



Tecnológico de Monterrey

Desarrollo de proyectos de ingeniería matemática [REDACTED]
INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

Proyecto de Sustentabilidad

Adrián Landaverde Nava



Jesús Yair Ramírez Islas



Luis Gerardo Lagunes Najera



Michael Steven Delgado Caicedo



Cristian Gonzaga López



Evelyn Geovanna Pérez Gómez



Deborah Tirado Hernández



Fecha de Entrega:

10 de junio de 2024

1. Índice

Índice

1. Índice	1
2. Introducción	2
3. Objetivo	2
4. Trabajo relacionado	2
5. Trabajo propuesto	3
5.1. Análisis de sequías	3
5.1.1. Librerías y Datos	3
5.1.2. Limpieza de Datos	4
5.1.3. Visualización de los datos	4
5.1.4. Modelo Prophet	5
5.2. Calidad del agua	6
5.2.1. Librerías y datos	6
5.2.2. Limpieza de Datos	6
5.2.3. Visualización de Datos	7
5.2.4. Modelado	7
5.2.5. Predicción y visualización	8
5.3. Inundaciones	9
5.3.1. Identificación de exceso de agua	9
5.3.2. Librerías utilizadas	9
5.3.3. Modelo de detección de agua	9
5.3.4. Propuestas referentes al modelo	10
6. Experimentos y resultados	10
6.1. Análisis de sequías	11
6.2. Calidad del agua	11
6.3. Inundaciones	12
7. Conclusiones	13

2. Introducción

México enfrenta una serie de desafíos relacionados con el agua que amenazan tanto a sus ecosistemas como a su población. Estos desafíos incluyen la escasez de agua, el control de la contaminación, la eficiencia en la distribución, la predicción y gestión de inundaciones, y la adaptación al cambio climático. Abordar estos problemas de manera efectiva requiere enfoques innovadores, y la Inteligencia Artificial (IA) presenta soluciones prometedoras. Este proyecto tiene como objetivo aprovechar la IA para enfrentar uno de estos desafíos críticos del agua en México.

3. Objetivo

El objetivo principal de este proyecto es llevar a cabo una serie de experimentos centrados en uno de los principales desafíos del agua en México. A través de la limpieza, integración y formateo de datos, seguidos por la modelización y validación, y culminando en la visualización, el proyecto busca desarrollar y demostrar soluciones efectivas impulsadas por IA. El desafío específico a abordar será elegido de entre los siguientes:

- **Escasez de Agua:** Optimizar el uso del agua en la agricultura mediante sistemas de riego de precisión que midan y respondan a las necesidades de los cultivos.
- **Control de la Contaminación:** Mejorar la monitorización y gestión de la calidad del agua prediciendo niveles de contaminación e identificando fuentes de contaminación.
- **Eficiencia en la Distribución:** Mejorar la gestión de las redes de distribución de agua detectando fugas e ineficiencias en infraestructuras envejecidas.
- **Predicción y Gestión de Inundaciones:** Predecir inundaciones y planificar respuestas, particularmente en regiones propensas a huracanes y lluvias estacionales intensas.
- **Adaptación al Cambio Climático:** Modelar y predecir los impactos del cambio climático en los recursos hídricos para ayudar en la gestión y conservación a largo plazo.

4. Trabajo relacionado

Rosanna Bonasia, especializada en Fluidodinámica Computacional y análisis de riesgos relacionados con fenómenos naturales e ingenieriles, mediante la aplicación de modelos numéricos y análisis estadístico, se ha especializado en la aplicación de modelos matemáticos a temas relacionados con los fluidos, como el aire o el agua. En 2010, llegó a México, originaria de Italia, y trajo consigo modelos de dispersión de ceniza volcánica, para calcular sus efectos.(TecScience, 2023)

En el caso de las inundaciones, estos modelos pueden ayudar a medir la probabilidad de que sucedan en una zona determinada, además, cuantifican el riesgo de daños que pueden ocasionar y sus causas, dependiendo de la región.

