto es la influencia que tienen los predictores sobre el modelo, no su relación con la variable de respuesta. Este incremento puede ser negativo (Amat Rodrigo, 2025).

9.2. Lógica difusa (Fuzzy Logic (FL))

La lógica difusa surgió con la teoría de conjuntos difusos, presentada por el matemático Lotfi Zadeh en 1965. FL aborda conceptos que no están definidos con precisión, sino que se sitúan en un espectro de posibilidades, es decir, asignan grados de posibilidad a las entradas. A esto se le llama función de pertenencia, que asocia los valores de entrada a su grado de pertenencia a un conjunto. Estos intervalos van de 0 a 1, en donde 1 significa totalmente posible, 0 completamente imposible y todos los números intermedios representan cierto grado de posibilidad (Kaskaya, 2022).

Al permitir valores intermedios obtenemos una toma de decisiones más realista, que imita los procesos de toma de decisiones humanos y subsana las deficiencias de los sistemas tradicionales de lógica binaria. Además, es esencial en la IA para gestionar la incertidumbre, la ambigüedad, la complejidad y los datos imprecisos (Kaskaya, 2022).

Los sistemas basados en reglas difusas evalúan reglas lingüísticas de si-entonces (*if-then*) usando la difusificación (fuzzification), inferencia y composición. Estas reglas de *if-then* expresan la relación entre las entradas y las salidas de manera difusa (Kaskaya, 2022). En esencia le decimos a la computadora “Si ocurre esto,

entonces haz esto”. Por ejemplo, si la roca es permeable, entonces puede haber infiltración.

Las reglas procesadas con sistemas de inferencia difusa combinan varias reglas difusas para tomar decisiones. Para realizar este proceso, primero se evalúa el grado de verdad de cada regla en función de los valores de entrada; posteriormente, se combinan los resultados de todas las reglas aplicables mediante operaciones lógicas. Finalmente se genera un resultado difuso que necesita convertirse en un resultado preciso mediante la defusificación (defuzzification) (Ferrer, 2025), como el promedio ponderado, el centroide de área (COA), el centro de gravedad (COG), entre otros (Kaskaya, 2022).

La arquitectura de un sistema de lógica difusa consta de cuatro componentes clave que funcionan conjuntamente para procesar entradas, aplicar un razonamiento similar al humano y generar salidas útiles (Ferrer, 2025):

1. Fusificación (Fuzzification): Convierte las entradas nítidas como valores precisos de temperatura o presión en valores difusos mediante funciones de pertenencia. Estos valores difusos representan grados de pertenencia a distintas categorías, lo que permite al sistema gestionar eficazmente la imprecisión, la incertidumbre y la vaguedad de los datos (Ferrer, 2025). La fusificación consta de varios pasos: Primero se introducen las entradas reales, posteriormente se obtienen las funciones de membresía o pertenencia, una función de pertenencia es cada valor de entrada que corresponde a un conjunto difuso que va de 0 a 1 (Mazraeh, 2025).

2. Base de conocimiento: Contiene un conjunto de reglas *If-then* y funciones de pertenencia proporcionadas por expertos. Estas reglas rigen el proceso de toma de decisiones al convertir las entradas en resultados aplicables (Ferrer, 2025).

3. Motor de inferencia: Emula el razonamiento humano evaluando el grado de coincidencia entre las entradas difusas y las reglas almacenadas en la base de conocimiento. Este proceso es el núcleo de la toma de decisiones en un sistema de lógica difusa, y a partir de esta evaluación se determina qué reglas activar y se combinan sus resultados para generar acciones de control (Ferrer, 2025).

4. Defusificación (Defuzzification): Transforma la salida difusa generada por el motor de inferencia en un valor nítido. Este paso traduce las acciones de control difuso en salidas precisas y aplicables, adecuadas para aplicaciones del mundo real. La elección del método de defusificación depende de los requisitos específicos del sistema (Ferrer, 2025). La técnica de defusificación más utilizada es el método del centroide. Esta técnica trata de encontrar el centro de gravedad o centroide del área bajo la curva definida por la salida difusa. Se obtiene promediando

ponderadamente los valores numéricos de la variable de salida, utilizando el nivel de certeza de las reglas como el peso de cada valor (Mazraeh, 2025).

Gracias a la combinación de estos cuatro componentes, la lógica difusa funciona con datos de entrada imprecisos, con ruido y distorsionados, requiriendo una cantidad mínima de datos y memoria, lo que la hace eficiente en cuanto al uso de recursos. Además, proporciona transiciones más fluidas entre estados, es fácil de comprender, aplicar y permite razonamientos detallados en entornos complejos (Kaskaya, 2022) (Ferrer, 2025).

A pesar de tener muchas ventajas, sus limitaciones destacan en la necesidad de tener un diseño y aplicación cuidadosos para garantizar su eficacia. Algunas de sus limitaciones son la falta de un diseño sistemático, ya que no existe un enfoque estandarizado para resolver problemas mediante la lógica difusa, lo que puede dar lugar a ambigüedades e incoherencias. Además, es complicado demostrar las características y la eficacia de los sistemas de lógica difusa, debido a la ausencia de descripciones matemáticas precisas. A pesar de gestionar bien los datos imprecisos, la lógica difusa puede comprometer en ocasiones la obtención de una alta precisión (Ferrer, 2025).

10. Evaluación de modelos de aprendizaje computacional

En el aprendizaje computacional, la utilidad de los modelos depende de la calidad de su capacidad predictiva y de su eficacia para resolver problemas reales. La evaluación de modelos es el proceso de cuantificar el rendimiento de un modelo de aprendizaje automático para encontrar el que mejor se ajuste al problema planteado. Esto es fundamental para garantizar que funcionen de forma óptima y correcta al implementarse en producción y así proporcionen predicciones fiables y confiables (Team Comet, 2022).

Existen diferentes métricas de evaluación, pero las más comunes son la exactitud, la precisión, la tasa de error, la exhaustividad, la matriz de confusión y la puntuación F1 (Team Comet, 2022).

Para este trabajo se usó la implementación de scikit learn y se validaron los modelos de aprendizaje automático, con las métricas derivadas de la matriz de confusión (para más información puede consultar la sección de análisis exploratorio de datos). La matriz de confusión es una herramienta fundamental para evaluar el desempeño de modelos de clasificación supervisada. Su objetivo es mostrar de forma clara y estructurada el comportamiento del algoritmo al comparar las predicciones realizadas contra los valores reales del conjunto de datos. Éstas tienen cuatro combinaciones posibles entre el valor real del modelo y el valor predicho:

- Verdadero Positivo (VP / TP – True Positive): El valor real es positivo y el modelo también predice positivo.
- Verdadero Negativo (VN / TN – True Negative): El valor real es negativo y el modelo predice negativo.
- Falso Negativo (FN – False Negative): El valor real es positivo, pero el modelo predice negativo.
- Falso Positivo (FP – False Positive): El valor real es negativo, pero el modelo predice positivo.

Los valores obtenidos de estos cuatro escenarios son la base de las métricas derivadas de la matriz de confusión, las cuales son:

F1-score: Es la media armónica entre precisión y sensibilidad. Penaliza modelos que optimizan una métrica a costa de otra (Barrios, 2019).

Exactitud (Accuracy): Mide la proporción total de predicciones correctas del modelo (Barrios, 2019).

Precisión (Precision): Es la métrica más popular para evaluar modelos (Team Comet, 2022). Indica qué proporción de las predicciones positivas son realmente correctas (Barrios, 2019).

Sensibilidad (Recall): Da la tasa de verdaderos positivos (TPR), es decir, indica la capacidad del modelo para detectar correctamente los valores positivos (Barrios, 2019).

Especificidad (Specificity): Da la tasa de verdaderos negativos (TNR), es decir, indica la capacidad del modelo para detectar correctamente los valores negativos (Barrios, 2019).

Combinando precisión y sensibilidad (*recall*), pueden darse cuatro escenarios (Barrios, 2019):

- Alta precisión y alto *recall*: El modelo detecta correctamente la clase y lo hace de manera confiable.
- Alta precisión y bajo *recall*: El modelo detecta pocos casos de la clase, pero cuando lo hace, acierta con mucha seguridad.
- Baja precisión y alto *recall*: El modelo detecta bien la clase, pero tiene muchos falsos positivos.
- Baja precisión y bajo *recall*: El modelo no logra identificar correctamente la clase.

11. Áreas de estudio

A continuación, se presenta una síntesis de la ubicación, geología, topografía, el clima, los tipos de acuífero, las condiciones actuales de los acuíferos de las zonas de estudio, con base en los datos de la CONAGUA y un breve análisis utilizando las herramientas de AQUEDUCT.

Las herramientas de AQUEDUCT son parte de la información que provee el Instituto de los recursos del mundo (*World Resources Institute*). Tienen diversas herramientas, que son mapas que analizan el estado actual y futuro de los recursos hídricos a nivel mundial. Las utilizadas en este trabajo fueron el *Aqueduct Water Risk Atlas* (AQUEDUCT b, 2023), que muestra los riesgos relacionados con el agua y evalúa la exposición a dichos riesgos en el mundo. La otra herramienta utilizada fue el *Aqueduct Country Rankings* (AQUEDUCT a, 2023), que compara los riesgos relacionados al agua a nivel nacional y subnacional. Este riesgo integra factores como cantidad física de agua, su calidad, incertidumbre por los cambios regulatorios, conflictos con la población debidos a problemas con el agua, entre otros factores. La escala de riesgo en ambas herramientas va de 0 a 5, donde 5 es riesgo hídrico extremadamente alto.

11.1. Acuífero de la Zona Metropolitana de la Ciudad de México

El acuífero de la Zona Metropolitana de la Ciudad de México se encuentra ubicado en el sur-poniente de la Cuenca del Valle de México. Abarca la Ciudad de México y una porción del Estado de México (Figura 4). La Zona Metropolitana de la Ciudad de México se encuentra limitada por elevaciones topográficas de origen volcánico: la Sierra de Guadalupe al norte, la Sierra de Las Cruces al poniente, La Sierra Chichinautzin al sur, el volcán del Ajusco al suroeste y la Sierra Nevada al oriente. La altitud de las sierras del sur alcanza los 2,240 metros, mientras que en el norte llegan a los 2,390 metros (Comisión Nacional del Agua (CONAGUA), 2024 b).

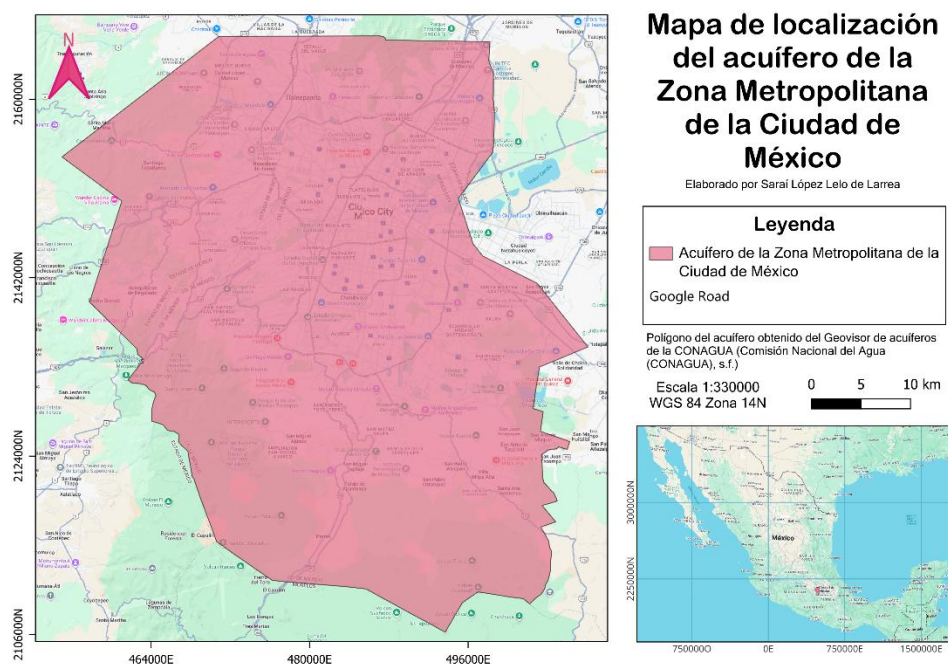


Figura 4 Mapa de localización del acuífero de la Zona Metropolitana de la Ciudad de México

La zona de estudio presenta varios climas debido a sus características topográficas. Dentro de esta variedad predomina el clima templado, que se concentra en los valles altos de la parte noroeste del Estado de México, así como en el centro y este de la subregión Valle de México; le siguen en importancia los climas secos y semisecos que predominan principalmente en la porción centro-norte del Estado de México (Comisión Nacional del Agua (CONAGUA), 2024 b).

El acuífero de la Zona Metropolitana de la Ciudad de México presenta de manera irregular material arcilloso que funge como confinante o semiconfinante, por lo que el acuífero se clasifica como semiconfinado. En este, su contenedor impermeable está constituido por rocas volcánicas y calizas, mientras que su paquete sedimentario da origen a un sistema acuífero complejo formado por tres grandes cuerpos: en la parte superior, un paquete arcilloso de alta porosidad, baja permeabilidad y gran heterogeneidad en su constitución, que forma un acuitardo de espesor variable y que actúa como semiconfinante en el centro de la cuenca (Comisión Nacional del Agua (CONAGUA), 2024 b).

Bajo este paquete se encuentra el acuífero actualmente en explotación, formado por material granular más grueso que el del acuitardo, que está formado principalmente por piroclastos y conglomerados de origen volcánico; su espesor es variable, generalmente mayor a 200 m. A este paquete le subyacen rocas volcánicas fracturadas, cuya base llega a estar a 2000 m en el centro, disminuyendo hacia los márgenes de la Cuenca (Comisión Nacional del Agua (CONAGUA), 2024 b).

Este acuífero se recarga por infiltraciones producto de la precipitación, donde las áreas de recarga más importantes son las sierras circundantes. No existen salidas naturales del acuífero, debido a que es una cuenca endorreica, la única salida es mediante la extracción por bombeo. En 2006 los niveles dentro del acuífero oscilaron entre 23.9 metros en Tláhuac hasta 250 m de profundidad en Milpa Alta (Comisión Nacional del Agua (CONAGUA), 2006). Estos niveles están relacionados con el espesor de la zona no drenada, ya que éste es el espesor que existe desde la superficie hasta el nivel en donde se encuentra el agua. En la figura 5 se muestran las isopacas de la zona no drenada del acuífero. Además, no se considera que existan salidas por evapotranspiración, ya que hay pavimentación en una gran área de este (Comisión Nacional del Agua (CONAGUA), 2024 b).

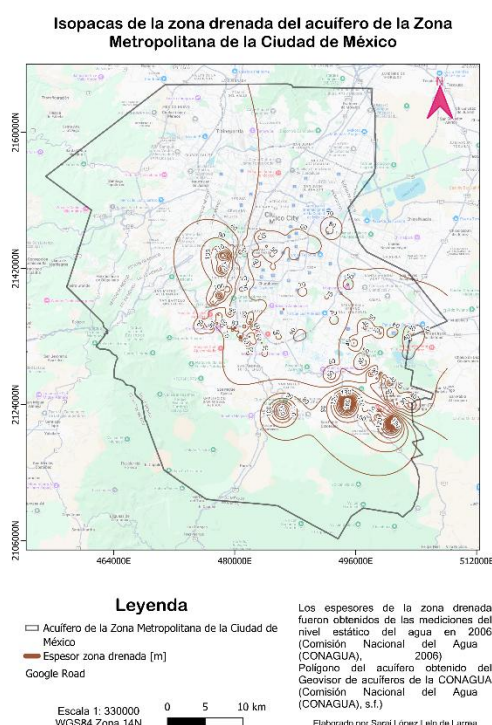


Figura 5 Mapa con los espesores de la zona no drenada del acuífero de la Zona Metropolitana de la Ciudad de México (Comisión Nacional del Agua (CONAGUA), 2006)

Este acuífero se encuentra sobreexplotado, de acuerdo con los datos publicados por la CONAGUA (Comisión Nacional del agua (CONAGUA), 2023 a). En específico, se tiene que la recarga media anual (R) es de 512.8 [hm³ anuales], mientras que el volumen de extracción de aguas subterráneas (VEAS) es de 993.23 [hm³ anuales], lo que significa que la disponibilidad media anual de agua subterránea (DMA) es de -480.43 [hm³ anuales], lo que significa que la extracción es 93.69% mayor que la recarga. Esto se puede corroborar con las herramientas de AQUEDUCT mencionadas al inicio de este capítulo (AQUEDUCT a, 2023) (AQUEDUCT b, 2023). En ellas se observa que la zona del acuífero del Valle de

México presenta estrés hídrico extremadamente alto, lo que se traduce en una descarga mayor al 80% de la recarga. Lo que implica que el riesgo relacionado con el agua de esta zona de estudio es alto (3-4) (AQUEDUCT a, 2023) (AQUEDUCT b, 2023) (Comisión Nacional del agua (CONAGUA), 2023 a).

11.2. Acuífero de La Sierra, Tabasco

El acuífero La Sierra se localiza en la porción centro-sur del estado de Tabasco (Figura 6), abarca los municipios de Teapa, Tacotalpa, Jalapa y, parcialmente, los de Macuspana y el Centro (Comisión Nacional del Agua (CONAGUA), 2024 a).

Se encuentra situado en dos provincias fisiográficas: en la Llanura Costera del Golfo Sur y las Sierras de Chiapas y Guatemala, la primera es una llanura formada por grandes cantidades de aluvión acarreado por el río Grijalva, el cual atraviesa la provincia para desembocar en la parte sur del Golfo de México, este río se une al río Usumacinta cerca de la ciudad de Frontera, Tabasco, donde tienen desembocadura común en el Golfo de México. Además, en la zona cruzan los ríos de La Sierra, Teapa y Puyacatengo, los cuales son de régimen permanente (Comisión Nacional del Agua (CONAGUA), 2024 a).

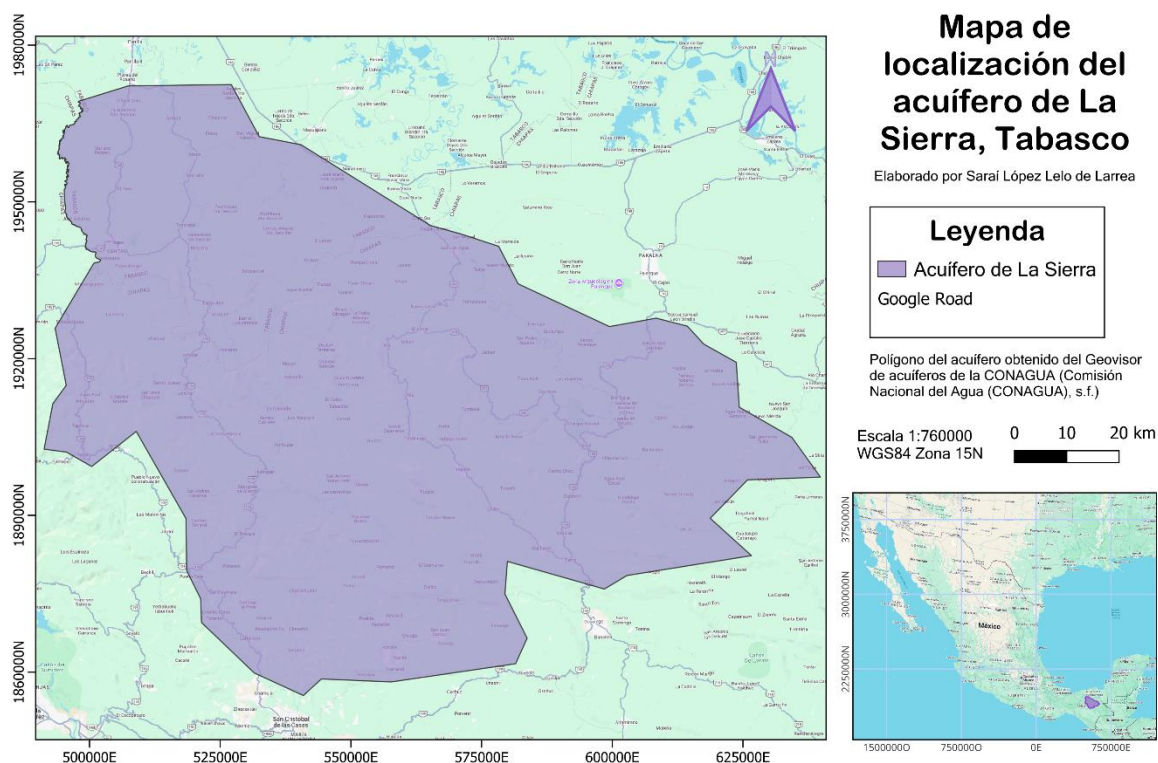


Figura 6 Mapa de localización del acuífero de La Sierra, Tabasco

Los principales materiales que afloran en el área son de tipo arcillo-arenoso, como areniscas, lutitas y calizas, predominando los materiales granulares provenientes de la erosión de la sierra de Chiapas. En el norte afloran depósitos constituidos por areniscas de grano grueso poco consolidadas. En la zona de estudio se encuentran depósitos aluviales formados por depósitos sin consolidar, de granulometría variable que va de arcillas a arenas y gravas. Superficialmente afloran rocas sedimentarias representadas por dos tipos de suelo, el primero de tipo arcilloso en el norte del acuífero y el segundo está constituido por material areno arcilloso en la parte sur. La cadena montañosa está conformada por estructuras de roca sedimentaria plegadas con tendencia a orientarse noroeste-sureste y se presentan una serie de fallas y fracturas (Comisión Nacional del Agua (CONAGUA), 2024 a).

El clima en la región es cálido húmedo con lluvias todo el año. Durante los meses de octubre a marzo frecuentemente se presentan “nortes” en la zona, son fríos y generalmente vienen acompañados de fuertes vientos y precipitaciones ligeras (Comisión Nacional del Agua (CONAGUA), 2024 a).

El sistema opera como un acuífero libre y en algunos sitios puede considerarse de tipo semiconfinado, ya que en algunas zonas se tiene un paquete arcillo-arenoso que sobreyace al paquete de arenas (Comisión Nacional del Agua (CONAGUA), 2024 a).

El acuífero recibe recarga vertical por infiltración en la porción sur, sureste y suroeste del estado de Tabasco, que es donde se encuentran las máximas elevaciones y altas precipitaciones, además la zona serrana de Teapa se considera una zona de recarga; y recarga horizontal proveniente del piedemonte de la Sierra Norte de Chiapas. Asimismo, las salidas se presentan del mismo modo: por flujo subterráneo hacia la parte norte del acuífero Samaría-Cunduacán y Macuspana, y vertical por evaporación (Comisión Nacional del Agua (CONAGUA), 2024 a).

Administrativamente, el acuífero La Sierra queda comprendido en la Zona decretada de Veda del Río Grijalva, publicada en el Diario Oficial de la Federación el 25 de octubre de 1957. De acuerdo con la Ley Federal de Derechos en Materia de Agua 2024, el acuífero se clasifica como zona de disponibilidad 4, lo que indica que el acuífero tiene una disponibilidad media anual de agua subterránea que permite un mayor margen de extracción, sin poner en riesgo a los ecosistemas ni sobrepasar los límites de sustentabilidad (Comisión Nacional del Agua (CONAGUA), 2024 a).

Este acuífero se encuentra con disponibilidad, de acuerdo con los datos publicados por la CONAGUA. En específico, se tiene que la recarga media anual (R) es de 771.9 [hm³ anuales], mientras que el volumen de extracción de agua subterránea (VEAS) es de 61.69 [hm³ anuales] y la descarga natural comprometida [DNC) es de 132.8 [hm³ anuales], lo que significa que la disponibilidad media anual de agua

subterránea (DMA) es de 577.41 [hm³anuales], lo anterior muestra que la extracción y la descarga natural son menores que la recarga y representan el 74.8% de la misma. Esto lo podemos corroborar con las herramientas de AQUEDUCT, en donde se observa que la zona del acuífero de La Sierra en Tabasco presenta estrés hídrico bajo, de hecho, es el de menor estrés hídrico en el país. Además, podemos observar que el riesgo relacionado con el agua de nuestra zona de estudio es bajo a medio (1-2) (AQUEDUCT a, 2023) (AQUEDUCT b, 2023) (Comisión Nacional del agua (CONAGUA), 2023 b).

12. Materiales y métodos

12.1. Creación de las bases de datos

La consolidación de la base de datos se realizó a partir de información proveniente del INEGI a partir de los mapas de topografía, relieve continental, hidrografía, edafología, fisiografía, geología, hidrología, uso de suelo y vegetación. De las cartas con escala 1:50 000 E14A28, E14A38, E14A29, E14A39, E14A49, E14B31 y E14B41 para el acuífero del Valle de México y de las cartas E15C19, E15C29, E15C39, E15C49, E15B81, E15D11, E15D21, E15D31, E15D41, E15D51, E15D22, E15D32, E15D42, E15D52, E15D23, E15D33, E15D43, E15D53, E15D34 y E15D44 para el acuífero de La Sierra en Tabasco. Además, los polígonos de las zonas de estudio se obtuvieron del geovisor de acuíferos de la CONAGUA (Comisión Nacional del Agua (CONAGUA), s.f.).

Los pasos para la creación de las bases de datos son los mismos para ambas zonas de estudio. En el anexo 2 se encuentra el procedimiento detallado de su creación. A continuación, se presenta un breve resumen del desarrollo.

1. Se subieron las bases de datos a QGis y se seleccionaron las capas de información relevantes que forman parte del área de estudio. Posteriormente, se generó una malla con separación de 660 m entre nodos para asignar las coordenadas geográficas a los datos que se encuentran en capas de información sin coordenadas.
2. A continuación, se unieron las bases de datos que representan la misma información, pero de distintas cartas geográficas en una sola capa. Por ejemplo, se juntaron todas las bases de datos sobre cuerpos de agua, manantiales, análisis químicos, entre otras.
3. Las bases de datos modificadas en QGis se exportaron en archivos csv, para posteriormente ordenarlas con base en sus coordenadas usando un script en *Python*.

4. En particular, los valores de porosidad y permeabilidad de las rocas se obtuvieron a partir de diversos artículos y libros sobre la geología, porosidad y permeabilidad de las rocas en las zonas de estudio.
5. Una vez unidas las capas de información, se eliminaron las filas con coordenadas duplicadas y las columnas sin información.
6. Posteriormente se realizó la codificación de variables. En las tablas del anexo 3 se muestra a detalle el nombre de la capa de información, las columnas que contiene y el método de codificación utilizado.
7. Se analizaron las bases de datos resultantes de este proceso para ambas zonas de estudio usando un script de *Python*. De estas se obtuvieron una matriz de correlación y una matriz de índices de Spearman, (puede consultar información sobre las matrices en el capítulo de análisis exploratorio de datos) con el fin de identificar las variables altamente correlacionadas. Se invita al lector a consultar las matrices en el enlace de github, incluido en el anexo 4.
8. Finalmente, se eliminaron las variables correlacionadas con coeficientes superiores a 0.8, manteniendo una variable representativa para cada asociado.

12.2. Procesamiento de datos

Una vez que las bases de datos estuvieron listas se procesaron usando implementaciones en *Python*. Se generaron dos modelos para la ubicación de las zonas de recarga para cada área de estudio, el primero fue usando *Random Forest* (RF) y el segundo fue utilizando un modelo híbrido de *Random Forest* con *Fuzzy Logic* (RF+FL).

El flujo de trabajo seguido en ambos modelos se muestra en el diagrama de la Figura 7, en el que se encuentran los pasos principales que siguen ambos modelos para analizar la base de datos del acuífero de la Zona Metropolitana de la Ciudad de México. En la figura 8 se muestra una representación visual resumida de los bosques aleatorios y los bosques aleatorios con lógica difusa empleados en el trabajo.

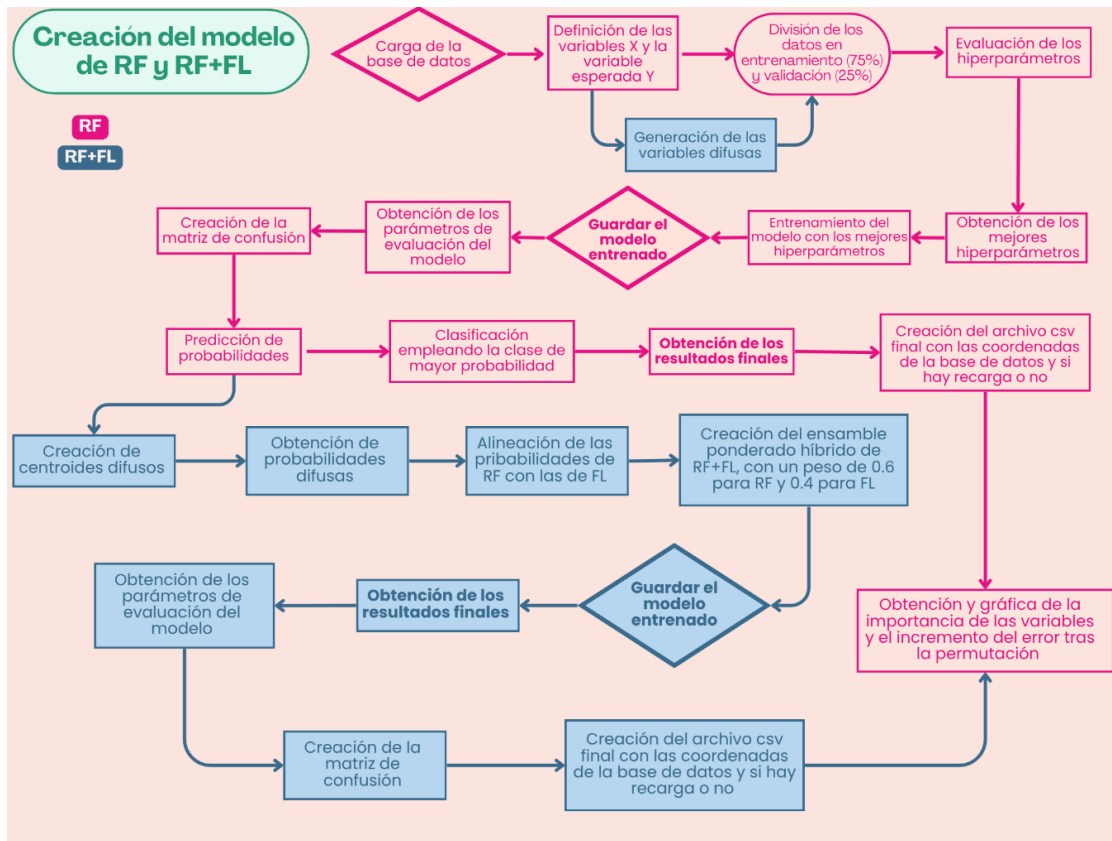


Figura 7 Diagrama de flujo de la creación de los modelos de Random Forest (RF) y de Random Forest híbrido con Fuzzy Logic (RF+FL)

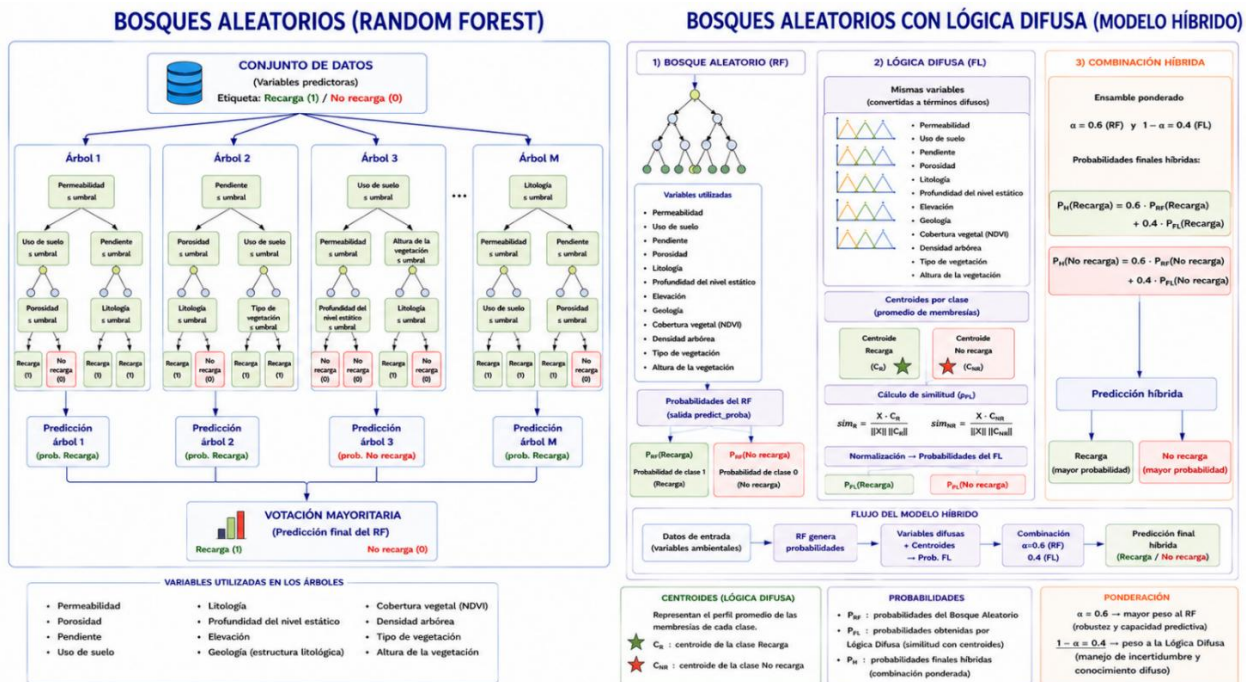


Figura 8 Representación visual resumida de los bosques aleatorios y los bosques aleatorios con lógica difusa empleados en el trabajo (generada en chatGPT).

La descripción de los pasos de ambos modelos se presenta a continuación:

1. Se agregaron las librerías de numpy, pandas, statsmodels.api, pickle, matplotlib.pyplot, RandomForestClassifier, accuracy_score, confusion_matrix, ConfusionMatrixDisplay, classification_report, ColumnTransformer, train_test_split, GridSearchCV, ParameterGrid, permutation_importance, oportuna, skfuzzy, seaborn.
2. Se cargó la base de datos del acuífero de la Zona Metropolitana de la Ciudad de México.
3. Se almacenaron y excluyeron las columnas de coordenadas X, Y de la base de datos.
4. Se definió la variable objetivo Y, la cual contiene las zonas de recarga del acuífero, a partir de ella se generó otra variable objetivo, en la que las zonas de recarga adquirieron el valor 1, y las de no recarga el de 0. Una vez creada la nueva variable respuesta, se descartó la original.
5. Para el modelo híbrido se utilizó skfuzzy para generar las variables difusas y se obtuvieron 3 variables difusas: baja, media y alta.
6. Se dividió la base de datos en entrenamiento (75%) y validación (25%).
7. Se evaluaron los hiperparámetros, se muestra una tabla con los parámetros tentativos que se evaluaron:

Parámetro	Valores tentativos de los parámetros			
n_estimators	100	300	500	
max_features	13	16	20	sqrt
max_depth	None	10	20	30
criterion	Gini	Entropy		

8. Se obtuvieron los mejores hiperparámetros utilizando el Out of bag error (oob_accuracy).
9. Se entrenó el modelo utilizando los mejores hiperparámetros obtenidos en el punto anterior.
10. Se guardó el modelo entrenado de RF usando pickle.
11. Creación de la matriz de confusión con sus parámetros de evaluación del modelo.
12. Se predijeron las probabilidades del modelo de RF para localizar correctamente las zonas de recarga.
13. Para el modelo de RF+FL se obtuvieron los centroides difusos, la función de probabilidades difusas y las probabilidades para localizar correctamente las zonas de recarga del modelo difuso. En esta etapa se aplicaron tres pasos para la lógica difusa:

- a. Las variables predictoras se transformaron mediante la fuzzificación, generando tres funciones de pertenencia (baja, media y alta) para representar gradualmente las condiciones hidrogeológicas.
 - b. Se calcularon los centroides de recarga y no recarga mediante la inferencia difusa. Los centroides representan el patrón promedio de las variables difusas para cada clase.
 - c. Se calculó la similitud de cada observación con ambos centroides y se obtuvieron las probabilidades difusas de pertenencia a cada clase.
14. Posteriormente, para el modelo híbrido, se alinearon las clases de árboles de RF con los de FL, es decir, se alinearon las columnas de probabilidad obtenidas con RF con las de probabilidad ordenadas de FL.
 15. Para el modelo de RF+FL se integraron las probabilidades difusas con las probabilidades obtenidas por RF, mediante un ensamble ponderado con un peso de 0.6 para las probabilidades del algoritmo de RF y un peso de 0.4 para el de FL. Con esto se obtuvieron las probabilidades finales utilizadas para clasificar los datos como zona de recarga o no recarga.
 16. Se guardó el modelo entrenado de RF+FL usando pickle.
 17. Se clasificaron los datos empleando la clase de mayor probabilidad y se obtuvieron los resultados finales del procesamiento.
 18. Para el modelo híbrido se creó la matriz de confusión con sus parámetros de evaluación del modelo (precisión, recall, f1-score, support).
 19. Se generó un archivo csv con las columnas de coordenadas X, Y que se guardaron al inicio y con los resultados del modelo para cada par de coordenadas, es decir, si hay o no recarga en esas coordenadas.
 20. Se obtuvieron y graficaron las 10 variables más importantes utilizadas por el modelo y las 10 variables más importantes tras la permutación, que indica qué tanto aumenta el error del modelo cuando los datos de esa variable se mezclan aleatoriamente. Las barras de error muestran la desviación estándar en múltiples permutaciones.

En el caso del acuífero de La Sierra, Tabasco, no se dividió la base de datos en entrenamiento y predicción, sino que se introdujo al modelo entrenado del acuífero del Valle de México, el cuál calculó las probabilidades de tener recarga en ambos modelos RF y RF+FL, y produjo el modelo final en un archivo csv con las coordenadas del área de estudio y si hay recarga o no.

13. Resultados y análisis de resultados

Se obtuvieron cuatro bases de datos con la predicción de zonas de recarga para las dos zonas de estudio: dos utilizando *Random Forest* y dos usando *Random Forest* híbrido con *Fuzzy Logic*. Para obtener los mejores resultados, se calcularon los

mejores hiperparámetros para cada modelo de aprendizaje computacional y posteriormente se usaron para la predicción de las zonas de recarga del acuífero de la Zona Metropolitana de la Ciudad de México.

En el caso del modelo de *Random Forest*, los mejores hiperparámetros fueron:

Oob_accuracy	0.905
Criterio	Entropía
max_depth	20
max_features	13
n_estimators	300

Tabla 1 Mejores hiperparámetros para el modelo de *Random Forest*

En cambio, los mejores hiperparámetros para el modelo de *Random Forest* con *Fuzzy Logic* fueron:

Oob_accuracy	0.901
Criterio	Gini
max_depth	20
max_features	Sqrt
n_estimators	100

Tabla 2 Mejores hiperparámetros para el modelo de *Random Forest* con *Fuzzy Logic*

Con esos parámetros se obtuvieron los modelos entrenados, los cuáles se evaluaron para validar los resultados. Con el antecedente de su desempeño, buscamos aplicar el modelo entrenado a zonas de estudio sin la identificación previa de las zonas de recarga, es decir, la variable objetivo y, que en este caso es el acuífero de La Sierra en Tabasco.