En este trabajo, Bonasia se enfoca en una sola ciudad y estado de la República Mexicana para aplicar estos modelos. Villahermosa, Tabasco, fue la ciudad seleccionada para analizar debido a las inundaciones que ocurren en este lugar, por los ríos que se encuentran al rededor de esta.

La ciudad de Villahermosa, un centro logístico en la economía del estado de Tabasco, se ve afectada por las recurrentes inundaciones de los ríos. En este estudio, se analizó el impacto de dos factores que son las causas más probables de este aumento del riesgo de inundaciones: los cambios en el uso de la tierra en las calendas hidrológicas aguas arriba de la ciudad, y la urbanización incontrolada de las llanuras de inundación adyacentes a los principales canales fluviales. Se evaluaron las descargas de inundaciones para diferentes períodos de retorno, teniendo en cuenta los usos de la tierra de las zanques, tanto en 1992 como en la actualidad. Estas descargas de inundación se utilizaron en un modelo 2D de aguas poco profundas para estimar el aumento de las profundidades de agua en la ciudad desde 1992 hasta la actualidad. Para evaluar la influencia de la expansión urbana en los niveles de inundación, se propusieron tres escenarios de urbanización futuros sobre la base de la previsión de la tasa de crecimiento urbano para 2050.

Los resultados confirmaron que el cambio en el uso de la tierra en las calas hidrológicas es el principal factor que explica el aumento de los eventos de inundación observados en los últimos años. Este estudio también proporcionó información útil para la futura planificación de la ciudad que podría ayudar a minimizar el impacto de la inundación en Villahermosa.(MDPI, 2023)

5. Trabajo propuesto

En el siguiente repositorio se encuentra la solución a este trabajo, incluyendo los códigos y bases de datos.
Liga de nuestro repositorio de *Github* con la solución: <https://github.com/AdrianLandaverde/water-ai>

El presente trabajo explora la aplicación de modelos de *Machine Learning* en la gestión de los recursos hídricos en México, con un enfoque particular en: Análisis de Sequías, Gestión de Inundaciones y Monitorización de la Calidad del Agua. Mediante la integración de estos modelos, se busca mejorar la resiliencia de las comunidades mexicanas frente a los desafíos hídricos actuales y futuros, promoviendo una gestión sostenible y eficiente de los recursos acuíferos.

Se llevaron a cabo una serie de pasos para la implementación de los modelos para el análisis de los tres enfoques anteriormente mencionados: sequías, inundaciones y calidad del agua.

5.1. Análisis de sequías

En esta sección se implementa un modelo entrenado para predecir los niveles de sequía en un período futuro para los estados del país.

5.1.1. Librerías y Datos

En primer lugar, se importaron varias librerías esenciales para el análisis y visualización de datos. Se utilizó `pandas` para la manipulación de datos, `plotly.express` para la creación de gráficos interactivos,

`prophet` para la previsión de series temporales, y `geopandas` para el manejo de datos geoespaciales.

A continuación, se cargaron los datos acerca de las sequías en México desde un archivo **Excel** que contiene información detallada por municipios, y los datos geográficos de los estados desde un archivo **Shapefile**.

5.1.2. Limpieza de Datos

La limpieza y preparación de los datos implicaron varios pasos clave:

1. **Selección de columnas relevantes:** De los datos, se seleccionaron únicamente las columnas necesarias para nuestro análisis. Esto incluye identificadores de entidades y fechas específicas.
2. **Reorganización de datos:** Se reorganizaron los datos para tener una estructura más manejable, convirtiendo las columnas de fechas en filas y asignando valores numéricos a las diferentes categorías de sequía. Además, se reemplazaron los valores nulos con ceros y se aseguraron de que todos los valores fueran de tipo entero.
3. **Conversión de fechas:** Las fechas en el conjunto de datos se convirtieron al formato de fecha adecuado y se extrajeron los componentes de mes y año para facilitar el análisis temporal.
4. **Agregación de datos:** Se calculó el promedio de los niveles de sequía por estado, año y mes. Esto permitió obtener una visión más clara de cómo varía la sequía a lo largo del tiempo y en diferentes regiones.
5. **Unión con datos geográficos:** Se combinaron los datos procesados de sequía con los datos geográficos de los estados, permitiendo una integración que facilita el análisis espacial.
6. **Exportación de datos:** Finalmente, se filtraron los datos para considerar solo aquellos a partir del año 2013 y se exportaron el conjunto de datos combinado a un nuevo archivo **Shapefile**, para su uso en aplicaciones de Sistemas de Información Geográfica (GIS).

5.1.3. Visualización de los datos

En esta sección, se visualizan los datos utilizando gráficos coropléticos y herramientas de visualización geoespacial:

5.1.3.1. Preparación de los datos para su visualización:

1. Se filtraron los datos para seleccionar aquellos correspondientes a la fecha específica de 1 de enero de 2023.
2. Se restablecieron los índices del `DataFrame` para que sean consecutivos y se empezó desde cero.

5.1.3.2. Conversión a GeoJSON:

1. Se convirtió la columna de geometría del **DataFrame** a formato **GeoJSON**. Esto se hace aplicando una función que convierte los objetos de geometría en cadenas de texto **JSON**.
2. Se creó un objeto **GeoJSON**, que es una estructura de datos para representar objetos geográficos. Este objeto contiene una colección de características (features), cada una con su propia geometría y un identificador único.

5.1.3.3. Creación del Gráfico coroplético:

1. Se utilizó la librería **plotly.express** para crear un gráfico coroplético, que muestra la intensidad de la sequía en diferentes estados de México.
2. Se configura el gráfico para que utilice los datos **GeoJSON** para la geometría, las ubicaciones de los estados y los niveles de sequía para la coloración.
3. Se añade título al gráfico, se configura la escala de colores y se habilita la animación por fechas para visualizar cómo cambia la sequía a lo largo del tiempo.

5.1.3.4. Ajustes del gráfico:

1. Se hace el ajuste del gráfico para que las regiones se ajusten a los límites geográficos de las ubicaciones y se desactiva la visibilidad de los bordes geográficos.
2. Se guardó el gráfico como un archivo **HTML**, lo que permite su visualización interactiva en un navegador web.

5.1.4. Modelo Prophet

En esta sección, se utiliza el modelo **Prophet** para la previsión de series temporales:

5.1.4.1. Preparación de los datos para la aplicación del modelo:

1. Se seleccionan los datos de sequía correspondientes a un estado específico (por ejemplo, Yucatán, identificado por el código 10).
2. Se renombran las columnas para que sean compatibles con **Prophet** (Fecha a *ds* y Sequía a *y*).
3. Se asegura que el **DataFrame** contenga únicamente las columnas necesarias (*ds* y *y*).

5.1.4.2. Entrenamiento del modelo:

1. Se crea el modelo **Prophet** con parámetros específicos para la priorización de puntos de cambio y modo de estacionalidad multiplicativa.
2. Se entrena el modelo con los datos preparados.

5.1.4.3. Predicción futura:

1. Se genera un `DataFrame` futuro para realizar predicciones para los próximos 12 meses con frecuencia mensual.
2. Se utiliza el modelo entrenado para predecir los niveles de sequía en este período futuro.

5.1.4.4. Visualización de la predicción:

1. Se utilizaron las funciones de visualización de `Prophet` para generar gráficos interactivos que muestran las predicciones.
2. Se configuraron los gráficos para incluir un título y etiquetas en los ejes, proporcionando una visión clara de la evolución y las predicciones de la sequía, un ejemplo el estado de Yucatán.