En las tablas 3 y 4 se observan los resultados de la evaluación de ambos modelos computacionales. Cabe resaltar que los valores de precisión, sensibilidad y F1-score son buenos para ambas clases (1-recarga, 0-no recarga), gracias al balance entre clases que tiene la base de datos, ya que no predomina una clase sobre la otra. Los valores antes mencionados son mayores o iguales a 0.85 en ambos modelos, por lo que éstos tienen un buen desempeño, siendo *Random Forest* con *Fuzzy Logic* ligeramente superior al modelo de *Random Forest*.

<i>Random Forest</i>			
	Precisión	Sensibilidad	F1-score
0	0.87	0.94	0.90
1	0.94	0.85	0.89
Exactitud: 89.90 %			

Tabla 3 Valores de los parámetros de evaluación del modelo de *Random Forest*

Random Forest + Fuzzy Logic			
	Precisión	Sensibilidad	F1-score
0	0.87	0.95	0.91
1	0.95	0.86	0.90
Exactitud: 90.55%			

Tabla 4 Valores de los parámetros de evaluación del modelo de Random Forest con Fuzzy Logic

Aunado a los parámetros para la evaluación de los modelos, se obtuvo la matriz de confusión, la cual muestra cuántos datos son clasificados correctamente y cuántos están mal clasificados. En la figura 9 se muestra la matriz de confusión para RF y en la figura 10 la de RF+FL.

Se observa que en el caso de la matriz del modelo de Random Forest (Figura 9), el modelo clasificó 1165 zonas de recarga correctamente, 1337 zonas de no recarga correctamente, 202 zonas que en realidad son de recarga, pero el modelo clasificó como no recarga y 79 zonas que son de no recarga, pero el modelo clasificó como de recarga, por lo que, en total, el modelo de Random Forest clasificó correctamente 2501 zonas y clasificó 281 zonas incorrectamente.

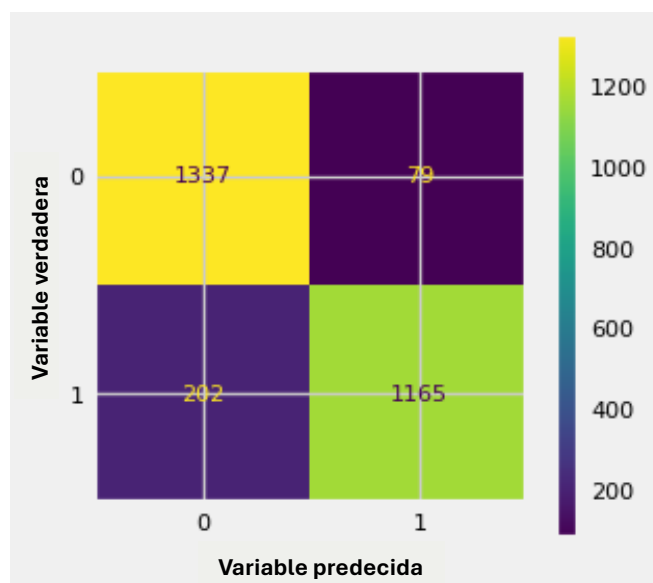


Figura 9 Matriz de confusión del modelo de Random Forest

En cambio, en la matriz del modelo híbrido de Random Forest con Fuzzy Logic (Figura 10), el modelo clasificó 1169 zonas de recarga correctamente, 1351 zonas de no recarga correctamente, 202 zonas que en realidad son de recarga, pero el modelo clasificó como no recarga y 79 zonas que son de no recarga, pero el modelo clasificó como de recarga, por lo que, en total, el modelo híbrido clasificó correctamente 2501 zonas y clasificó 281 zonas incorrectamente.

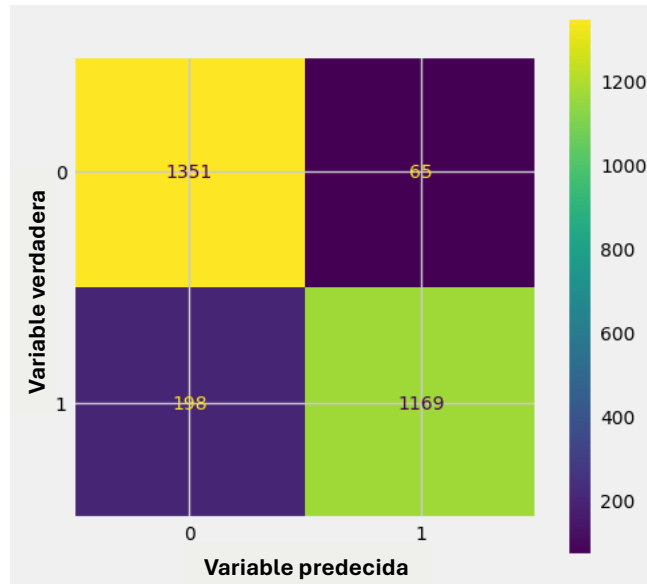
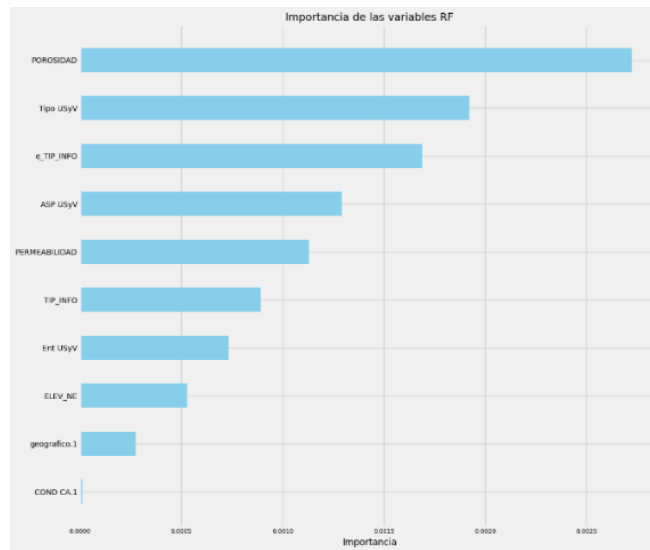


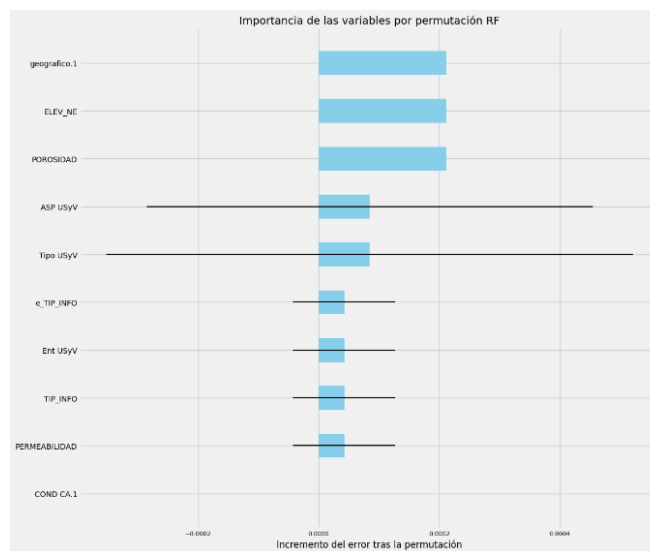
Figura 10 Matriz de confusión del modelo de Random Forest con Fuzzy Logic

Se observa que el modelo de RF+FL predice correctamente más datos, comparado con RF. Los resultados en la validación de los modelos sugieren que su desempeño es muy similar, por lo que podríamos elegir cualquiera de ellos para realizar la clasificación de las zonas de recarga; sin embargo, más adelante mostraremos las diferencias geográficas en la aplicación de ambos modelos.

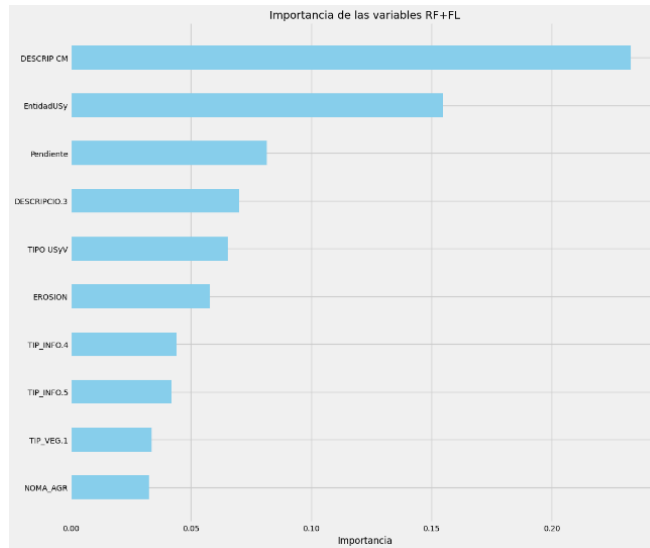
Para comprender cómo funcionan los modelos, a partir de su entrenamiento podemos identificar las variables más relevantes para la clasificación de las zonas de recarga (Gráfica 3 para RF, gráfica 5 para RF+FL) y el incremento del error tras la permutación (Gráfica 4 para RF, gráfica 6 para RF+FL). Con el fin de identificar las variables más importantes y asegurarnos de que estén presentes en la base de datos de La Sierra, para obtener buenos resultados.



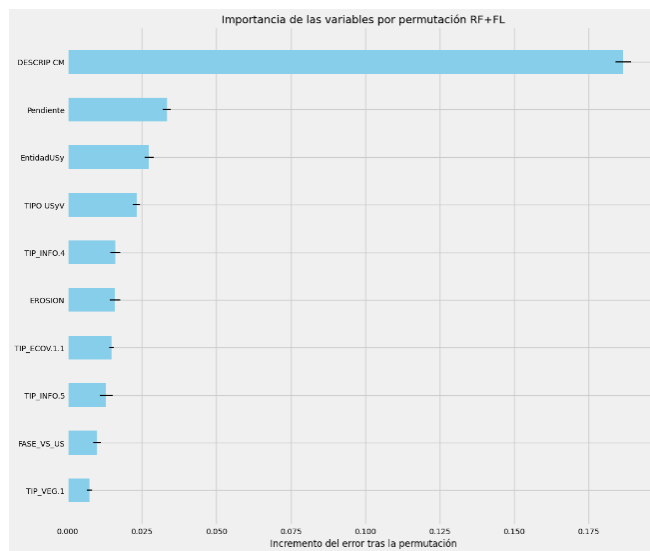
Gráfica 3 Importancia de las variables en el modelo de Random Forest



Gráfica 4 Incremento del error tras la permutación en el modelo de Random Forest

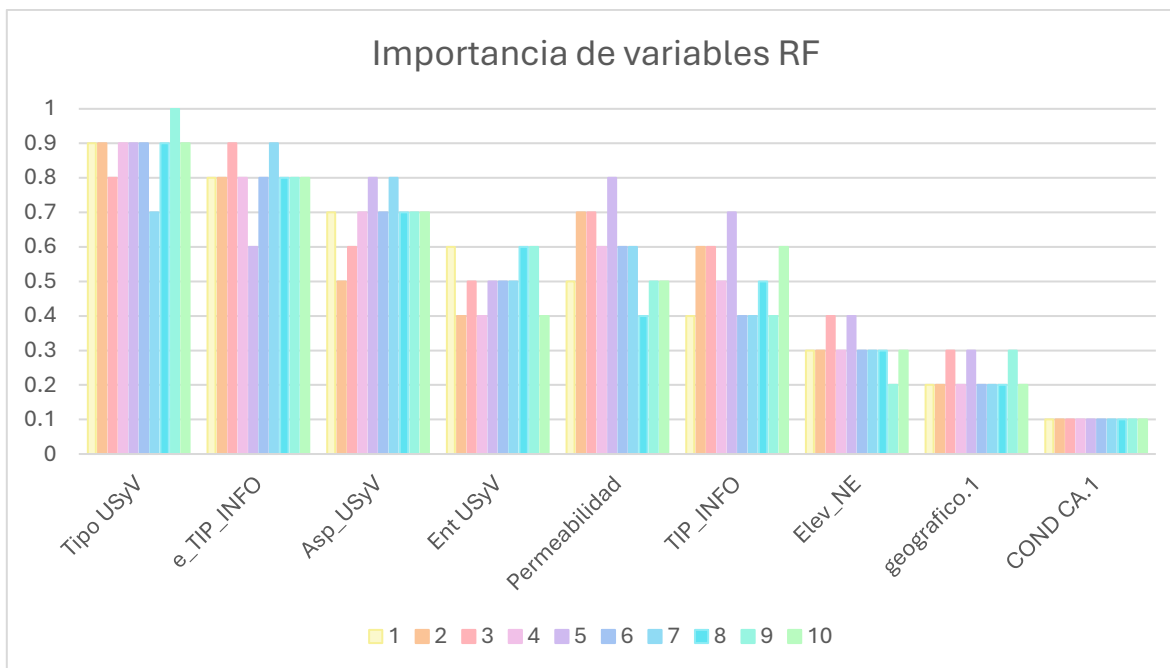


Gráfica 5 Importancia de las variables en el modelo de Random Forest con Fuzzy Logic

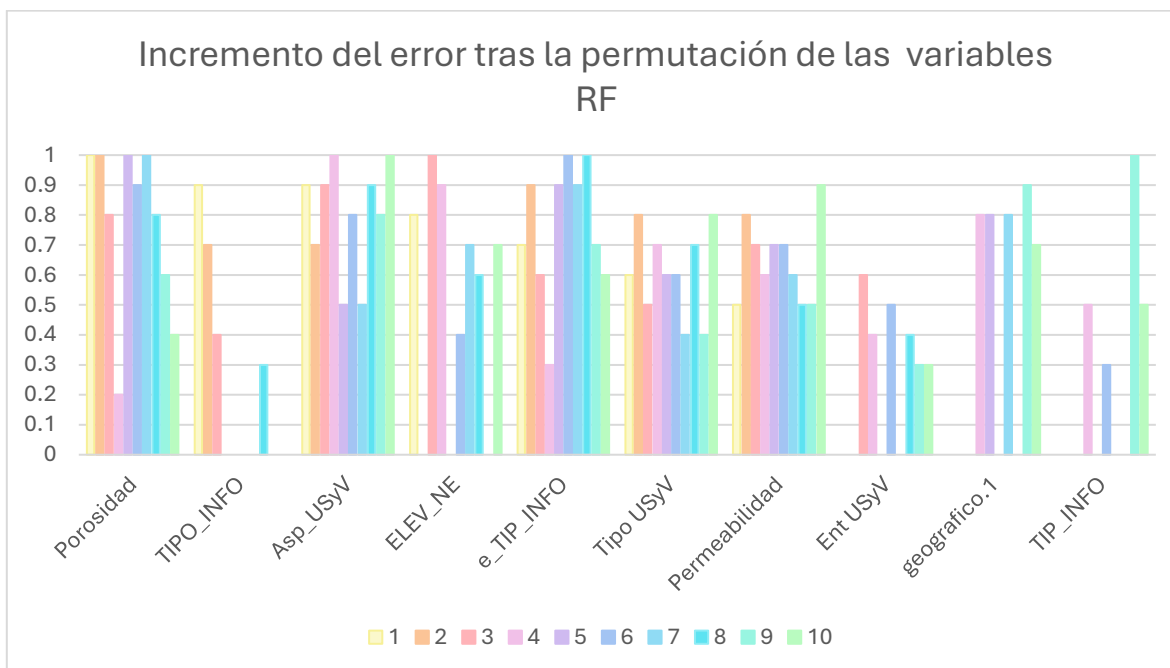


Gráfica 6 Incremento del error tras la permutación en el modelo de Random Forest con Fuzzy Logic

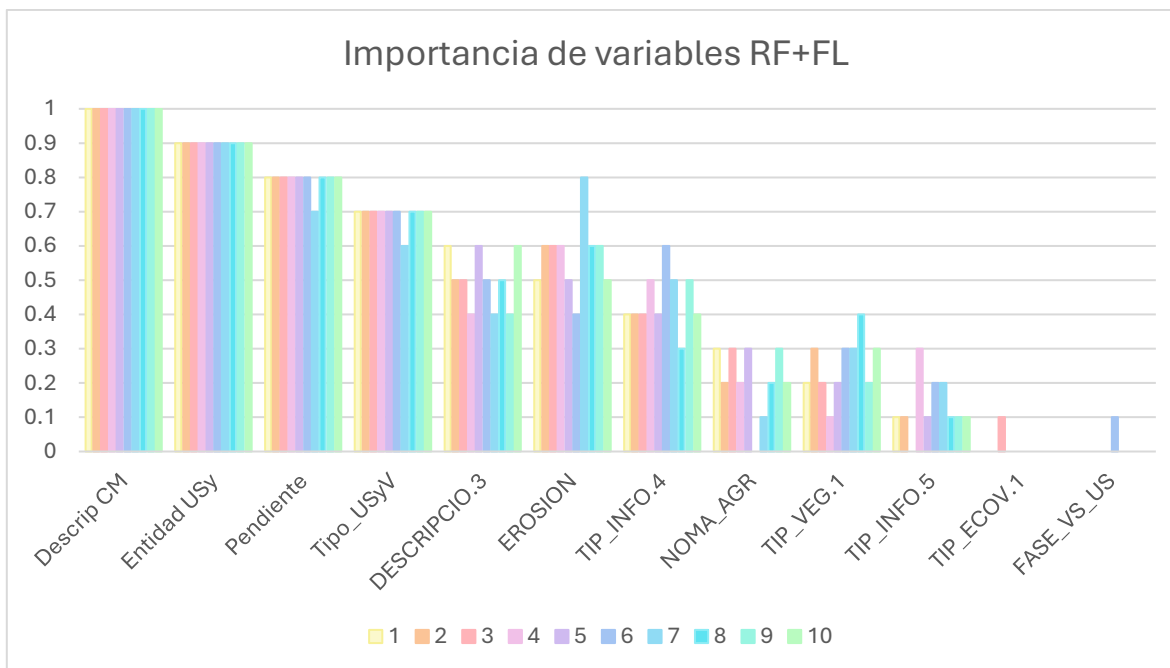
Las variables más importantes cambian entre ambos modelos y cada vez que se ejecuta el código, debido a que se seleccionan aleatoriamente en los árboles de decisión para hacer las predicciones. Por ello, cada código se ejecutó 10 veces, con el fin de identificar a las variables más importantes para cada modelo y a las que aumentan el error tras la permutación. Los resultados obtenidos se observan en las gráficas 7 y 8 para *Random Forest* y en las gráficas 9 y 10 para *Random Forest* con *Fuzzy Logic*. En estas gráficas, el valor en el eje Y representa su importancia en el modelo, siendo 1 la más importante y 0.1 la variable con menor importancia entre las 10 más relevantes.