5.2. Calidad del agua

En esta sección se implementa un modelo para la visualización detallada y análisis de la calidad del agua subterránea.

5.2.1. Librerías y datos

En esta etapa, importamos varias librerías esenciales para el análisis y visualización de datos:

- `Pandas`, para la manipulación y análisis de datos.
- `Folium`, para la creación de mapas interactivos.
- `Plotly.express` permite la visualización de datos interactivos.
- `Sklearn.manifold.TSNE`, para la reducción de dimensionalidad y visualización.
- `Matplotlib.pyplot`, para gráficos estáticos.
- `Geopandas` funciona para el manejo de datos geoespaciales.
- `Pykrige.ok.OrdinaryKriging` y `Pykrige.uk.UniversalKriging`, para la interpolación espacial.

Se configura `pandas` para manejar conversiones de tipo de datos de manera silenciosa y se cargan los datos acerca de la calidad del agua subterránea en el país desde un archivo `Excel`, así como los datos geoespaciales desde un archivo `Shapefile`.

5.2.2. Limpieza de Datos

La limpieza y preparación de los datos involucraron varios pasos:

1. **Reemplazo y conversión de valores:** Se reemplazó los valores «.en el `DataFrame` y se convirtieron varias columnas a tipo `float` para asegurar un análisis correcto.

2. **Manejo de valores nulos:** Para cada parámetro de calidad del agua, se hace el reemplazo de los valores nulos con el valor mínimo de la respectiva columna.
3. **Conversión de códigos de calidad a números:** Se reemplazaron las descripciones de calidad del agua por valores numéricos para facilitar el análisis. Cada descripción de calidad se convirtió en una escala numérica según su gravedad o impacto.
4. **Normalización de datos:** Se hizo la normalización de las variables y calidades de agua utilizando la escala $[0, 1]$. Esta normalización permite comparar diferentes parámetros en una escala común.

5.2.3. Visualización de Datos

En esta etapa, se visualizan los datos de diferentes maneras:

- **Visualización de la distribución de variables:** Se utilizó `plotly.express` para crear diagramas de caja (box plots) que muestran la distribución de diferentes variables de calidad del agua. Los gráficos se organizan en una cuadrícula con varias columnas y filas para facilitar la comparación visual.
- **Visualización de la distribución de calidades:** Se crearon diagramas de caja similares para las diferentes calidades del agua, permitiendo observar cómo varían las evaluaciones de calidad entre las distintas métricas.
- **Creación de un mapa interactivo:** Se utilizó `folium` para crear un mapa centrado en las coordenadas medias de los puntos de muestreo. Además, se definió un diccionario de colores para representar visualmente los niveles de calidad del agua con un semáforo: verde (calidad buena), amarillo (calidad media) y rojo (calidad mala). Se añadieron marcadores circulares en el mapa para cada punto de muestreo, coloreados según la evaluación del semáforo, cada marcador también incluye información de la latitud y longitud de los puntos de muestreo.

5.2.4. Modelado

Reducción de dimensionalidad con TSNE

TSNE (t-Distributed Stochastic Neighbor Embedding) es una técnica de reducción de dimensionalidad que se utiliza para visualizar datos de alta dimensión en un espacio de menor dimensión (2D o 3D). Este método se aplica a dos conjuntos de datos: variables numéricas y variables de calidad.

- **TSNE en variables numéricas:** Primero, se aplica TSNE a las variables numéricas normalizadas para reducir las dimensiones a un espacio 2D. Los resultados se almacenan en un `DataFrame` con columnas 'x' y 'y'. Se añade una columna 'SEMÁFORO' para clasificar visualmente los puntos según la calidad del agua. Se utilizó `plotly.express` para crear un diagrama de dispersión donde los puntos están coloreados según su clasificación en el semáforo.

- **TSNE en variables de calidad:** De manera similar, se aplica TSNE a las variables de calidad normalizadas. Los resultados se visualizan en un diagrama de dispersión con las mismas clasificaciones de semáforo. Ambos gráficos se guardan como archivos HTML para facilitar su visualización interactiva.

Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) es un método de *Clustering* basado en densidad que encuentra grupos de puntos en un conjunto de datos.

- **Aplicación de DBSCAN:**
 - Se utiliza DBSCAN para identificar *clústeres* en el espacio TSNE resultante de las variables de calidad.
 - Los resultados del *Clustering* se añaden al `DataFrame` original, incluyendo tanto el número de *clúster* como la clasificación de semáforo.

- **Visualización de *clústeres***

Se crea un gráfico de dispersión para visualizar los *clústeres* en el espacio TSNE. Un gráfico de *sunburst* muestra la distribución de semáforos dentro de cada *clúster*.

- **Mapa interactivo con *clústeres***

Los *clústeres* se representan en un mapa interactivo utilizando `folium`. Cada punto en el mapa está coloreado según el promedio del semáforo dentro de su *clúster*.

Interpolación de *Kriging* Ordinario

Kriging Ordinario es un método de interpolación geoespacial que predice valores en puntos no muestreados basándose en la distancia y la correlación entre los puntos muestreados.

- **Configuración de *Kriging*** Se utilizaron las coordenadas y clasificaciones de semáforo para entrenar un modelo de *Kriging* Ordinario. Se definió una cuadrícula de puntos sobre la cual se evaluó el modelo de *Kriging*.

5.2.5. Predicción y visualización

El modelo de *Kriging* se usa para predecir los valores del semáforo en toda la cuadrícula. Se crearon gráficos de contorno para visualizar la interpolación sobre un mapa, mostrando las áreas con diferentes calidades de agua.

- **Mapas de interpolación**

Un primer gráfico de contorno muestra la interpolación de *Kriging* sobre la región de estudio. Un segundo gráfico añade los puntos de datos originales sobre el mapa de interpolación, permitiendo comparar las predicciones del modelo con los datos reales. Estos pasos permiten una visualización detallada y análisis de la calidad del agua subterránea, tanto en términos de reducción de dimensionalidad como de identificación de patrones geoespaciales a través de clustering e interpolación.

5.3. Inundaciones

5.3.1. Identificación de exceso de agua

Se realizó la identificación de inundaciones mediante un **dataset** de imágenes donde se tenía una ruta para las máscaras y otra para las tomas originales. Para entender mejor el concepto de máscara, se pueden visualizar imágenes en blanco y negro que son la representación en estas dos tonalidades de la imagen real, en una máscara binaria en blanco y negro, los píxeles blancos (valor 255 o 1) indican las áreas de interés (por ejemplo, el objeto que quieres segmentar), mientras que los píxeles negros (valor 0) representan el fondo o las áreas que no interesan.

5.3.2. Librerías utilizadas

Las siguientes librerías se utilizan comúnmente en la implementación de Redes Neuronales Convolucionales (CNN).

`Os` se usa para la gestión de directorios y archivos del sistema operativo, `numpy` para operaciones matemáticas y manejo de arreglos, y `pandas` para la manipulación y análisis de datos. `Tensorflow` y su módulo `keras` proporcionan herramientas y estructuras para construir y entrenar modelos de aprendizaje profundo. `Tqdm` se usa para mostrar barras de progreso en bucles. `Glob` para buscar archivos que coincidan con un patrón especificado. `Tensorflow.image` contiene funciones para manipulación de imágenes y `tensorflow.keras.utils` proporciona utilidades como `load_img` e `img_to_array` para cargar y convertir imágenes a arreglos `numpy`. `Matplotlib.pyplot` es una librería de visualización que permite la creación de gráficos.