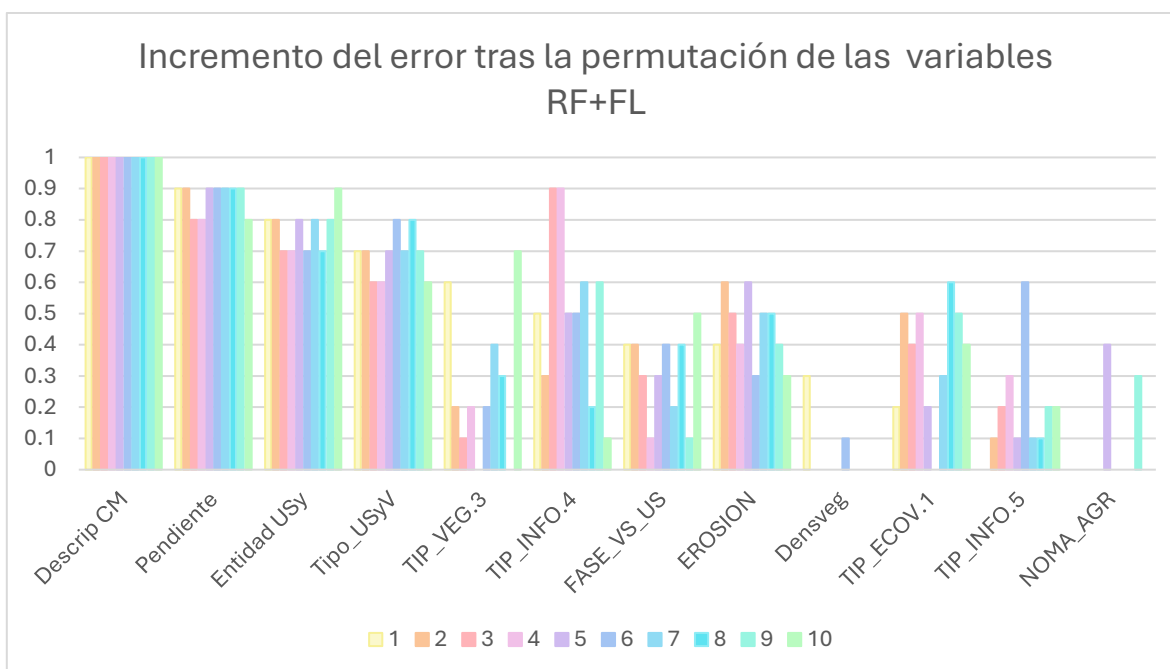
Gráfica 7 Evaluación de la importancia de las variables en cada ejecución del modelo de Random Forest



Gráfica 8 Evaluación del incremento del error tras la permutación de las variables en cada ejecución del modelo de Random Forest



Gráfica 9 Evaluación de la importancia de las variables en cada ejecución del modelo de Random Forest con Fuzzy Logic



Gráfica 10 Evaluación del incremento del error tras la permutación de las variables en cada ejecución del modelo de Random Forest con Fuzzy Logic

La información que tiene cada variable se encuentra en la tabla 5.

Variable	Información que contiene	Ejemplo de la información
Porosidad	Porosidad	15-30, 25-40, 7-15
Tipo USyV	Subtipo de vegetación	Bosque de encino, bosque de pino, bosque de oyamel
e_TIP_INFO	Tipo de información	Agrícola, pecuaria, florística, ecológica
Asp_USyV	Aspecto de la vegetación matorral	Matorral subinermes, sin aspecto matorral
Ent USyV	Tipo de vegetación	Bosque, matorral, pastizal, localidad
Permeabilidad	Permeabilidad	Alta, baja, media-baja
TIP_INFO	Tipo de información	Agrícola, pecuaria, florística, ecológica
Elev_NE	Elevación del nivel estático del agua	Valores del nivel estático del agua en pozos
geografico.1	Canales de agua	Si hay canal o no
COND CA.1	Condición del cuerpo de agua	Intermitente, perenne
Descrip CM	Descripción del material consolidado	Material consolidado, material poco consolidado
Entidad USy	Tipo de vegetación	Bosque, matorral, pastizal, localidad
Pendiente	Pendiente	3,5,15,17

DESCRIPCIO.3	Áreas con pozos	Área con pozos, área libre
EROSION	Erosión	Erosión, sin erosión aparente
TIP_INFO.4	Tipo de información	Agrícola, pecuaria, florística, ecológica
NOMA_AGR	Nomadismo agrícola	Nomadismo agrícola, sin nomadismo agrícola
TIP_VEG.1	Tipo de vegetación	Bosque, matorral, pastizal, localidad
TIP_INFO.5	Tipo de información	Agrícola, pecuaria, florística, ecológica
TIP_ECOV.1	Subtipo de vegetación	Bosque de encino, bosque de pino, vegetación inducida
FASE_VS_US	Tamaño de la vegetación	Arbustiva, arbórea

Tabla 5 Nombre e información de las 10 variables más importantes de los modelos de Random Forest y Random Forest híbrido con Fuzzy Logic

En las gráficas 3, 5, 7 y 9 podemos identificar que las variables más importantes para la predicción de la localización de las zonas de recarga en ambos modelos contienen información sobre el tipo de vegetación, el uso de suelo, la porosidad, permeabilidad, pendiente, erosión, si hay canales, la condición del cuerpo de agua y la descripción del material.

Los resultados de la aplicación del modelo se cargaron a QGis, para comparar los resultados de ambos modelos (RF, RF+FL) del acuífero del Valle de México con los datos oficiales del INEGI de las zonas de recarga que se tienen en la zona de estudio y se obtuvieron los mapas 11 y 12, siendo el primero el generado con *Random Forest*, y el segundo el creado con *Random Forest* híbrido con *Fuzzy Logic*.

Se observa que el modelo generado con *Random Forest* híbrido se aproxima mucho a lo reportado por el INEGI, sin embargo, hay algunas zonas que no considera de recarga, que el INEGI sí, pero en su mayor parte los resultados son muy buenos. En cambio, el modelo generado con *Random Forest*, a pesar de tener buenos parámetros de evaluación del modelo, el modelo final generado es menos realista que el modelo híbrido, ya que no identifica muchas zonas de recarga que el INEGI si toma en cuenta. Esto puede deberse a la diferencia entre la muestra de entrenamiento utilizada para el modelo de *Random Forest* y para el modelo híbrido, ya que pueden utilizar algunos datos de distintas coordenadas. Además, hay diferencia en las variables más importantes para el entrenamiento de ambos modelos, lo cual influye en los resultados, ya que el modelo aprende relaciones estadísticas con base en esas variables de entrada y, por ende, ambos modelos pueden tener métricas similares y las zonas de recarga pueden diferir.

El modelo de *Random Forest* define zonas de recarga en el norte del acuífero, las cuales no se consideran de recarga en la información del INEGI, ni en el modelo híbrido. Sin embargo, esa zona podría presentar recarga urbana inducida, específicamente en donde se encuentran los puntos azules, que abarcan los municipios de Tlanepantla, Ecatepec, Naucalpan y la alcaldía Gustavo A. Madero, en donde hay subsidencia del suelo, la cuál puede generar fracturas y por ende facilitar el proceso de infiltración del agua. Además, se presentan socavones en la zona, los cuales también facilitan la infiltración del agua. Aunado a esto, en la zona se puede presentar infiltración al subsuelo proveniente de fugas en redes de agua potable, alcantarillado y drenaje pluvial, los cuales pueden generar una recarga incidental en el acuífero. Aproximadamente el 40% del agua suministrada en el Estado de México se pierde debido a numerosas fugas en las redes hidráulicas municipales, las cuales son causadas por el deterioro de la red de distribución de agua potable, por los hundimientos de suelo, el tránsito de carga pesada, la calidad y el mantenimiento de las calles y vías, y por la antigüedad. Los municipios con mayor incidencia de fuga son Tlanepantla, Naucalpan y Ecatepec (Gobierno del estado de México, 2024), los cuales coinciden con las zonas de posible recarga obtenidas por el modelo de *Random Forest*.

Por ello, a pesar de que el modelo de *Random Forest* no coincida con todas las zonas de recarga definidas por el INEGI en el norte, sí podría representar una posible zona de recarga inducida para el acuífero.

Por lo tanto, después de analizar los resultados obtenidos con ambos modelos, el que mejor identifica las zonas de recarga es el modelo híbrido de *Random Forest* con *Fuzzy Logic*, a pesar de que es más complicado de programar.

Comparación de las zonas de recarga definidas por el INEGI con las obtenidas con el modelo de Random Forest

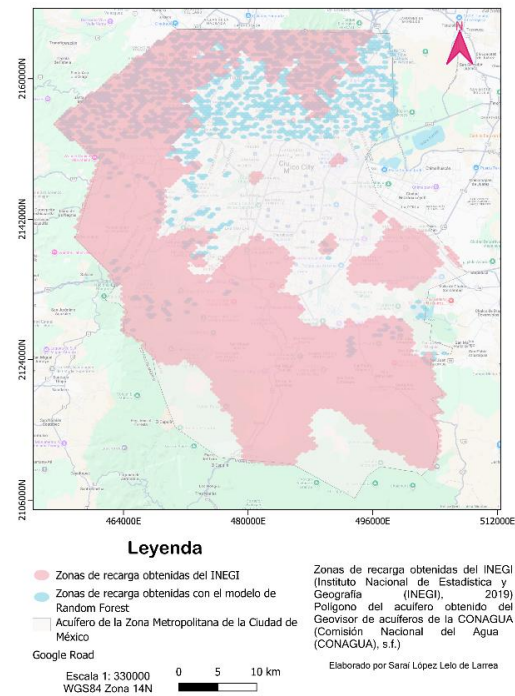


Figura 11 Mapa de comparación de las zonas de recarga del acuífero de la Zona Metropolitana de la Ciudad de México definidas por el INEGI con las obtenidas con el modelo de Random Forest

Comparación de las zonas de recarga definidas por el INEGI con las obtenidas con el modelo de Random Forest con Fuzzy Logic

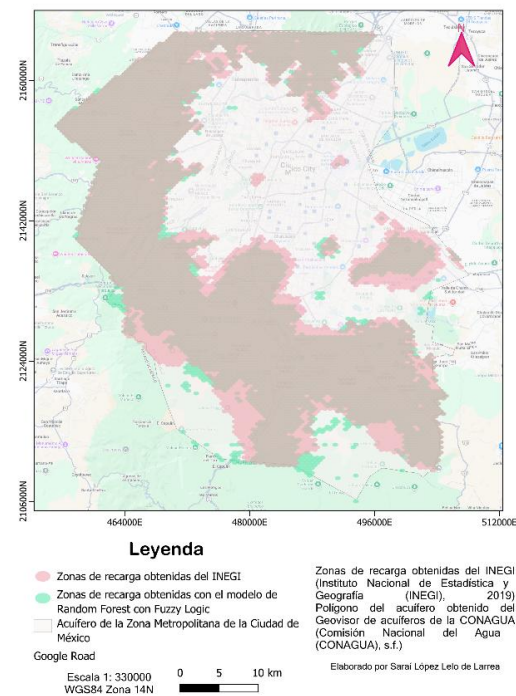


Figura 12 Mapa de comparación de las zonas de recarga del acuífero de la Zona Metropolitana de la Ciudad de México definidas por el INEGI con las obtenidas con el modelo de Random Forest híbrido con Fuzzy Logic

Finalmente, los modelos entrenados con la base de datos del acuífero de la Zona Metropolitana de la Ciudad de México se emplearon para evaluar y generar los modelos de las zonas de recarga del acuífero de La Sierra, en Tabasco. Los resultados de los modelos de RF y RF+FL se observan en el mapa de la figura 13. Los puntos azules son el modelo generado con RF, los rosas son los del modelo de RF+FL y los morados es en donde se traslapan ambos modelos. Debido a que hay poca información disponible sobre las zonas de recarga del acuífero de La Sierra, Tabasco, los resultados obtenidos se pueden corroborar de manera limitada con datos verificados. Los modelos obtenidos con RF y RF+FL coinciden en que las zonas de Teapa, el sur de Tabasco y el piedemonte de la Sierra Norte de Chiapas son de recarga.

Dichas zonas abarcan la mayor parte del acuífero y son congruentes con la geología del lugar, ya que el norte del acuífero estudiado coincide con la zona de recarga del sur de Tabasco, que es en donde se encuentran las máximas elevaciones del estado y altas precipitaciones, y con la zona de recarga de Teapa, mencionada en la síntesis de información geográfica del estado de Tabasco del INEGI (INEGI, 2001). Además, las zonas de recarga de los modelos generados coinciden con las zonas de recarga horizontal provenientes del piedemonte de la sierra norte de Chiapas.

A partir de la comparación del modelo del acuífero de la Zona Metropolitana de la Ciudad de México con los datos obtenidos del INEGI, se puede inferir que el resultado generado con el modelo de Random Forest híbrido es el que da el mejor resultado.

Comparación de las zonas de recarga obtenidas con el modelo de Random Forest con las del modelo de Random Forest con Fuzzy Logic

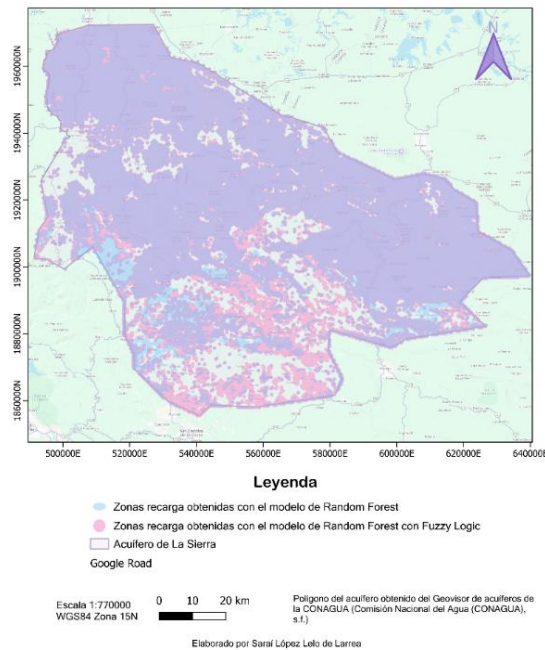


Figura 13 Mapa de comparación de las zonas de recarga del acuífero de La Sierra, Tabasco obtenidas con el modelo de Random Forest con las del modelo de Random Forest con Fuzzy Logic

14. Conclusiones

El diseño de la base de datos, partiendo desde la búsqueda de información, es un paso muy importante en la implementación de modelos de aprendizaje automático para la solución de problemáticas relacionadas con la hidrología. Los métodos propuestos son capaces de identificar las variables con mayor importancia, por lo que la base de datos tiene una gran influencia en el modelo resultante. Por ello, contar con información confiable, completa y actualizada es esencial para obtener resultados apegados a la realidad.

Aunado a lo anterior, es esencial darle el tratamiento adecuado a los datos recopilados, el cual empieza en la selección de las variables que va a utilizar el algoritmo, además, para evitar una alta dimensionalidad, el sobreajuste, el sesgo en árboles demasiado complejos, entre otras desventajas del modelo de Random Forest, es primordial eliminar las variables que no tienen información, las que tienen muy pocos datos, juntar las que cuentan con la misma información y asegurarnos que la variable objetivo esté balanceada, es decir, que la cantidad de datos de cada categoría, en este caso recarga y no recarga, sea muy similar, para no darle preferencia a ninguna durante el entrenamiento del modelo.

Los métodos de aprendizaje computacional que se desarrollaron en este trabajo fueron medianamente costosos computacionalmente, por lo que es importante cuidar la dimensionalidad de los datos, para que el procesamiento sea más rápido y utilice menos recursos computacionales.

Además, para que el modelo pueda analizar los datos, fue esencial codificarlos utilizando *embeddings*, *ordinal encoding* y *standardscaler*, y así convertir los datos alfanuméricos en numéricos.

En los resultados del acuífero de la Zona Metropolitana de la Ciudad de México se observa que ambos modelos obtuvieron un desempeño satisfactorio, sin embargo, el modelo híbrido de *Random Forest* con *Fuzzy Logic* obtuvo mejores indicadores de clasificación y una distribución espacial más consistente con la dinámica hidrogeológica observada en la zona de estudio. Esto podría deberse a que el modelo híbrido es mejor para manejar la incertidumbre de los datos hidrogeológicos e integra las ventajas de ambos modelos, por ejemplo, los bosques aleatorios mejoran la precisión de clasificación del modelo, mientras que la lógica difusa permite una interpretación matizada de los parámetros cualitativos del modelo.

Teniendo esto en mente, se puede inferir que los resultados de la implementación del modelo híbrido entrenado con la base de datos del acuífero de la Zona Metropolitana de la Ciudad de México en la base de datos del acuífero de La Sierra, Tabasco son mejores que los obtenidos por el modelo de *Random Forest*, sin embargo, ambos modelos coinciden con las zonas de recarga definidas en la literatura. Es necesario realizar más estudios y corroborar los resultados obtenidos por los modelos de aprendizaje computacional con información de campo, con el fin de corroborar la validez de ambos modelos.

Identificar las zonas de recarga es esencial para el diseño de estrategias de protección de la cantidad y calidad del agua en los acuíferos; además, es útil para su restauración y manejo sustentable. Sin embargo, debido a la complejidad de los sistemas hidrogeológicos, la delimitación de zonas de recarga continúa siendo un desafío. Por ello, este trabajo busca contribuir con el desarrollo de metodologías innovadoras para la gestión del agua subterránea en el país, en este caso en específico, con el modelo híbrido de *Random Forest* con *Fuzzy Logic*, el cual no ha sido implementado de esta manera en artículos mexicanos.

15. Recomendaciones y perspectiva a futuro

El uso de herramientas de aprendizaje computacional ha tenido un auge en el campo de la hidrología en los últimos años, sin embargo, aún falta mucho desarrollo en el área, la cual tiene un gran potencial para subsanar la falta de información que se tiene en una gran parte del mundo, además de que podría ser una buena

herramienta que contribuya en el modelado de recursos hídricos con el fin de generar planes de gestión de estos recursos, además de una herramienta que ayude a mitigar las problemáticas relacionadas con estos recursos, como la escasez y la contaminación del agua.

Para implementar estos modelos es importante contar con información suficiente, para ello se podrían definir algunas variables esenciales dependiendo de su objetivo. Estas variables se podrían definir con base en las que tengan mayor importancia en distintos estudios del mismo tema, pero en distintas ubicaciones geográficas. Aunado a esto, se podría utilizar información satelital y de percepción remota para complementar la información, monitoreando y analizando los cambios que se tengan en el área de estudio a lo largo del tiempo.

En el futuro sería recomendable que los trabajos que modelen recursos hídricos tomen en cuenta las condiciones actuales en las que se encuentran, considerando las modificaciones y estructuras antropogénicas que se han instalado a lo largo del tiempo y que podrían modificar la dinámica de los recursos hídricos. Algunos ejemplos son la entubación de los ríos, la colocación de asfalto y concreto, la construcción de infraestructura urbana, el secado de cuerpos de agua, la impermeabilización urbana, la deforestación, la recarga incidental, la recarga artificial, etc.

Además, es importante experimentar con distintos modelos de aprendizaje computacional, incluso generar distintos modelos híbridos, con el fin de encontrar los que generen los resultados más apegados a la realidad y que se puedan implementar en zonas de difícil acceso, donde por alguna razón adquirir información es complicado o no se tengan los recursos para hacerlo. Para modelar y comprender de mejor manera los recursos hídricos del mundo y administrarlos, distribuirlos, protegerlos y restaurarlos de la mejor manera posible.

16. Anexos

16.1. Anexo 1

Legislación de recursos hídricos en México

Existen diversas leyes y normas en México enfocadas en los recursos hídricos, las cuales se pueden consultar en la Figura 14. Algunos de los aspectos más importantes con respecto a la situación actual legislativa de los acuíferos, su gestión sustentable y conservación son:

- La Constitución Política de nuestro país faculta al Poder Ejecutivo Federal para establecer medidas regulatorias para mantener el control de la extracción de las aguas nacionales subterráneas mediante la expedición de vedas, reglamentos, reservas y rescates. (Comisión Nacional del Agua (CONAGUA), 2024 c)
- La norma NOM-011-Conagua-2015, que es de las más relevantes, ya que en ella se fundamenta el cálculo de la disponibilidad de agua en cuencas y acuíferos y hace posible el cumplimiento de una de las obligaciones legales de la Conagua. (Comisión Nacional del Agua (CONAGUA), 2024 c)
- Las normas expedidas por la Conagua que establecen las disposiciones, especificaciones y métodos de prueba que garantizan que los productos y servicios asociados al sector hídrico cumplan, preservando el agua en cantidad y calidad. (Comisión Nacional del Agua (CONAGUA), 2024 c)

No.	Grupo: CONAGUA -
1	NOM-001-CONAGUA-2011 - Sistemas de agua potable, toma domiciliar y alcantarillado sanitario-Hermeticidad-Especificaciones y métodos de prueba
2	NOM-003-CONAGUA-1996 - Requisitos para construcción de pozos para prevención de contaminación de acuíferos.
3	NOM-004-CONAGUA-1996 - Requisitos para la protección de acuíferos durante mantenimiento y rehabilitación de pozos de agua y cierre de pozos en general.
4	NOM-006-CONAGUA-1997 - Especificaciones y métodos de prueba para fosas sépticas prefabricadas.
5	NOM-008-CONAGUA-1998 - Especificaciones y métodos de prueba para regaderas.
6	NOM-009-CONAGUA-2001 - Especificaciones y métodos de prueba para inodoros.
7	NOM-010-CONAGUA-2000 - Especificaciones y métodos de prueba para válvulas de inodoros.
8	NOM-014-CONAGUA-2003 - Requisitos para la recarga artificial de acuíferos con agua residual tratada.
9	NOM-011-CONAGUA-2015 - Conservación del recurso agua. Especificaciones y método para determinar la disponibilidad media anual de las aguas nacionales.
10	NOM-015-CONAGUA-2007 - Características y especificaciones de las obras y del agua para infiltración artificial a acuíferos.
11	NMX-AA-175-SCFI-2015 Operación segura de presas. Parte 1 y 2.- Análisis de riesgo, clasificación de presas e inspecciones de seguridad
12	NOM-002-CONAGUA-2021 Aparatos y accesorios de uso sanitario.
No.	Grupo: SEMARNAT
1	NOM-060-SEMARNAT-1994 - Especificaciones para mitigar los efectos adversos ocasionados en los suelos y cuerpos de agua por el aprovechamiento forestal.
2	NOM-001-SEMARNAT-1996 - Límites máximos permisibles de contaminantes en las descargas de aguas residuales en aguas y bienes nacionales.
3	NOM-002-SEMARNAT-1996 - Límites máximos permisibles de contaminantes en las descargas de aguas residuales a los sistemas de alcantarillado urbano o municipal.
4	NOM-003-SEMARNAT-1997 - Límites máximos permisibles de contaminantes para las aguas residuales tratadas que se reusen en servicios al público.
5	NOM-004-SEMARNAT-2002 - Protección ambiental. Lodos y biosólidos. Especificaciones y límites máximos permisibles de contaminantes para su aprovechamiento y disposición final.
6	NOM-022-SEMARNAT-2003 - Preservación, conservación, aprovechamiento sustentable y restauración de los humedales costeros en zonas de manglar.
7	NOM-150-SEMARNAT-2017 - Establece las especificaciones técnicas de protección ambiental que deben observarse en las actividades de construcción y evaluación preliminar de pozos geotérmicos para exploración, ubicados en zonas agrícolas, ganaderas y eriales, fuera de áreas naturales protegidas y terrenos forestales.
8	NOM-001-SEMARNAT-2021. Que establece los límites permisibles de contaminantes en las descargas de aguas residuales en cuerpos receptores propiedad de la nación.
No.	Grupo: Energía
1	NOM-006-ENER-2015 - Eficiencia energética electromecánica en sistemas de bombeo para pozo profundo en operación. Límites y método de prueba.
No.	Grupo: Salud
1	NOM-117-SSA1-1994 - Método de prueba para la determinación de cadmio, arsénico, plomo, estaño, cobre, hierro, zinc y mercurio en alimentos, agua potable y agua purificada por espectrometría de absorción atómica.
2	NOM-127-SSA1-1994 - Salud ambiental. Agua para uso y consumo humano. Límites permisibles de calidad y tratamientos a que debe someterse el agua para su potabilización.
3	NOM-179-SSA1-1998 - Vigilancia y evaluación del control de calidad del agua potable en redes.
4	NOM-201-SSA1-2002 - Productos y servicios. Agua y hielo para consumo humano, envasados y a granel. Especificaciones sanitarias.
5	NOM-230-SSA1-2002 - Requisitos sanitarios para manejo del agua en las redes de agua potable.
6	NOM-244-SSA1-2008 - Equipos y sustancias germicidas para tratamiento doméstico de agua. Requisitos sanitarios
7	NOM-127-SSA1-2021. Agua para uso y consumo humano. Límites permisibles de la calidad del agua.
No.	Grupo: Normas Mexicanas -
1	NMX-AA-120-SCFI-2006 - Requisitos y especificaciones de sustentabilidad de calidad de playas.
2	NMX-AA-147-SCFI-2008 - Metodología de evaluación de las tarifas de agua potable, drenaje y saneamiento.
3	NMX-AA-148-SCFI-2008 - Metodología para evaluar la calidad de los servicios de agua potable, drenaje y saneamiento. Directrices para la evaluación y la mejora del servicio a los usuarios.
4	NMX-AA-149/1-SCFI-2008 - Metodología para evaluar la eficiencia de los prestadores de servicios de agua potable, drenaje y saneamiento. Directrices para la prestación y evaluación de los servicios de agua residual.
5	NMX-AA-149/2-SCFI-2008 - Metodología para evaluar la eficiencia de los prestadores de servicios de agua potable, drenaje y saneamiento. Directrices para la prestación y evaluación de los servicios de agua potable.
6	NMX-AA-017-SCFI-2021. Análisis de agua-medición de color verdadero en aguas naturales, residuales, residuales tratadas y marinas-mediante coeficientes de absorción espectral-método de prueba (cancela a la NMX-AA-017-1980)
7	NMX-AA-028-SCFI-2021. Análisis de agua-medición de demanda bioquímica de oxígeno (DBO5) en aguas naturales, residuales y residuales tratadas-dilución y método de siembra-método de prueba (cancela a la NMX-AA-028-SCFI-2001).
8	NMX-AA-131/1-SCFI-2021. Análisis de agua-medición de elementos por espectrometría de plasma acoplado inductivamente (ICP), en aguas naturales, potables, residuales y residuales tratadas-método de prueba. Parte 1-medición de elementos por espectrometría de emisión óptica con plasma acoplado inductivamente (ICP-OES).
9	NMX-AA-131/2-SCFI-2021. Análisis de agua-medición de elementos por espectrometría de plasma acoplado inductivamente (ICP), en aguas naturales, potables, residuales y residuales tratadas-método de prueba-parte 2-amplificación de la espectrometría de masas con plasma acoplado inductivamente (ICP-MS)-medición de 73 elementos (cancela al PROY-NMX-AA-131/2-SCFI-2013 publicado en el diario oficial de la federación el 25 de julio de 2014).
10	NMX-AA-187-SCFI-2021. Análisis de agua-medición de carbono orgánico total en aguas naturales, salinas, residuales y residuales tratadas.

Figura 14 Leyes enfocadas en los recursos hídricos de México (Comisión Nacional del Agua (CONAGUA), 2024 c)

Para llevar a cabo las funciones que le han sido conferidas, la CONAGUA trabaja en conjunto con diversas instancias en el ámbito federal, estatal y municipal; asociaciones de usuarios y empresas; instituciones del sector privado y social, así como organizaciones internacionales con las que CONAGUA tiene coordinación para cumplir los objetivos del Programa Hídrico Nacional (Figura 15) (Comisión Nacional del Agua (CONAGUA), 2024 c).

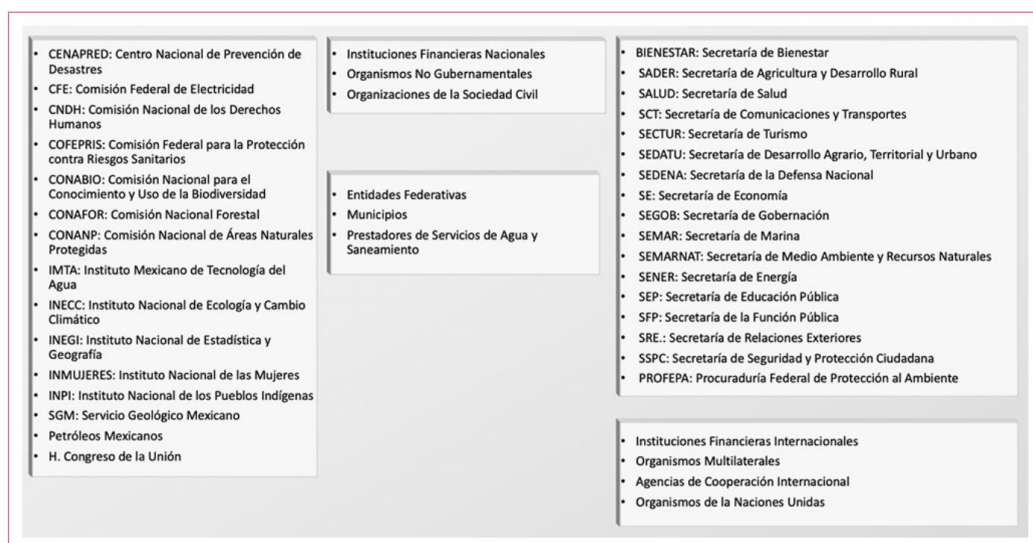


Figura 15 Principales instituciones, entidades y dependencias que trabajan en coordinación con la CONAGUA (Comisión Nacional del Agua (CONAGUA), 2024)

16.2. Anexo 2

Pasos para la creación de la base de datos

Las bases de datos se obtuvieron del INEGI, de los mapas de topografía, relieve continental, hidrografía, climatología, edafología, fisiografía, geología, hidrología, y uso de suelo y vegetación. De las cartas E14A28, E14A38, E14A29, E14A39, E14A49, E14B31 y E14B41 para el acuífero del Valle de México, y de las cartas E15C19, E15C29, E15C39, E15C49, E15B81, E15D11, E15D21, E15D31, E15D41, E15D51, E15D22, E15D32, E15D42, E15D52, E15D23, E15D33, E15D43, E15D53, E15D34 y E15D44 para el acuífero de la Sierra, Tabasco. Además, del geovisor de acuíferos de la CONAGUA, se obtuvieron los polígonos de las áreas de estudio.

Posteriormente, se subieron las bases de datos a QGIS, y se seleccionaron las capas de información útiles para este trabajo, como la geología, uso de suelo, el modelo digital de elevación, los cuerpos de agua, pozos, estudios químicos del agua subterránea, entre otros.

Una vez que se filtró la información obtenida del INEGI, y se conservó la relevante para el trabajo, se cortaron las capas de información en QGIS con la herramienta de corte de capas vectoriales, usando como capa de corte el polígono del acuífero del Valle de México. A continuación, se revisaron todas las capas de información cortadas, y se eliminaron las que en su tabla de atributos ya no tenían información.

Después, se generó una malla triangular con separación de 660 m entre nodos para asignar las coordenadas geográficas a las bases de datos que no tenían esta información. Además, se implementó esta malla con el fin de unificar la resolución espacial de las variables, ya que la escala de las capas de información obtenidas del INEGI tenía distintas resoluciones. La información de topografía, hidrografía, edafología y uso de suelo y vegetación tiene una escala 1:50 000, en cambio, la de relieve continental, geología e hidrología tiene una escala 1:250 000. En la figura 16 se muestra una representación visual de la malla mencionada y de algunas variables utilizadas en la creación de la base de datos del acuífero de La Sierra, Tabasco.

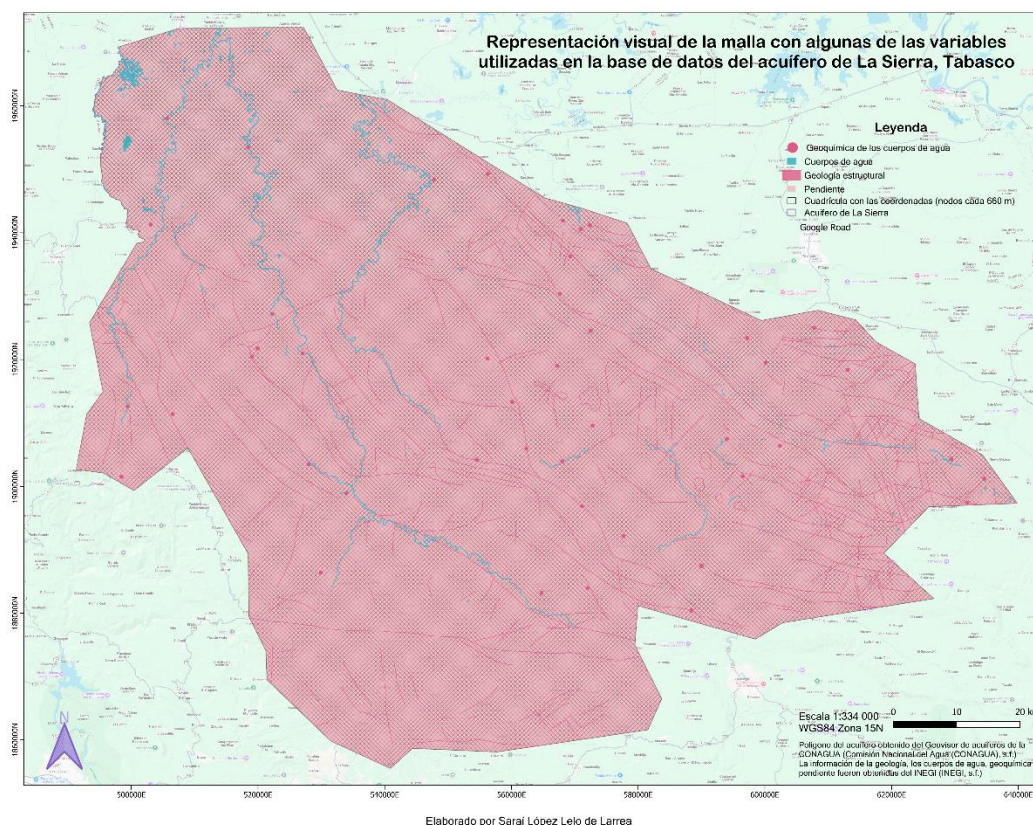


Figura 16 Mapa con la malla generada en QGis con una separación de 660 m entre nodos, y algunas de las variables utilizadas en la base de datos del acuífero de La Sierra, Tabasco.

Posteriormente, se cambiaron los nombres de algunas columnas de las capas de información, para poder identificar de qué información se trata a la hora de unir las bases de datos.

Para generar la base de datos de las pendientes, se utilizó el modelo digital de elevaciones, y se usó la herramienta de pendientes que se encuentra en la caja de herramientas, en análisis de terreno ráster. Después, las bases de datos modificadas en QGis se exportaron en archivos csv para juntarlas usando código

de Python en Google Colab. Debido al gran tamaño de la base de datos, se obtuvieron 3 archivos con información.

Cabe mencionar que en la base de datos de geología se añadieron dos columnas, una de porosidad y otra de permeabilidad, cuyos valores se obtuvieron a partir de diversos artículos y libros sobre la geología, porosidad, y permeabilidad de las rocas en el Valle de México, y algunos valores teóricos generales de estos parámetros. Dicho proceso se realizó en la capa csv exportada de QGis, y a continuación se unió al resto de las capas.

Debido al gran tamaño de las bases de datos, se eliminaron columnas innecesarias con código de Python en Google Colab, para realizar la correlación entre variables, como el id, área, perímetro, columnas sin información, entre otras.

Posteriormente se realizó la codificación de variables categóricas

16.3. Anexo 3

16.3.1. Índice de los archivos

A continuación, se presenta la información de los archivos modificados en QGis para la creación de la base de datos, los nombres de los archivos son los que se pueden encontrar en el enlace de github que se encuentra en el anexo 4.

16.3.1.1. Acuífero de la Zona Metropolitana de la Ciudad de México

a2606: Condición del acuífero, coordenadas UTM X, coordenadas UTM Y.

e1402pg_coord: Tipo de roca, nombre de roca, Era, sistema, porosidad, permeabilidad.

actividades_forestales_pec_coord: Actividad forestal o pecuaria, tipo de actividad.

acueducto_l_cort1_fin_copiar, acueducto_l_cort2_fin_copiar,
acueducto_l_cort3_coord, acueducto_l_cort4_copiar_copiar,
acueducto_l_cort5_copiar_copiar, acueducto50_l_cort7_coord,
acueducto50_l_cort8_coord: Tipo de estructura, relación en el suelo.

ae1402v_cort_fin_copiar: Forma de captación, tipo de obra, valores de distintas medidas de un análisis del agua: Ca, Mg, Na, K, CaCO₃, RAS, pH, Ce, SO₄, HCO₃, NO₃, CO₃, Cl, sólidos, calidad del agua, agresividad del agua, nivel estático del agua, nivel dinámico del agua (nivel del agua cuando el pozo está bombeando), caudal (cantidad del agua que produce el pozo), temperatura, uso del agua.

agricultura_y_vegetacion_coord: Uso de suelo, Tipo de uso de suelo, fisionomía, vegetación secundaria, erosión, cultivo primario, cultivo secundario, cultivo terciario.

areacultiv50_a_cort_coord, areacultiv50_a_cort2_coord: Uso de suelo.

c2606_coord: Elevación del nivel estático del agua, descripción.

canal_a_cort1_fin_copiar, canal_a_cort2_fin_copiar, canal_a_cort3_fin_copiar,
canal_a_cort4_coord: Término general, condición.

canal_l_cort1_fin_copiar, canal_l_cort2_fin_copiar, canal_l_cort3_coord,
canal50_l_cort2_coord: Término general, condición.

clave_de_cultivos_cort_copiar_c: Entidad, nombre común 1, nombre común 2,
nombre común 3.

clave_de_especie_vegetales_cort: Entidad, nombre científico 1, nombre científico
2, nombre científico 3, nombre científico 4, nombre científico 5.

corriente_ag_a_cort_fin_copiar, corriente_ag_l_cort1_fin_copiar,
corriente_ag_l_cort2_coord, corriente_ag_l_cort3_fin_copiar,
corriente_ag_l_cort4_fin_copiar, corriente_ag50_l_cort1_coord,
corriente_ag50_l_cort2_coord, corriente_ag50_l_cort3_coord: Tipo de formación
natural, condición del cuerpo de agua, tipo de cuerpo de agua.

corriente_de_p_cort1_fin_copiar, corriente_de_p_cort2_fin_copiar,
corriente_de_p_cort3_fin_copiar, corriente_de_p_cort4_copiar_cop,
corriente_de50_p_cort2_fin_copiar, corriente_de50_p_cort3_copiar_c: Tipo de
corriente, tipo de rasgo geológico.

cuerpo_agua_a_cort1_fin_copiar, cuerpo_agua_a_cort2_fin_copiar,
cuerpo_agua_a_cort3_fin_copiar, cuerpo_agua_a_cort4_fin_copiar,
cuerpo_agua_a_cort5_coord, cuerpo_agua50_a_cort3_coord: Condición del
cuerpo de agua, tipo de rasgo geológico, tipo de cuerpo de agua.

e1402_usv250s5a_coord: Tipo de información, tipo de ecosistema, tipo de uso de
suelo.

e1402_usv250s5d_coord: Tipo de información, tipo de ecosistema, densidad
vegetal, cobertura de árboles, tipo de uso de suelo.

e1402_usv250s5g_coord: Tipo de información, tipo de ecosistema, edad del
ecosistema, tipo de planta, tipo de cultivo 1, tipo de cultivo 2, otros.

e1402_usv250s5h_coord: Tipo de información, tipo de ecosistema, densidad
vegetal, altura de vegetación, otros.

e1402_usv250s5v_coord: Tipo de información, tipo de ecosistema, tipo de
vegetación, densidad vegetal, fase, otros.

e1402e_coord: Clave del suelo, nombre del suelo 1, nombre del subsuelo 1, nombre
del suelo 2, nombre del subsuelo 2, nombre del suelo 3, nombre del subsuelo 3,

clase textural, fase física del suelo, fase química del suelo, entidad del suelo, clave de interpretación.

e1402lg_coord: Longitud de fractura, tipo de estructura geológica, movimiento horizontal, movimiento vertical, inclinación, buzamiento.

e1402ptg_fin_copiar: Código del objeto, entidad geológica, tipo, inclinación, rumbo, buzamiento, dirección de la formación geológica, mineral principal, mineral secundario, tercer mineral, uso del material geológico.