Para la construcción de modelos, `keras.models.Model` y `keras.models.Sequential` se utilizan para definir modelos secuenciales y funcionales, mientras que `keras.layers` incluye capas como `Input`, `Conv2D`, `Concatenate`, `UpSampling2D`, `AveragePooling2D`, `BatchNormalization`, `LeakyReLU` y `ReLU` para construir las arquitecturas de las redes. `Tensorflow.keras.applications.ResNet50` se usa como modelo pre-entrenado para funciones de extracción de características. `Tensorflow.keras.utils.plot_model` permite visualizar la arquitectura del modelo. El optimizador `Adam` es usado para la optimización del modelo, y `keras.callbacks` incluye herramientas como `ModelCheckpoint` y `Callback` para manejar el guardado y control del entrenamiento. Finalmente, `tf_explain.core.grad_cam.GradCAM` se utiliza para la visualización de capas del modelo a través de técnicas como Grad-CAM.

5.3.3. Modelo de detección de agua

El modelo implementado es una variante del modelo *DeepLab V3*, diseñado para la segmentación semántica de imágenes, específicamente adaptado para la detección de áreas de inundación. La arquitectura del modelo se puede describir de la siguiente manera:

- **Capa de Entrada:** La entrada del modelo es una imagen de tamaño 256×256 con 3 canales de color (RGB).

- **Backbone:** Se utiliza **ResNet50** preentrenado en **ImageNet** como red base para extraer características de alto nivel de las imágenes. En particular, se toman las características de la capa `conv4_block6_2_relu`.
- **Atrous Spatial Pyramid Pooling (ASPP):** Para captar información de contexto a diferentes escalas, se implementa un módulo ASPP que consta de:
 - **Image Pooling:** Se realiza un **AveragePooling2D** global, seguido de una convolución (**ConvBlock**) y un **UpSampling2D** para ajustar el tamaño de la característica de entrada.
 - **Convoluciones Atrous:** Se aplican convoluciones con diferentes tasas de dilatación (1, 6, 12, y 18) para captar características a diferentes resoluciones.
 - **Concatenación:** Las salidas de las capas anteriores se concatenan y se pasa por una convolución adicional (**ConvBlock**).
- **Características de bajo nivel:** Se toman características adicionales de una capa más baja de la red base (`conv2_block3_2_relu`) y se pasan por un bloque de convolución (**ConvBlock**).
- **Combinación de características:** Las características de bajo nivel y las características procesadas por el módulo ASPP se concatenan y pasan por varias capas de convolución (**ConvBlock**).
- **Upsampling:** La salida se pasa por un **UpSampling2D** para restaurar el tamaño original de la imagen de entrada.
- **Capa de Salida:** Finalmente, una capa de convolución con un filtro y activación **sigmoid** genera el mapa de segmentación binario, indicando las áreas de interés (inundación).

El modelo se compila con la función de pérdida `binary_crossentropy` y el optimizador **Adam**. Para el entrenamiento, se utilizan *callbacks* como **ModelCheckpoint** para guardar el mejor modelo y **ShowProgress** para monitorear el progreso del entrenamiento. El modelo se entrena en imágenes de entrenamiento y sus respectivas máscaras de segmentación, con un conjunto de validación para evaluar su rendimiento durante el entrenamiento.

5.3.4. Propuestas referentes al modelo

Esta forma de hacer la detección de agua puede llegar a ayudar a sectores de la población donde se conoce que existen inundaciones, logrando alertar a su brevedad a agentes de la policía, al igual que tomar el dato de la región para tomar medidas preventivas sobre inundaciones en la zona.

6. Experimentos y resultados

Se muestran los resultados obtenidos en los modelos implementados para el análisis de sequías, gestión de inundaciones y monitorización de la calidad del agua.

6.1. Análisis de sequías

Como se mencionó anteriormente, se utilizó el modelo **Prophet** para la previsión de series temporales. El modelo entrenado predice los niveles de sequía para un período futuro, proporcionando una visión clara de la evolución y las predicciones de la sequía, para un estado en específico.