Pendiente: Valor de la pendiente.

e1402u1v_coord: Tipo, fisionomía, vegetación secundaria, erosión, cultivo primario, cultivo secundario, cultivo terciario.

e2606_coord: Descripción.

estanque_a_cort1_fin_copiar, estanque_a_cort4_fin_copiar,
estanque_a_cort5_fin_copiar: Tipo de estanque, tipo de estructura.

f2606_fin_copiar: Descripción de la capa, azimutal de la dirección del flujo de agua.

fe1402v_coord: Longitud de la dirección de flujo, descripción.

h2606_coord: Profundidad del nivel estático del agua, descripción.

manantial_p_cort1_copiar_copiar, manantial_p_cort2_fin_copiar,
manantial_p_cort4_1_copiar_copi, manantial_p_cort5_copiar_copiar: Tipo de rasgo geológico, tipo de rasgo hidrográfico.

manantial50_p_cort8_copiar_copi: Tipo de rasgo hidrográfico.

ne1402v_cort_copiar_copiar: Tipo de obra.

o2606_cort_copiar_copiar: Tipo de obra, caudal (cantidad de agua que se produce en el pozo), nivel estático del agua, nivel dinámico del agua (pozo está bombeando), uso del agua, observaciones.

p2606_cort_copiar_copiar: Análisis fisicoquímico del agua: Ca, Mg, K, Dureza como CaCO_3 , Relación de Adsorción de Sodio (SAR), pH medido en campo, pH medido en laboratorio, conductividad eléctrica medida en campo, conductividad eléctrica medida en laboratorio, SO_4 , HCO_3 , CO_3 , Cl, NO_3 , sólidos totales disueltos medidos en campo, sólidos totales disueltos medidos en laboratorio, clasificación de agua para riego, tipo de agresividad química del agua, caudal del pozo, nivel estático del agua, nivel dinámico del agua, uso del agua, temperatura, termalismo, composición hidroquímica, salinidad, familia hidroquímica del agua, observaciones.

pe1402v_coord: Descripción.

pozo_p_cort1_fin_copiar, pozo_p_cort2_fin_copiar, pozo_p_cort3_fin_copiar,
pozo_p_cort4_fin_copiar: Condición del pozo, tipo de pozo, tipo de estructura.

referencia_g_p_cort1_copiar_cop, referencia50_p_cort1_copiar_cop: Rasgo geológico.

t2606_cort_copiar_copiar: Caudal, nivel estático del agua, nivel dinámico del agua, nivel máximo del agua estática, caudal de agua subterránea, transmisividad del acuífero, coeficiente de almacenamiento, espesor del acuífero, conductividad hidráulica, comportamiento hidráulico del acuífero, método, observaciones.

terreno_suje_a_cort1_copiar_cop, terreno_suje_a_cort2_coord,
terreno_suje_a_cort3_copiar_cop: Rasgo geológico, tipo.

u2606_coord: Descripción del área del acuífero.

ue1402v_coord: Descripción del material.

us250s3a_coord: Tipo de información, tipo de ecosistema, otros.

us250s3c_cort_copiar_copiar: Tipo de información, cultivo 1, cultivo 2, cultivo 3.

us250s3d_4_coord: Tipo de información, tipo de vegetación, descripción vegetal, cobertura de árboles, otros.

us250s3e_fin_copiar: Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.

us250s3g_coord: Tipo de información, tipo de cultivo 1, tipo de cultivo 2, otros.

us250s3h_coord: Tipo de información, descripción vegetal, altura, otros.

us250s3i_fin_copiar: Tipo de información, tipo de ecosistema, tipo de vegetación.

us250s3n_coord: Tipo de información, nomadismo agrícola, otros.

us250s3r_coord: Tipo de información, erosión, otros.

us250s3v_coord: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de vegetación, fase de vegetación, otros.

us250s4a_coord: Tipo de información, tipo de ecosistema, aspecto, otros.

us250s4d_coord: Tipo de información, tipo de ecosistema, descripción de la vegetación, cobertura de árboles, otros.

us250s4e_fin_copiar: Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.

us250s4g_coord: Tipo de información, tipo de cultivo 1, tipo de cultivo 2, otros.

us250s4h_coord: Tipo de información, tipo de ecosistema, descripción vegetal, altura arbórea, otros.

us250s4i_fin_copiar: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de la vegetación, otros.

us250s4v_3_coord: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de la vegetación, fase, otros.

vegeta_dens50_a_cort1_coord, vegeta_dens50_a_cort2_coord,
vegeta_dens50_a_cort3_coord: Densidad vegetal.

z2606_coord: Nombre del acuífero, zonas de recarga y extracción del acuífero.

16.3.1.2. Acuífero de La Sierra, Tabasco

actividades_forestales_pec copiar: Entidad, tipo de actividad.

acueducto50_l_combinada: Tipo de estructura, relación con el suelo.

ae1509v copiar: Tipo de obra, valores de distintas medidas de un análisis del agua: Ca, Mg, Na, K, CaCO₃, RAS, pH, Ce, SO₄, HCO₃, NO₃, CO₃, Cl, sólidos, calidad del agua, agresividad del agua, diámetro de la tubería de descarga, nivel estático del agua, nivel dinámico del agua, caudal, temperatura, uso, observaciones.

agricultura_nomada: Entidad.

agricultura_y_vegetacion: Entidad, tipo de agricultura, fisionomía, vegetación secundaria, erosión, cultivo primario, cultivo secundario, cultivo terciario.

area_cultiv50_a_combinada: Entidad.

area_ver_ur50_a copiar: Entidad.

bordo_l_combinada: Tipo de estructura.

canal50_l_combinada: Tipo de cuerpo de agua.

clave_de_cultivos: Entidad, nombre común 1, nombre común 2, nombre común 3.

clave_de_especie_vegetales: Entidad, nombre científico 1, nombre científico 2, nombre científico 3, nombre científico 4, nombre científico 5.

corriente_ag_a_combinada: Tipo de rasgo geológico, tipo de cuerpo de agua, condición del cuerpo de agua.

corriente_ag_l_combinada, corriente_ag_l_combinada2: Tipo de rasgo geológico, tipo de cuerpo de agua, condición del cuerpo de agua.

corriente_ag50_l_combinada: Tipo de cuerpo de agua, condición del cuerpo de agua.

corriente_de_p_combinada: Tipo de rasgo geológico, tipo de cuerpo de agua.

corriente_de50_p_combinada: Tipo de cuerpo de agua.

cuerpo_agua_a_combinada: Tipo de rasgo geológico, tipo de cuerpo de agua, condición del cuerpo de agua.

eu1v: Entidad, tipo de vegetación, fisonomía, vegetación secundaria, erosión, cultivo primario, cultivo secundario, cultivo terciario.

e1_usv250s5a: Tipo de información, aspecto, otros.

e1_usv250s5d: Tipo de información, tipo de ecosistema, descripción de la vegetación, cobertura arbórea, otros.

e1_usv250s5e: Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.

e1_usv250s5f: Tipo de información.

e1_usv250s5g: Tipo de información, clasificación de la información, descripción de la información, tipo de plantación, tipo de cultivo 1, tipo de cultivo 2, otros.

e1_usv250s5h: Tipo de información, tipo de ecosistema, descripción vegetal, altura, otros.

e1_usv250s5l: Tipo de información, clave.

e1_usv250s5n: Tipo de información, nomadismo agrícola, otros.

e1_usv250s5o: Tipo de información, clave.

e1_usv250s5v: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de la vegetación, otros.

e15_combinada: Clave, nombre de suelo 1, nombre del subsuelo 1, nombre de suelo 2, nombre del subsuelo 2, nombre de suelo 3, nombre del subsuelo 3, clase textural, fase física, fase química, entidad.

e15d22_toponimo50_p_utm copiar: Tipo de información, tipo de rasgo geológico.

e1508_usv250s5c copiar: Tipo de información, tipo de cultivo 1, tipo de cultivo 2, tipo de cultivo 3.

e1508_usv250s5i copiar: Tipo de información, clave.

e1508_usv250s5p copiar: Tipo de información, clave.

E15080911_combinada: Clave, grupo 1, calificador del grupo 1, calificador primario del grupo 1, grupo 2, calificador del grupo 2, calificador primario del grupo 2, grupo 3, calificador del grupo 3, calificador primario del grupo 3, clase textural, frudica.

ee_combinada: Coeficiente de escurrimiento.

est_geo_grav_p_combinada: Gravedad, altura ortométrica.

estanque50_a_combinada: Tipo de estructura, tipo de estanque.

fe1508v copiar: Tipo de información (flujo de agua).

geologia_area: Tipo de roca, porosidad, permeabilidad.

geologia_linea: Longitud, tipo de estructura geológica, subtipo de estructura geológica, movimiento horizontal, movimiento vertical, inclinación, buzamiento.

geologia_punto: Entidad, inclinación, rumbo, buzamiento, tipo de mineralización asociada.

manantial_p_combinada: Tipo de rasgo geológico, tipo de cuerpo de agua.

pe_combinada: Tipo de cuerpo de agua, valores de distintas medidas de un análisis del agua: Ca, Mg, Na, K, CaCO₃, RAS, pH, Ce, SO₄, HCO₃, NO₃, CO₃, Cl, sólidos, calidad del agua, agresividad del agua, temperatura, uso, descripción.

pev_combinada: Descripción de uso de suelo.

pozo_p_combinada: Tipo de estructura, tipo de pozo.

referencia_g_p_2 copiar: Tipo de información, tipo de rasgo geológico.

referencia50_p_combinada: Tipo de información, tipo de rasgo geográfico.

salto_agua_l_combinada: Tipo de rasgo geológico, tipo de cuerpo de agua.

salto_agua50_p_combinada: Tipo de rasgo geológico, tipo de cuerpo de agua.

tanque_agua50_p_combinada_final: Tipo de estructura, tipo de tanque, relación con el suelo.

terreno_suj50_a_combinada: Terreno sujeto a inundación.

ue_combinada: Descripción del material.

Pendiente: Pendiente.

us250s3a: Tipo de información, tipo de ecosistema, aspecto, otros.

us250s3c: Tipo de información, tipo de cultivo 1, tipo de cultivo 2, tipo de cultivo 3.

us250s3d: Tipo de información, tipo de ecosistema, descripción vegetal, cobertura arbórea, otros.

us250s3e: Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.

us250s3f: Tipo de información, tipo de actividad.

us250s3g: Tipo de información, tipo de actividad, subtipo de actividad, tipo de plantación, tipo de cultivo 1, tipo de cultivo 2, otros.

us250s3h: Tipo de información, tipo de ecosistema, descripción vegetal, altura, otros.

us250s3i_1 copiar: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal.

us250s3l_1 copiar: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal.

us250s3n: Tipo de información, nomadismo agrícola, otros.

us250s3r: Tipo de información, erosión, otros.

us250s3v: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal, fase, otros.

us250s4a: Tipo de información, tipo de ecosistema, aspecto, otros.

us250s4c_1 copiar: Tipo de información, tipo de cultivo 1, tipo de cultivo 2, tipo de cultivo 3.

us250s4d: Tipo de información, tipo de ecosistema, descripción vegetal, cobertura arbórea, otros.

us250s4e_1 copiar: Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.

us250s4f_2 copiar: Tipo de información, tipo de actividad.

us250s4g: Tipo de información, tipo de actividad, subtipo de actividad, tipo de plantación, tipo de cultivo 1, tipo de cultivo 2, otros.

us250s4h: Tipo de información, tipo de ecosistema, descripción vegetal, altura, otros.

us250s4i: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal, otros.

us250s4l: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal.

us250s4n: Tipo de información, nomadismo agrícola, otros.

us250s4o: Tipo de actividad, tipo de información.

us250s4p, us250s4p_1 copiar: Tipo de información, tipo de ganado.

us250s4v: Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal, fase, otros.

ve_combinada: Tipo de área.

vegeta_dens50_a_combinada: Densidad de vegetación.

zona50_a copiar: Zona arenosa

16.3.2. Tipo de codificación de las variables

16.3.2.1. Acuífero de la Zona Metropolitana de la Ciudad de México

Nombre del archivo	Variables nominales (Embedding)	Variables ordinales (Ordinal Encoding)	Variables numéricas (Standard Scaling)
a2606		Condición del acuífero.	
e1402pg_coord	Tipo de roca, nombre de roca.	Porosidad, permeabilidad.	
actividades_forestales_pec_coord	Actividad forestal o pecuaria, Tipo de actividad.		
acueducto_l_cort1_fin_copiar , acueducto_l_cort2_fin_copiar , acueducto_l_cort3_coord , acueducto_l_cort4_copiar_copiar , acueducto_l_cort5_copiar_copiar , acueducto50_l_cort7_coord , acueducto50_l_cort8_coord	Tipo de estructura.	Relación en el suelo.	
ae1402v_cort_fin_copiar	Tipo de obra, uso del agua, observaciones.	Agresividad del agua.	valores de distintas medidas de un análisis del agua: Ca, Mg, Na, K, CaCO ₃ , RAS, pH, Ce, SO ₄ , HCO ₃ , NO ₃ , CO ₃ , Cl, sólidos, calidad del agua, diámetro de la tubería de descarga,

			profundidad total de la obra, nivel estático del agua, nivel dinámico del agua (nivel del agua cuando el pozo está bombeando), caudal (cantidad del agua que produce el pozo), temperatura.
agricultura_y_vegetacion_coord	Uso de suelo, tipo de uso de suelo, fisionomía, vegetación secundaria, cultivo primario, cultivo secundario, cultivo terciario.	Erosión.	
areacultiv50_a_cort_coord, areacultiv50_a_cort2_coord	Uso de suelo.		
c2606_coord	Descripción.		Elevación del nivel estático del agua.
canal_a_cort1_fin_copiar, canal_a_cort2_fin_copiar, canal_a_cort3_fin_copiar, canal_a_cort4_coord	Término general.		
canal_l_cort1_fin_copiar, canal_l_cort2_fin_copiar, canal_l_cort3_coord, canal50_l_cort2_coord	Término general.		
clave_de_cultivos_cort_copiar_c	Entidad, asociación, nombre común 1, nombre común 2, nombre común 3.		
clave_de_especie_vegetales_cort:	Entidad, asociación, nombre científico 1, nombre científico 2, nombre científico 3, nombre científico 4, nombre científico 5.		
corriente_ag_a_cort_fin_copiar, corriente_ag_l_cort1_fin_copiar, corriente_ag_l_cort2_coord,	Tipo de formación natural, tipo de cuerpo de agua.	Condición del cuerpo de agua.	

corriente_ag_l_cort3_fin_copiar, corriente_ag_l_cort4_fin_copiar, corriente_ag50_l_cort1_coord, corriente_ag50_l_cort2_coord, corriente_ag50_l_cort3_coord			
corriente_de_p_cort1_fin_copiar, corriente_de_p_cort2_fin_copiar, corriente_de_p_cort3_fin_copiar, corriente_de_p_cort4_copiar_cop, corriente_de50_p_cort2_fin_copi, corriente_de50_p_cort3_copiar_c	Tipo de corriente, tipo de rasgo geológico.		
cuerpo_agua_a_cort1_fin_copiar, cuerpo_agua_a_cort2_fin_copiar, cuerpo_agua_a_cort3_fin_copiar, cuerpo_agua_a_cort4_fin_copiar, cuerpo_agua_a_cort5_coord, cuerpo_agua50_a_cort3_coord	Tipo de rasgo geológico, tipo de cuerpo de agua.	Condición del cuerpo de agua.	
e1402_usv250s5a_coord	Tipo de información, Tipo de ecosistema, Tipo de uso de suelo.		
e1402_usv250s5d_coord	Tipo de información, tipo de ecosistema, tipo de uso de suelo.	Densidad vegetal, cobertura de árboles	
e1402_usv250s5g_coord	Tipo de información, Tipo de ecosistema, área geoestadística, tipo de planta, tipo de cultivo 1, tipo de cultivo 2, otros.		
e1402_usv250s5h_coord	Tipo de información, tipo de ecosistema, otros.	Densidad vegetal, altura de vegetación	
e1402_usv250s5v_coord	Tipo de información, tipo de ecosistema, tipo de vegetación, fase, otros.	Densidad vegetal	
e1402e_coord	Clave del suelo, nombre del suelo 1, nombre del		

	subsuelo 1, nombre del suelo 2, nombre del subsuelo 2, nombre del suelo 3, nombre del subsuelo 3, clase textural, fase física del suelo, fase química del suelo, entidad del suelo, clave de interpretación.		
e1402lg_coord	Tipo de estructura geológica, tipo, movimiento horizontal, movimiento vertical.		Longitud de fractura, inclinación, buzamiento.
e1402ptg_fin_copiar	Código del objeto, entidad geológica, análisis realizado, tipo, mineral principal, mineral secundario, tercer mineral, uso del material geológico.		Inclinación, rumbo, buzamiento, dirección de la formación geológica.
Pendiente			Valor de la pendiente
e1402u1v_coord	Entidad, tipo, fisionomía, vegetación secundaria, cultivo primario, cultivo secundario, cultivo terciario	Erosión	
e2606_coord	Descripción.		
estanque_a_cort1_fin_copiar, estanque_a_cort4_fin_copiar, estanque_a_cort5_fin_copiar	Tipo de estanque, tipo de estructura.		
f2606_fin_copiar	Descripción de la capa		Azimutal de la dirección del flujo de agua
fe1402v_coord:	Descripción		Longitud de la dirección de flujo
h2606_coord	Descripción		Profundidad del nivel estático del agua

manantial_p_cort1_copiar_copiar, manantial_p_cort2_fin_copiar, manantial_p_cort4_1_copiar_copi, manantial_p_cort5_copiar_copiar	Tipo de rasgo geológico, tipo de rasgo hidrográfico.		
manantial50_p_cort8_copiar_copi	Tipo de rasgo hidrográfico.		
ne1402v_cort_copiar_copiar	Tipo de obra.		
o2606_cort_copiar_copiar	Tipo de obra, uso del agua, observaciones.		Caudal, nivel estático del agua, nivel dinámico del agua.
p2606_cort_copiar_copiar	Tipo de agresividad química del agua, uso del agua, termalismo, composición hidroquímica, familia hidroquímica del agua, observaciones.	Clasificación de agua para riego	Análisis fisicoquímico del agua: Ca, Mg, K, Dureza como CaCO ₃ , Relación de Adsorción de Sodio (RAS), pH medido en campo, pH medido en laboratorio, conductividad eléctrica medida en campo, conductividad eléctrica medida en laboratorio, SO ₄ , HCO ₃ , NO ₃ , CO ₃ , Cl, sólidos totales disueltos medidos en campo, Sólidos totales disueltos medidos en laboratorio, caudal del pozo, nivel estático del agua, nivel dinámico del agua, temperatura, salinidad.
pe1402v_coord	Descripción		
pozo_p_cort1_fin_copiar, pozo_p_cort2_fin_copiar, pozo_p_cort3_fin_copiar, pozo_p_cort4_fin_copiar:	Condición del pozo, tipo de pozo, tipo de estructura.		
referencia_g_p_cort1_copiar_cop, referencia50_p_cort1_copiar_cop	Rasgo geológico.		
t2606_cort_copiar_copiar	Comportamiento hidráulico del acuífero, método, observaciones		Caudal, nivel estático del agua, nivel dinámico del agua, nivel máximo del agua estática, caudal de agua subterránea, transmisividad del acuífero, coeficiente de

			almacenamiento, espesor del acuífero, conductividad hidráulica
terreno_suje_a_cort1_copiar_cop, terreno_suje_a_cort2_coord, terreno_suje_a_cort3_copiar_cop u2606_coord	Rasgo geológico, tipo.		
ue1402v_coord	Descripción del área del acuífero.		
us250s3a_coord		Descripción del material.	
us250s3a_coord	Tipo de información, tipo de ecosistema, otros.		
us250s3c_cort_copiar_copiar	Tipo de información, asociación, cultivo 1, cultivo 2, cultivo 3.		
us250s3d_4_coord	Tipo de información, tipo de vegetación, descripción vegetal, otros.	Cobertura de árboles	
us250s3e_fin_copiar	Tipo de información, asociación, especie 1, especie 2, especie 3, especie 4, especie 5.		
us250s3g_coord	Tipo de información, tipo de cultivo 1, tipo de cultivo 2, otros.		
us250s3h_coord	Tipo de información, descripción vegetal, otros.	Altura	
us250s3i_fin_copiar	Tipo de información, tipo de ecosistema, tipo de vegetación.		
us250s3n_coord .	Tipo de información,		

	nomadismo agrícola, otros		
us250s3r_coord	Tipo de información, otros	Erosión	
us250s3v_coord	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de vegetación, fase de vegetación, otros.		
us250s4a_coord	Tipo de información, tipo de ecosistema, aspecto, otros.		
us250s4d_coord	Tipo de información, tipo de ecosistema, descripción de la vegetación, otros.	Cobertura de árboles	
us250s4e_fin_copiar	Tipo de información, asociación, especie 1, especie 2, especie 3, especie 4, especie 5.		
us250s4g_coord	Tipo de información, tipo de cultivo 1, tipo de cultivo 2, otros.		
us250s4h_coord	Tipo de información, tipo de ecosistema, descripción vegetal, otros.	Altura arbórea	
us250s4i_fin_copiar	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de la vegetación, otros.		
us250s4v_3_coord	Tipo de información, tipo de ecosistema, tipo de vegetación,		

	descripción de la vegetación, fase, otros.		
vegeta_dens50_a_cort1_coord, vegeta_dens50_a_cort2_coord, vegeta_dens50_a_cort3_coord:		Densidad vegetal.	
z2606_coord	Zonas de recarga y extracción del acuífero.		