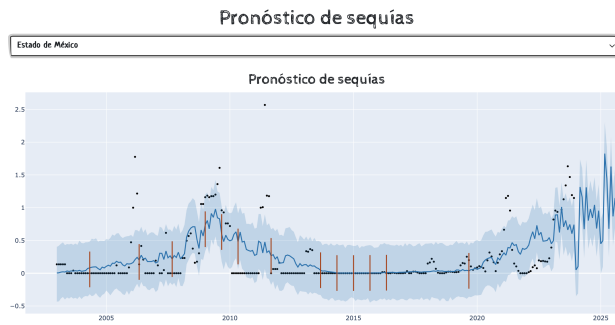


Figura 1: Pronóstico de sequías para el Estado de México

Como se muestra en la imagen, se muestra la predicción de sequía para el Estado de México, se puede observar que los niveles de sequías irán aumentando para el siguiente año.

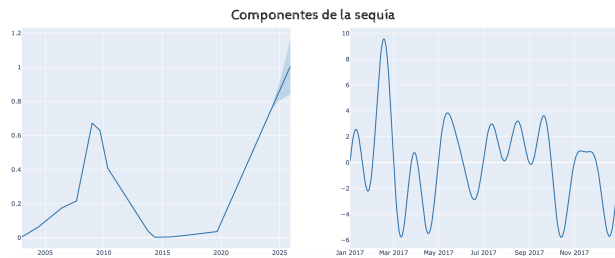


Figura 2: Componentes de la sequía

6.2. Calidad del agua

Los puntos en el mapa están agrupados en diferentes colores (verde, amarillo, naranja, rojo) que representan distintos *clústeres*. En este mapa, se observan las regiones en las que la calidad del agua es excelente o muy mala. El color verde representa buena calidad y así, hasta el rojo de muy mala calidad, siendo aguas residuales.

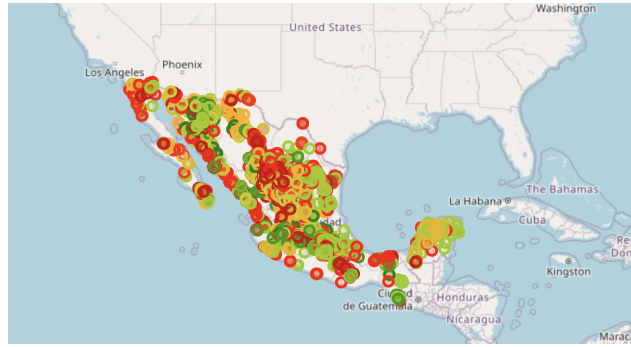


Figura 3: *Clustering*

Por ejemplo se puede observar, en el siguiente mapa, que la región del sur, incluyendo estados como Quintana Roo, Yucatán y Campeche, el nivel de la calidad del agua es bastante bueno. Mientras que regiones de los estados de Chihuahua, Zacatecas y Durango, la calidad del agua llega a ser regular y muy mala.

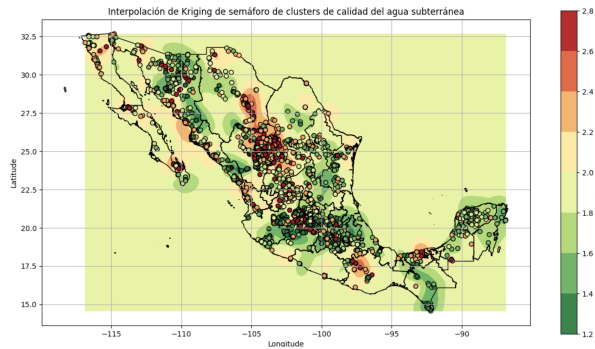


Figura 4: Interpolación de *Kriging* de semáforo de *clústeres* de calidad del agua subterránea

6.3. Inundaciones

Estos son los resultados obtenidos sobre el modelo de detección de imágenes. Se puede observar un `loss` y un `val_loss` bastante bajos, lo que indica que es un buen modelo, tal que se puede observar claramente qué es agua dentro de una imagen y lo que no, así saber si existe o no una inundación en la zona.