16.3.2.2. Acuífero de La Sierra, Tabasco

Nombre del archivo	Variables nominales (Embedding)	Variables ordinales (Ordinal Encoding)	Variables numéricas (Standard Scaling)
actividades_forestales_pec copiar	Entidad, tipo de actividad.		
acueducto50_l_combinada	Tipo de estructura, relación con el suelo.		
ae1509v copiar	Tipo de obra, agresividad del agua, uso, observaciones.	calidad del agua	valores de distintas medidas de un análisis del agua: Ca, Mg, Na, K, CaCO ₃ , RAS, PH, Ce, SO ₄ , HCO ₃ , NO ₃ , CO ₃ , Cl, sólidos, diámetro de la tubería de descarga, nivel estático del agua, nivel dinámico del agua, caudal, temperatura
agricultura_nomada	Entidad		
agricultura_y_vegetacion	Entidad, tipo de agricultura, fisionomía, vegetación secundaria, cultivo primario, cultivo secundario, cultivo terciario	Erosión	
area_cultiv50_a_combinada	Entidad		
area_ver_ur50_a copiar	Entidad		
bordo_l_combinada	Tipo de estructura		
canal50_l_combinada	Tipo de cuerpo de agua		

clave_de_cultivos	Entidad, nombre común 1, nombre común 2, nombre común 3.		
clave_de_especie_vegetales	Entidad, nombre científico 1, nombre científico 2, nombre científico 3, nombre científico 4, nombre científico 5.		
corriente_ag_a_combinada	Tipo de rasgo geológico, tipo de cuerpo de agua	condición del cuerpo de agua	
corriente_ag_l_combinada, corriente_ag_l_combinada2	Tipo de rasgo geológico, tipo de cuerpo de agua	condición del cuerpo de agua	
corriente_ag50_l_combinada	Tipo de cuerpo de agua	condición del cuerpo de agua	
corriente_de_p_combinada	Tipo de rasgo geológico, tipo de cuerpo de agua.		
corriente_de50_p_combinada	Tipo de cuerpo de agua.		
cuerpo_agua_a_combinada	Tipo de rasgo geológico, tipo de cuerpo de agua	condición del cuerpo de agua	
eu1v	Entidad, tipo de vegetación, fisonomía, vegetación secundaria, cultivo primario, cultivo secundario, cultivo terciario.	erosión	
e1_usv250s5a	Tipo de información, aspecto, otros.		
e1_usv250s5d	Tipo de información, tipo de ecosistema, descripción de la vegetación, otros.	cobertura arbórea	
e1_usv250s5e	Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.		
e1_usv250s5f	Tipo de información.		
e1_usv250s5g	Tipo de información, clasificación de la		

	información, descripción de la información, tipo de plantación, tipo de cultivo 1, tipo de cultivo 2, otros.		
e1_usv250s5h	Tipo de información, tipo de ecosistema, descripción vegetal, otros.	altura	
e1_usv250s5l	Tipo de información, clave.		
e1_usv250s5n	Tipo de información, nomadismo agrícola, otros.		
e1_usv250s5o	Tipo de información, clave.		
e1_usv250s5v	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción de la vegetación, fase, otros.		
e15_combinada	Clave, nombre de suelo 1, nombre del subsuelo 1, nombre de suelo 2, nombre del subsuelo 2, nombre de suelo 3, nombre del subsuelo 3, fase física, fase química, entidad.	clase textural	
e15d22_toponimo50_p _utm copiar	Tipo de información, tipo de rasgo geológico.		
e1508_usv250s5c copiar	Tipo de información, tipo de cultivo 1, tipo de cultivo 2, tipo de cultivo 3.		
e1508_usv250s5i copiar	Tipo de información, clave.		
e1508_usv250s5p copiar	Tipo de información, clave.		
E15080911_combinada	Clave, grupo 1, calificador del grupo 1, calificador primario del grupo 1, grupo 2, calificador del grupo 2, calificador primario del grupo 2, grupo 3, calificador del grupo 3, calificador primario del grupo 3, frudica	clase textural	
ee_combinada		Coeficiente de escurrimiento.	
est_geo_grav_p_combinada			Gravedad, altura ortométrica

estanque50_a_combinada	Tipo de estructura, tipo de estanque.		
fe1508v copiar	Tipo de información (flujo de agua).		
geologia_area	Tipo de roca	porosidad, permeabilidad.	
geologia_linea	tipo de estructura geológica, subtipo de estructura geológica, movimiento horizontal, movimiento vertical, inclinación, buzamiento.		Longitud
geologia_punto	Entidad, inclinación, rumbo, buzamiento, tipo de mineralización asociada.		
manantial_p_combinada	Tipo de rasgo geológico, tipo de cuerpo de agua.		
pe_combinada	Tipo de cuerpo de agua, agresividad del agua, uso, descripción.	calidad del agua	valores de distintas medidas de un análisis del agua: Ca, Mg, Na, K, CaCO ₃ , RAS, PH, Ce, SO ₄ , HCO ₃ , NO ₃ , CO ₃ , Cl, sólidos, temperatura
pev_combinada:.	Descripción de uso de suelo		
pozo_p_combinada	Tipo de estructura, tipo de pozo.		
referencia_g_p_2 copiar	Tipo de información, tipo de rasgo geológico.		
referencia50_p_combinada	Tipo de información, tipo de rasgo geográfico.		
salto_agua_l_combinada	Tipo de rasgo geológico, tipo de cuerpo de agua.		
salto_agua50_p_combinada	Tipo de rasgo geológico, tipo de cuerpo de agua.		
tanque_agua50_p_combinada_final	Tipo de estructura, tipo de tanque, relación con el suelo.		
terreno_suj50_a_combinada	Terreno sujeto a inundación.		
ue_combinada		Descripción del material.	
Pendiente			Pendiente

us250s3a	Tipo de información, tipo de ecosistema, aspecto, otros.		
us250s3c	Tipo de información, tipo de cultivo 1, tipo de cultivo 2, tipo de cultivo 3.		
us250s3d, us250s4d	Tipo de información, tipo de ecosistema, descripción vegetal, otros.	cobertura arbórea	
us250s3e	Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.		
us250s3f	Tipo de información, tipo de actividad.		
us250s3g	Tipo de información, tipo de actividad, subtipo de actividad, tipo de plantación, tipo de cultivo 1, tipo de cultivo 2, otros.		
us250s3h	Tipo de información, tipo de ecosistema, descripción vegetal, otros.	altura	
us250s3i_1 copiar	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal.		
us250s3l_1 copiar	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal.		
us250s3n	Tipo de información, nomadismo agrícola, otros.		
us250s3r	Tipo de información, otros.	erosión	
us250s3v	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal, fase, otros.		
us250s4a:	Tipo de información, tipo de ecosistema, aspecto, otros.		
us250s4c_1 copiar	Tipo de información, tipo de cultivo 1, tipo de cultivo 2, tipo de cultivo 3.		

us250s4d	Tipo de información, tipo de ecosistema, descripción vegetal, otros.	cobertura arbórea	
us250s4e_1 copiar	Tipo de información, especie 1, especie 2, especie 3, especie 4, especie 5.		
us250s4f_2 copiar	Tipo de información, tipo de actividad.		
us250s4g	Tipo de información, tipo de actividad, subtipo de actividad, tipo de plantación, tipo de cultivo 1, tipo de cultivo 2, otros.		
us250s4h	Tipo de información, tipo de ecosistema, descripción vegetal, otros.	altura	
us250s4i	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal, otros.		
us250s4l	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal.		
us250s4n	Tipo de información, nomadismo agrícola, otros.		
us250s4o	Tipo de actividad, tipo de información.		
us250s4p, us250s4p_1 copiar	Tipo de información, tipo de ganado.		
us250s4v	Tipo de información, tipo de ecosistema, tipo de vegetación, descripción vegetal, fase, otros.		
ve_combinada	Tipo de área.		
vegeta_dens50_a_co mbinada		Densidad de vegetación.	
zona50_a copiar	Zona arenosa		

16.4. Anexo 4

Enlace de github: <https://github.com/sarailopez-23/Localizador-de-zonas-de-recarga-de-acuiferos.git>

17. Referencias

- Kuzma, S., Saccoccia, L., & Chertock, M. (16 de Agosto de 2023). *25 Countries, Housing One-Quarter of the Population, Face Extremely High Water Stress*. Obtenido de World Resources Institute: <https://www.wri.org/insights/highest-water-stressed-countries>
- Ortega Martinez, R. A., Ramos Zenteno, R. M., Garcés Hazou, C. L., Pardo Ledezma, A., Cuadros Pariente, N. A., & Poma Cruz, N. (Noviembre de 2025). *Family satisfaction in the Intensive Care Unit: A multidimensional network analysis approach*. Obtenido de ResearchGate: https://www.researchgate.net/publication/397433011_Family_satisfaction_in_the_Intensive_Care_Unit_A_multidimensional_network_analysis_approach
- Agua.org.mx. (8 de Noviembre de 2023). *Derecho Humano al Agua y al Saneamiento*. Obtenido de Agua.org.mx: <https://agua.org.mx/tus-derechos-sobre-el-agua/#contexto-nacional>
- Al Masud, A. (2026). *Correlation Matrix: What is it, How It Works & Examples*. Obtenido de QuestionPro: <https://www.questionpro.com/blog/correlation-matrix/>
- Alatríste, A. (20 de Enero de 2025). *Conagua ignora niveles de aguas subterráneas a nivel nacional*. Obtenido de EME EQUIS: <https://emeequis.com/investigaciones/conagua-ignora-niveles-de-aguas-subterraneas-a-nivel-nacional/>
- Amat Rodrigo, J. (Agosto de 2025). *Random Forest con Python*. Obtenido de Ciencia de datos: https://cienciadedatos.net/documentos/py08_random_forest_python#Ejemplo_clasificaci%C3%B3n
- Amat Rodrigo, J. (Agosto de 2025). *Random Forest con Python*. Obtenido de Ciencia de datos: https://cienciadedatos.net/documentos/py08_random_forest_python#Ejemplo_clasificaci%C3%B3n
- AQUEDUCT a. (2023). *Aqueduct Country Rankings*. Obtenido de World Resources Institute: <https://www.wri.org/applications/aqueduct/country-rankings/?country=MEX>
- AQUEDUCT b. (2023). *Aqueduct Water Risk Atlas*. Obtenido de World Resources Institute: <https://www.wri.org/applications/aqueduct/water-risk->

atlas/#/?advanced=false&basemap=hydro&geoStore=a32bff8dec2cf0fa3ee2d050a18a8ea3&indicator=w_awr_def_tot_cat&lat=19.509573511550222&lng=-99.59930453449489&mapMode=analysis&month=1&opacity=0.5&ponderation

Barrios, J. I. (26 de Julio de 2019). *La matriz de confusión y sus métricas*. Obtenido de Health Big Data: <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>

Brownlee , J. (28 de Agosto de 2020). *How to Use StandardScaler and MinMaxScaler Transforms in Python*. Obtenido de Machine learning mastery: <https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/>

codificandobits. (22 de Enero de 2024). *¿Cómo codificar datos categóricos?* Obtenido de codificandobits: <https://codificandobits.com/blog/como-codificar-datos-categoricos/>

Comisión Federal de Electricidad (CFE). (s.f.). *Historia de la CFE*. Recuperado el 1 de Julio de 2026, de Gobierno de México |CFE: <https://www.cfe.gob.mx/nuestraempresa/pages/historia.aspx>

Comisión Nacional del Agua (CONAGUA). (2006). *CENSO DE APROVECHAMIENTOS DE AGUA SUBTERRÁNEA, ACUÍFEROS ZMCM Y CUAUTITLÁN-PACHUCA CONAGA 2006*. Ciudad de México: Comisión Nacional del Agua (CONAGUA).

Comisión Nacional del Agua (CONAGUA). (Noviembre de 2009). *Semblanza histórica del agua en México*. (S. d. (SEMARNAT), Ed.) Obtenido de CONAGUA: https://web.archive.org/web/20131019125614if_/http://www.conagua.gob.mx/CONAGUA07/Publicaciones/Publicaciones/SGP-28SemblanzaHist%C3%B3ricaM%C3%A9xico.pdf

Comisión Nacional del Agua (CONAGUA). (31 de Mayo de 2014). *Vedas, Reservas y Reglamentos de Aguas Nacionales Superficiales*. Obtenido de Gobierno de México: <https://www.gob.mx/conagua/documentos/vedas-reservas-y-reglamentos-de-aguas-nacionales-superficiales>

Comisión Nacional del agua (CONAGUA). (20 de Abril de 2015). *Actualización de la disponibilidad media anual de agua en el acuífero Zona Metropolitana de la Ciudad de México (0901)*. Obtenido de Comisión Nacional del agua (CONAGUA): https://www.gob.mx/cms/uploads/attachment/file/102942/DR_0901.pdf

- Comisión Nacional del agua (CONAGUA). (2023 a). *Disponibilidad por acuífero: Ciudad de México*. Obtenido de Comisión Nacional del agua /CONAGUA): <https://sigagis.conagua.gob.mx/gas1/sections/Edos/cdmx/cdmx.html>
- Comisión Nacional del agua (CONAGUA). (2023 b). *Disponibilidad por acuífero: Tabasco*. Obtenido de Comisión Nacional del agua (CONAGUA): <https://sigagis.conagua.gob.mx/gas1/sections/Edos/tabasco/tabasco.html>
- Comisión Nacional del Agua (CONAGUA). (2024 a). *ACTUALIZACIÓN DE LA DISPONIBILIDAD MEDIA ANUAL DE AGUA EN EL ACUÍFERO LA SIERRA (2705), ESTADO DE TABASCO*. Ciudad de México: Comisión Nacional del Agua (CONAGUA).
- Comisión Nacional del Agua (CONAGUA). (2024 b). *ACTUALIZACIÓN DE LA DISPONIBILIDAD MEDIA ANUAL DE AGUA EN EL ACUÍFERO ZONA METROPOLITANA DE LA CD. DE MÉXICO (0901), CIUDAD DE MÉXICO*. Ciudad de México: Comisión Nacional del Agua (CONAGUA).
- Comisión Nacional del Agua (CONAGUA). (2024 c). *Estadísticas del Agua en México 2023*. México: Comisión Nacional del Agua (CONAGUA).
- Comisión Nacional del Agua (CONAGUA). (2026). *Información estadística para la planeación hídrica*. Obtenido de Comisión Nacional del Agua (CONAGUA): <https://sinav30.conagua.gob.mx:8080/Descargas/pdf/InformacionEstadisticaParaLaPlaneacionHidrica.pdf>
- Comisión Nacional del Agua (CONAGUA). (s.f.). *Geovisor de acuífero*. Obtenido de Comisión Nacional del Agua (CONAGUA): <https://sigagis.conagua.gob.mx/dma230911/>
- Dashti Latif, S., & Najah Ahmed, A. (2023). A review of deep learning and machine learning techniques for hydrological inflow forecasting. *Environment, Development and Sustainability*.
- Disfruta las matemáticas. (2025). *Diagrama de Dispersión*. Obtenido de Disfruta las matemáticas: <https://www.disfrutalasmatematicas.com/datos/grafica-dispersion-xy.html>
- Disfruta las matemáticas. (2025). *Gráficas de Barras*. Obtenido de Disfruta las matemáticas: <https://www.disfrutalasmatematicas.com/datos/graficas-de-barras.html>

- Duffield, G. M. (23 de Noviembre de 2019). *Representative Values of Hydraulic Properties*. Obtenido de AQTESOLV: https://www.aqtesolv.com/aquifer-tests/aquifer_properties.htm?utm_source=chatgpt.com
- El agua subterránea representa 97% del líquido disponible*. (24 de Marzo de 2022). Obtenido de Gaceta UNAM: <https://www.gaceta.unam.mx/el-agua-subterranea-representa-97-del-liquido-disponible/>
- Espíndola, G. (2 de Marzo de 2023). *¿Qué son los embeddings y cómo se utilizan en la inteligencia artificial con python?* Obtenido de Medium: <https://gustavo-espindola.medium.com/qu%C3%A9-son-los-embeddings-y-c%C3%B3mo-se-utilizan-en-la-inteligencia-artificial-con-python-45b751ed86a5>
- Fang, W., Ren, K., Liu, T., Shang, J., Jia, S., Jiang, X., & Zhang, J. (2024). An evaluation of random forest based input variable selection methods for one month ahead streamflow forecasting. *Scientific reports*.
- Feature-engine developers. (s.f.). *Categorical Encoding*. Obtenido de Feature-engine: https://feature-engine.trainindata.com/en/1.7.x/user_guide/encoding/#categorical-encoding
- Feature-engine. (s.f.). *Ordinal Encoding*. Obtenido de Feature-engine: https://feature-engine.trainindata.com/en/1.8.x/user_guide/encoding/OrdinalEncoder.html
- Ferrer, J. (16 de Enero de 2025). *Fuzzy Logic in AI: Principles, Applications, and Python Implementation Guide*. Obtenido de datacamp: <https://www.datacamp.com/tutorial/fuzzy-logic-in-ai>
- Fetter, C. W. (2001). Chapter 3: Properties of Aquifers. En C. W. Fetter, *Applied Hydrogeology* (págs. 69-78). New Jersey: Prentice-Hall.
- Freeze, A., & Cherry A., J. (1979). Capítulo 2: Principios y propiedades físicas. En A. Freeze., & J. Cherry A., *Agua Subterránea*. The Groundwater Project.
- Gerencia de aguas subterráneas. (Junio de 2026). *GeoVisor de Consulta de Mediciones Piezométricas*. Obtenido de Gobierno de México: <https://sigagis.conagua.gob.mx/rp20/>
- Gobierno del estado de México. (18 de Septiembre de 2024). Programa Sectorial: Bienestar ambiental y acceso universal al agua 2024-2029. *Periódico oficial Gaceta del gobierno, gobierno del Estado de México*, págs. 1500-1518.

- González-Ortigoza, S., Hernández-Espriú, A., & Arciniega-Esparza, S. (2023). Regional modeling of groundwater recharge in the Basin of Mexico: new insights from satellite observations and global data sources. *Hydrogeology Journal*.
- Huet, P. (12 de Julio de 2024). *Embeddings: Qué son y cómo transforman datos en información*. Obtenido de OpenWebinars:
<https://openwebinars.net/blog/embeddings/>
- IBM. (20 de Noviembre de 2020). *¿Qué es el análisis exploratorio de datos (EDA)?*
Obtenido de IBM: <https://www.ibm.com/mx-es/think/topics/exploratory-data-analysis>
- INEGI. (2001). *Síntesis de la Información geográfica del estado de Tabasco*. Obtenido de INEGI:
https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/historicos/2104/702825223939/702825223939_7.pdf
- Instituto de geografía, UNAM. (s.f.). *Atlas Nacional de México, 6.3 Hidrogeología*. Obtenido de Instituto de geografía, UNAM:
https://geodigital.geografia.unam.mx/atlas_nacional/index.html/grals/Tomo_II/IV.Naturaleza/IV.6.Hidrogeografia/IV.6.3.jpg
- Instituto Mexicano de Tecnología del Agua. (11 de Octubre de 2020). *Repaso histórico del agua en México – Parte I (1888-1917)*. Obtenido de Gobierno de México:
<https://www.gob.mx/imta/es/articulos/repaso-historico-del-agua-en-mexico-parte-i-1888-1917?idiom=es>
- Instituto Nacional de Estadística y Geografía (INEGI). (1983). *Conjunto de datos vectoriales Geológicos serie I. Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/geologia/>
- Instituto Nacional de Estadística y Geografía (INEGI). (1984). *Conjunto de datos vectoriales Geológicos serie I. Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/geologia/>
- Instituto Nacional de Estadística y Geografía (INEGI). (1985). *Conjunto de datos vectoriales Geológicos serie I. Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/geologia/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2000 a). *Modelos Digitales de Elevación. Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/relieve/continental/#descargas>

- Instituto Nacional de Estadística y Geografía (INEGI). (2000 b). *Modelos Digitales de Elevación. Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/relieve/continental/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2000 c). *Modelos Digitales de Elevación. Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/relieve/continental/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2000 d). *Modelos Digitales de Elevación. Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/relieve/continental/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2001). *Síntesis de Información geográfica del estado de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/historicos/2104/702825224028/702825224028_12.pdf?utm_source=chatgpt.com
- Instituto Nacional de Estadística y Geografía (INEGI). (2007). *Conjunto de Datos Vectorial Edafológico. Escala 1:250 000 Serie II Continuo Nacional Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2019 a). *Conjunto de Datos Vectoriales de Información Topográfica E14A38 Toluca de Lerdo escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2019 b). *Conjunto de Datos Vectoriales de Información Topográfica E14B41 Ameca de Juárez escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2019 c). *Conjunto de Datos Vectoriales de Información Topográfica E15B81 Ocuilzapotlán escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2019). *Carta hidrológica aguas subterráneas, México, escala 1:1 000 000, Informe técnico, Serie II*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):

https://www.inegi.org.mx/contenido/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825109639.pdf

Instituto Nacional de Estadística y Geografía (INEGI). (2019). *Conjunto de Datos Vectoriales de Información Topográfica E15B81 Ocuilzapatlán escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 d). *Conjunto de Datos Vectoriales de Información Topográfica E15C19 Reforma escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 e). *Conjunto de Datos Vectoriales de Información Topográfica E15D11 Villahermosa escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 f). *Conjunto de Datos Vectoriales de Información Topográfica E15D21 Teapa escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 g). *Conjunto de Datos Vectoriales de Información Topográfica E15D42 Petalcingo escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 h). *Conjunto de Datos Vectoriales de Información Topográfica E15D43 Yajalón escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 i). *Conjunto de Datos Vectoriales de Información Topográfica E15D51 Bochil escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI). (2019 j). *Conjunto de Datos Vectoriales de Información Topográfica E15D52 Oxchuc escala 1:50 000, 2019*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>

- Instituto Nacional de Estadística y Geografía (INEGI). (2021 a). *Conjunto de Datos Vectoriales de Información Topográfica E14A39 (Ciudad de México) Escala 1:50 000, 2021*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2021 b). *Conjunto de Datos Vectoriales de Información Topográfica E14A49 (Milpa Alta) Escala 1:50 000, 2021*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2021 c). *Conjunto de Datos Vectoriales de Información Topográfica E15D23 (Catazajá) Escala 1:50 000, 2021*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2021). *Conjunto de Datos Vectoriales de Información Topográfica E14A29 (Cuautitlán) Escala 1:50 000, 2021*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2021 d). *Conjunto de Datos Vectoriales de Información Topográfica E15D33 (Tumbalá) Escala 1:50 000, 2021*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2021 e). *Conjunto de Datos Vectoriales de Información Topográfica E15D34 (Río Chancalá) Escala 1:50 000, 2021*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#mapas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2023). *Conjunto de datos vectoriales de información topográfica a escala 1:50 000, 2023 E14B31*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2025 a). *Carta Hidrológica de Aguas Subterráneas 1:250 000 E1401*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI). (2025 b). *Carta Hidrológica de Aguas Subterráneas 1:250 000 E1402*. Obtenido de Instituto Nacional de

Estadística y Geografía (INEGI):

<https://www.inegi.org.mx/temas/hidrologia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI). (2025 c). *Carta Hidrológica de Aguas Subterráneas 1:250 000 E1405*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):

<https://www.inegi.org.mx/temas/hidrologia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI). (22 de Marzo de 2025).

Estadísticas a propósito del Día Mundial del Agua. Obtenido de INEGI:

https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2025/EAP_DM_unAgua.pdf

Instituto Nacional de Estadística y Geografía (INEGI). (2025). *Información Topográfica 1:250 000 E1402*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/topografia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) a. (s.f.). *Carta edafológica Escala 1:250 000 Serie III E1402*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) aa. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie II. Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ab. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie III. Continuo Nacional. Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) ac. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie III. Continuo Nacional. Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ad. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie III. Continuo Nacional. Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ae. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie III. Continuo*

Nacional. Villahermosa. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) af. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación. Ciudad de México.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) ag. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación. Tenosique.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ah. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación. Tuxtla Gutierrez.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ai. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación. Villahermosa.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) aj. (s.f.). *Conjunto de datos vectoriales de la carta Edafológica escala 1:250 000 serie I Ciudad de México.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) ak. (s.f.). *Conjunto de datos vectoriales de la carta Edafológica escala 1:250 000 serie I Tenosique.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/>

Instituto Nacional de Estadística y Geografía (INEGI) al. (s.f.). *Conjunto de datos vectoriales de la carta Edafológica escala 1:250 000 serie I Tuxtla Gutierrez.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/>

Instituto Nacional de Estadística y Geografía (INEGI) am. (s.f.). *Conjunto de datos vectoriales de la carta Edafológica escala 1:250 000 serie I Villahermosa.* Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/>

Instituto Nacional de Estadística y Geografía (INEGI) an. (s.f.). *Conjunto de datos vectoriales de uso del suelo y vegetación escala 1:250 000 serie V Conjunto Nacional Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) ao. (s.f.). *Conjunto de datos vectoriales de uso del suelo y vegetación escala 1:250 000 serie V Conjunto Nacional Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ap. (s.f.). *Conjunto de datos vectoriales de uso del suelo y vegetación escala 1:250 000 serie V Conjunto Nacional Tuxtla Gutiérrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) aq. (s.f.). *Conjunto de datos vectoriales de uso del suelo y vegetación escala 1:250 000 serie V Conjunto Nacional Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

Instituto Nacional de Estadística y Geografía (INEGI) ar. (s.f.). *Conjunto de datos vectoriales Geológicos serie I. Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/geologia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) b. (s.f.). *Conjunto de Datos Vectorial Edafológico. Escala 1:250 000 Serie II Continuo Nacional Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/#descargas>

Instituto Nacional de Estadística y Geografía (INEGI) c. (s.f.). *Conjunto de Datos Vectorial Edafológico. Escala 1:250 000 Serie II Continuo Nacional Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/>

Instituto Nacional de Estadística y Geografía (INEGI) d. (s.f.). *Conjunto de Datos Vectorial Edafológico. Escala 1:250 000 Serie II Continuo Nacional Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/edafologia/>

Instituto Nacional de Estadística y Geografía (INEGI) e. (s.f.). *Conjunto de datos vectoriales de información topográfica E15C29 Pichucalco escala 1:50 000*

serie III. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) f. (s.f.). *Conjunto de datos vectoriales de información topográfica E15C39 Ixhuatán escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) g. (s.f.). *Conjunto de datos vectoriales de información topográfica E15C49 Tecpatán escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) h. (s.f.). *Conjunto de datos vectoriales de información topográfica E15D22 Salto del Agua escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) i. (s.f.). *Conjunto de datos vectoriales de información topográfica E15D31 Amatán escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) j. (s.f.). *Conjunto de datos vectoriales de información topográfica E15D32 Tila escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) k. (s.f.). *Conjunto de datos vectoriales de información topográfica E15D41 Jitotol de Zaragoza escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) l. (s.f.). *Conjunto de datos vectoriales de información topográfica E15D44 San Jerónimo Tulijá escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

Instituto Nacional de Estadística y Geografía (INEGI) m. (s.f.). *Conjunto de datos vectoriales de información topográfica E15D53 Ocosingo escala 1:50 000 serie III*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI):
<https://www.inegi.org.mx/temas/topografia/#mapas>

- Instituto Nacional de Estadística y Geografía (INEGI) n. (s.f.). *Conjunto de datos vectoriales de la carta de Aguas subterráneas. Escala 1:250 000. Serie I. Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/>
- Instituto Nacional de Estadística y Geografía (INEGI) o. (s.f.). *Conjunto de datos vectoriales de la carta de Aguas subterráneas. Escala 1:250 000. Serie I. Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/>
- Instituto Nacional de Estadística y Geografía (INEGI) p. (s.f.). *Conjunto de datos vectoriales de la carta de Aguas subterráneas. Escala 1:250 000. Serie I. Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/>
- Instituto Nacional de Estadística y Geografía (INEGI) q. (s.f.). *Conjunto de datos vectoriales de la carta de Aguas superficiales. Escala 1:250 000. Serie I. Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/>
- Instituto Nacional de Estadística y Geografía (INEGI) r. (s.f.). *Conjunto de datos vectoriales de la carta de Aguas superficiales. Escala 1:250 000. Serie I. Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/>
- Instituto Nacional de Estadística y Geografía (INEGI) s. (s.f.). *Conjunto de datos vectoriales de la carta de Aguas superficiales. Escala 1:250 000. Serie I. Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/hidrologia/>
- Instituto Nacional de Estadística y Geografía (INEGI) t. (s.f.). *Conjunto de datos vectoriales de la carta de uso del suelo y vegetación escala 1:250 000 serie IV Conjunto Nacional Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI) u. (s.f.). *Conjunto de datos vectoriales de la carta de uso del suelo y vegetación escala 1:250 000 serie IV Conjunto Nacional Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>

- Instituto Nacional de Estadística y Geografía (INEGI) v. (s.f.). *Conjunto de datos vectoriales de la carta de uso del suelo y vegetación escala 1:250 000 serie IV Conjunto Nacional Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>
- Instituto Nacional de Estadística y Geografía (INEGI) w. (s.f.). *Conjunto de datos vectoriales de la carta de uso del suelo y vegetación escala 1:250 000 serie IV Conjunto Nacional Villahermosa*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>
- Instituto Nacional de Estadística y Geografía (INEGI) x. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie II. Ciudad de México*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/#descargas>
- Instituto Nacional de Estadística y Geografía (INEGI) y. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie II. Tenosique*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>
- Instituto Nacional de Estadística y Geografía (INEGI) z. (s.f.). *Conjunto de datos vectoriales de la carta de Uso del suelo y vegetación serie II. Tuxtla Gutierrez*. Obtenido de Instituto Nacional de Estadística y Geografía (INEGI): <https://www.inegi.org.mx/temas/usosuelo/>
- Instituto Nacional de Estadística y Geografía. (2020). *Diccionario de datos zonas hidrogeológicas de México*. México.
- JMP statistical discovery. (2026 a). *Diagrama de barras*. Obtenido de JMP statistical discovery: <https://www.jmp.com/es/statistics-knowledge-portal/exploratory-data-analysis/bar-chart>
- JMP statistical discovery. (2026). *Análisis exploratorio de datos*. Obtenido de JMP statistical discovery: <https://www.jmp.com/es/statistics-knowledge-portal/exploratory-data-analysis>
- JMP statistical discovery a. (2026 b). *Gráfico de dispersión*. Obtenido de JMP statistical discovery: <https://www.jmp.com/es/statistics-knowledge-portal/exploratory-data-analysis/scatter-plot#content>
- JMP statistical discovery b. (2026 c). *Histograma*. Obtenido de JMP statistical discovery: <https://www.jmp.com/es/statistics-knowledge-portal/exploratory-data-analysis/histogram>

- Kaskaya, S. (8 de Febrero de 2022). *Introduction to Fuzzy Logic Implementation with Python (Heart Disease Diagnosis Project)*. Obtenido de Medium: <https://medium.com/@sevde.kaskaya/heart-disease-diagnosis-with-fuzzy-logic-b214e27492c4>
- Kuzma, S., & Saccoccia, L. (16 de Agosto de 2023). *25 Countries, Housing One-Quarter of the Population, Face Extremely High Water Stress*. Obtenido de World Resources Institute: <https://www.wri.org/insights/highest-water-stressed-countries>
- Kuzma, S., P. Bierkens, M., Lakshman, S., Tianyi, L., Saccoccia, L., H. Sutanudjaja, E., & Van Beek, R. (16 de Agosto de 2023). *Aqueduct 4.0: Updated Decision-Relevant Global Water Risk Indicators*. Obtenido de World Resources Institute: <https://www.wri.org/research/aqueduct-40-updated-decision-relevant-global-water-risk-indicators?campaign=752713&auHash=74cRjEQPsH0NDpgT1NqIfNpqV-QpYNR4oiPo1HRhpGs>
- Madrigal, E. (27 de Enero de 2023). *La Correlación en el Análisis de Datos: Conceptos, Aplicaciones y Ejemplos con Python*. Obtenido de growup: <https://www.growupcr.com/post/correlacion-analisis-datos>
- Martínez Guerra, R. (s.f.). Recarga Vertical De Los Acuíferos por Infiltración Natural de Lluvia. Magdalena Contreras, Distrito Federal, México.
- Mazraeh, A. (11 de Febrero de 2025). *Semana 3: Fuzzificación y desfuzzificación*. Obtenido de Medium: <https://medium.com/@adnan.mazraeh1993/fuzzification-and-defuzzification-6fbfe059047e>
- Mejía Pineda, N. G., & Murillo Álvarez, M. (2025). Identificación de zonas potenciales de recarga de agua subterránea en el acuífero Siguatepeque, Honduras, mediante evaluación multicriterio. *Boletín de la Sociedad Geológica Mexicana*.
- Mejía Pineda, N. G., & Murillo Álvarez, M. (8 de Agosto de 2025). Identification of potential recharge zones of groundwater in the Siguatepeque aquifer, Honduras, by means of multicriteria evaluation. *Boletín de la Sociedad Geológica Mexicana*, 77. doi:<https://doi.org/10.18268/bsgm2025v77n1a260624>
- numiqo. (2026). *Coeficiente de correlación de Spearman*. Obtenido de numiqo: Online Statistics Calculator: <https://numiqo.es/tutorial/spearman-correlation>

- Ordoñez Gálvez, J. J. (2011). *Cartilla Técnica: Contribuyendo al desarrollo de una Cultura del Agua y la Gestión Integral de Recurso Hídrico*. Lima: Sociedad Geográfica de Lima.
- Parra, A. (2026). *¿Qué es el coeficiente de correlación de Spearman?* Obtenido de QuestionPro: <https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-spearman/>
- Pérez, I. (27 de Abril de 2020). *Acuíferos, nuestra valiosa fuente de agua* . Obtenido de Ciencia UNAM: <https://ciencia.unam.mx/leer/988/acuiferos-una-valiosa-fuente-de-agua->
- Plan de Mejora. (2022). *Histograma de Frecuencias – La guía completa*. Obtenido de Plan de Mejora: <https://www.plandemejora.com/histograma-de-frecuencias/>
- Portillo, G. (6 de Septiembre de 2024). *Cuenca endorreica: qué es, cómo se forman, su evolución y ejemplos*. Obtenido de Ecología Verde: https://ecologiaverde.elperiodico.com/cuenca-endorreica-que-es-como-se-forman-su-evolucion-y-ejemplos-5098.html#anchor_0
- Presidencia de la República. (18 de Mayo de 2026). *Programa Nacional Hídrico 2026-2030*. Obtenido de Diario Oficial de la Federación: https://www.gob.mx/cms/uploads/attachment/file/1079404/PROGRAMA_Nacional_H_drico_2026-2030.pdf
- Salgado-Albiter, A. C., Olea-Olea , S., Ramírez-Serrato, N. L., Morales-Casique , E., Ramírez-González, L., Llanos-Solis , A. G., . . . Avellán, D. R. (2026). Mapping groundwater-dependent ecosystems in a groundwater flow system: A case in central Mexico. *ELSEVIER*.
- scikit learn . (2026a). *1.10. Decision Trees*. Obtenido de scikit learn: <https://scikit-learn.org/stable/modules/tree.html>
- scikit learn. (2026b). *1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking*. Obtenido de scikit learn: <https://scikit-learn.org/stable/modules/ensemble.html>
- scikit learn. (2026c). *Random Forest Classifier*. Obtenido de scikit learn: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

- Secretaría de agricultura y fomento. (6 de Enero de 1926). *Ley Forestal 1926*.
Obtenido de Consejo Civil Mexicano para la Silvicultura Sostenible (CCMSS):
<https://www.ccmss.org.mx/wp-content/uploads/Ley-forestal-1926.pdf>
- Team Comet. (2 de Noviembre de 2022). *Why is Model Evaluation Important in Machine Learning?* Obtenido de comet:
<https://www.comet.com/site/blog/why-is-model-evaluation-important-in-machine-learning/>
- Tipos de acuíferos*. (31 de marzo de 2017). Obtenido de InTeGrate Interdisciplinary Teaching about Earth for a Sustainable Future: <https://serc.carleton.edu/176321>
- UN-Water Members and Partners. (2022). *The United Nations World Water Development Report 2022: Groundwater making the invisible visible*. París: United Nations Educational, Scientific and Cultural Organization.
- Upadhyay, A. (13 de Junio de 2020). *StandardScaler and Normalization with code and graph*. Obtenido de Medium: <https://medium.com/analytics-vidhya/standardscaler-and-normalization-with-code-and-graph-ba220025c054>
- Water Science School . (6 de Noviembre de 2018 a). *Groundwater: What is Groundwater?* Obtenido de USGS science for a changing world:
<https://www.usgs.gov/water-science-school/science/groundwater-what-groundwater#overview>
- Water Science School. (s.f.). *The Edible Aquifer* . Obtenido de USGS science for a changing world: <https://www.usgs.gov/water-science-school/edible-aquifer>
- Water Science School. (8 de Junio de 2018 b). *Pesticides in Groundwater*. Obtenido de USGS science for a changing world: <https://www.usgs.gov/water-science-school/science/pesticides-groundwater#overview>
- Water Science School. (16 de Octubre de 2019). *Aquifers and Groundwater* . Obtenido de USGS science for a changing world: <https://www.usgs.gov/water-science-school/science/aquifers-and-groundwater#overview>
- Woessner, W. W., & Poeter, E. P. (2020). *Hydrogeologic Properties of Earth materials and principles of Groundwater Flow*. Ontario: The Groundwater Project.
- Wolff, R. G. (1982). *Physical Properties of rocks-Porosity, permeability, distribution coefficients, and dispersivity*. Virginia: U.S. Geological survey.

Xu, T., & Liang, F. (2021). Machine learning for hydrologic sciences: An introductory. *Wiley Periodicals LLC.* .

Zaresefat, M., & Derakhshani, R. (2023). Revolutionizing Groundwater Management with Hybrid AI Models: A Practical Review. *Water*.