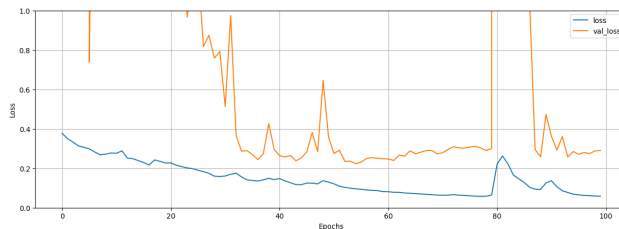


Figura 5: `loss` comparado contra el `val_loss`

Esto mismo lo podemos observar con la comparación entre la imagen, la máscara y la máscara predicha por el modelo, como se observa a continuación.

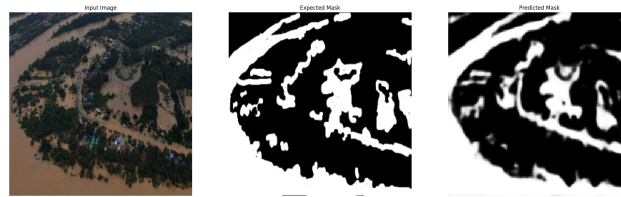


Figura 6: Imagen original, máscara y máscara predicha

7. Conclusiones

El análisis de la calidad del agua subterránea en México, a través de diversas técnicas de procesamiento y modelado de datos, ofrece una visión comprensible y detallada de la situación actual. La visualización de las distribuciones de las variables mediante diagramas de caja ayuda a identificar la dispersión y posibles valores atípicos, facilitando la comprensión de la variabilidad de cada parámetro. La creación de mapas interactivos permite localizar espacialmente las mediciones y clasificar visualmente las áreas según un sistema de semáforo, identificando rápidamente regiones con problemas potenciales. La reducción de dimensionalidad con TSNE ofrece una visualización manejable de los datos de calidad del agua en un espacio 2D, permitiendo observar patrones y clústeres que indican agrupaciones naturales en diferentes regiones. El uso de DBSCAN en los datos TSNE muestra cómo se agrupan las mediciones de calidad del agua en distintas categorías, detectando áreas con similitudes y posibles problemas.

La interpolación espacial mediante *Kriging* Ordinario predice valores en puntos no muestreados, creando una superficie continua de calidad del agua y destacando regiones que requieren atención especial. Los mapas de contorno validan la precisión del modelo y resaltan discrepancias significativas.

Es esencial continuar el monitoreo de la calidad del agua subterránea, especialmente en áreas problemáticas, y utilizar estos resultados para informar a los responsables de políticas y gestión del agua, implementando medidas correctivas y preventivas. Además, se recomienda realizar estudios adicionales con técnicas avanzadas y datos más recientes para mejorar la precisión y aplicabilidad de los resultados. En resumen, este análisis integral proporciona una base sólida para la evaluación y gestión de la calidad del agua subterránea en México, utilizando herramientas avanzadas de visualización y modelado de datos para obtener una comprensión profunda de la situación actual.

Referencias

- [1] BONASIA, R. (2019). *Impacto del crecimiento urbano y los cambios en el uso de la tierra en el peligro de las inundaciones de los ríos en Villahermosa, Tabasco (México)*. [MDPI]. <https://www.mdpi.com/2073-4441/11/2/304>
- [2] GUTIÉRREZ, I. (2023). *Modelos matemáticos para prevenir inundaciones*. <https://tecscience.tec.mx/es/humano-social/usar-modelos-matematicos-para-prevenir-inundaciones/>

- [3] ITESM. (s.f). *Escuela de Ingeniería y Ciencias*. <https://centrodelagua.tec.mx/es/semblanza/dra-rosanna-bonasia>