

Fundamentos del Aprendizaje Automático para las Ciencias de la Tierra

Parte 3: Ajustes de Modelos, Optimización de Parámetros y Algoritmos de Aprendizaje Automático Adicionales

Instructores: Jordan A. Caraballo-Vega, Mark L. Carroll, Jules Kouatchou, Jian Li, Caleb S. Spradlin
4 de mayo de 2023



Objetivos de esta Capacitación

Al final de la capacitación, los participantes podrán

- Reconocer los métodos de aprendizaje automático más comúnmente utilizados para procesar datos de observación de la Tierra
- Describir los beneficios y las limitaciones del aprendizaje automático para el análisis de datos de observación de la Tierra
- Explicar cómo aplicar algoritmos y técnicas de aprendizaje automático básicos de manera significativa a datos de teledetección
- Usar datos de entrenamiento para evaluar las condiciones y soluciones para un estudio de caso determinado
- Completar los procedimientos básicos para interpretar, refinar y evaluar la precisión de los resultados del análisis de aprendizaje automático



Recordatorio- Pre-requisitos

- Prerrequisitos:
 - La Sesión 1 de nuestra serie disponible a pedido, Fundamentos de la Percepción Remota (Teledetección) o contar con experiencia equivalente (https://appliedsciences.nasa.gov/sites/default/files/2023-02/Fundamentals_of_RS_Span.pdf).
 - Los participantes necesitarán tener acceso a Google Drive y Google Colab. Para acceder a estos recursos, deben utilizar un correo que termine en 'gmail.com'.
 - Pondremos la grabación de esta sesión a su disposición dentro de 48 horas después de la presentación.



Agenda

Parte 1:
Introducción al
Aprendizaje
Automático

20 de abril de 2023

Parte 2:
Ejemplo de Datos
de Entrenamiento y
Clasificación de la
Cobertura Terrestre

27 de abril de 2023

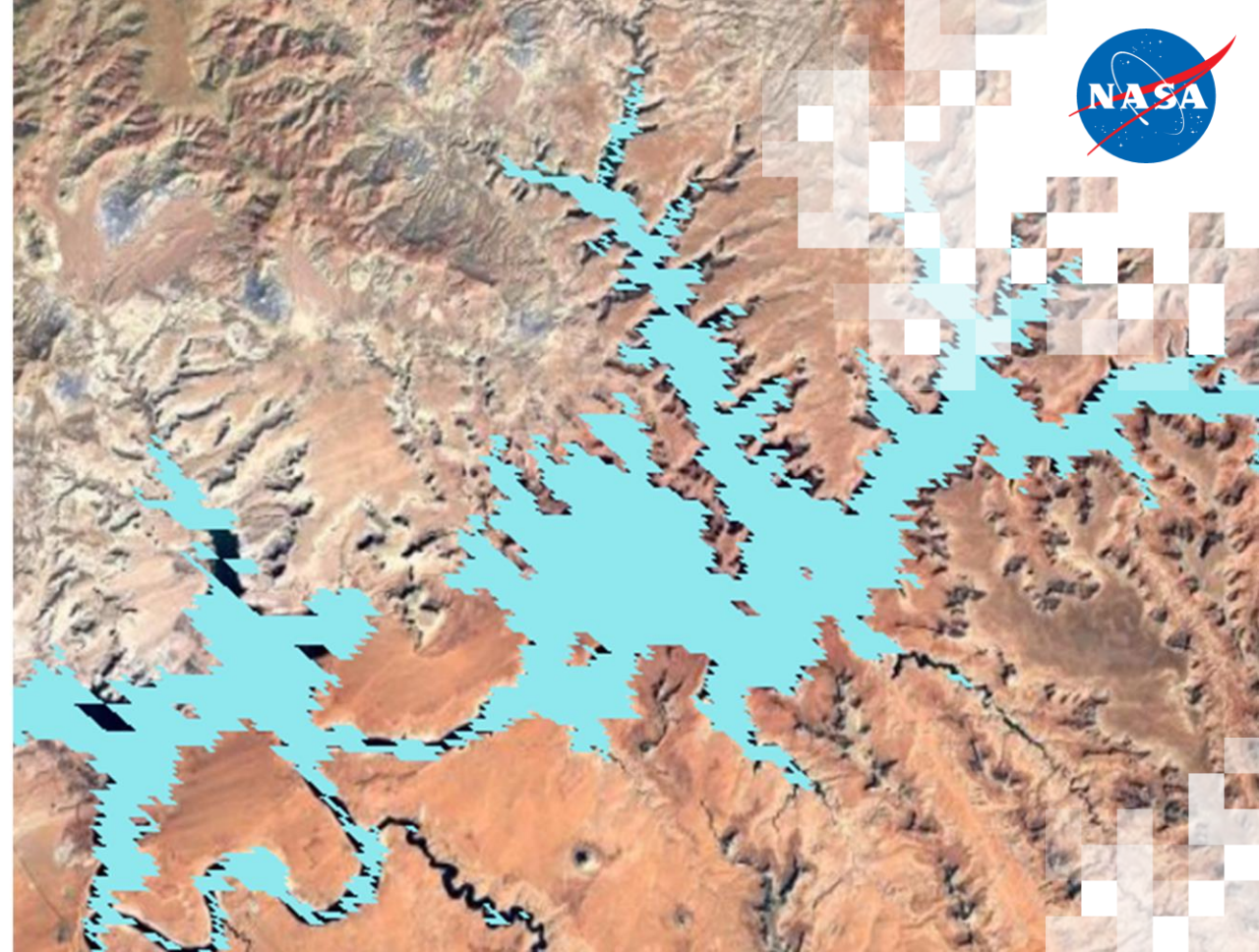
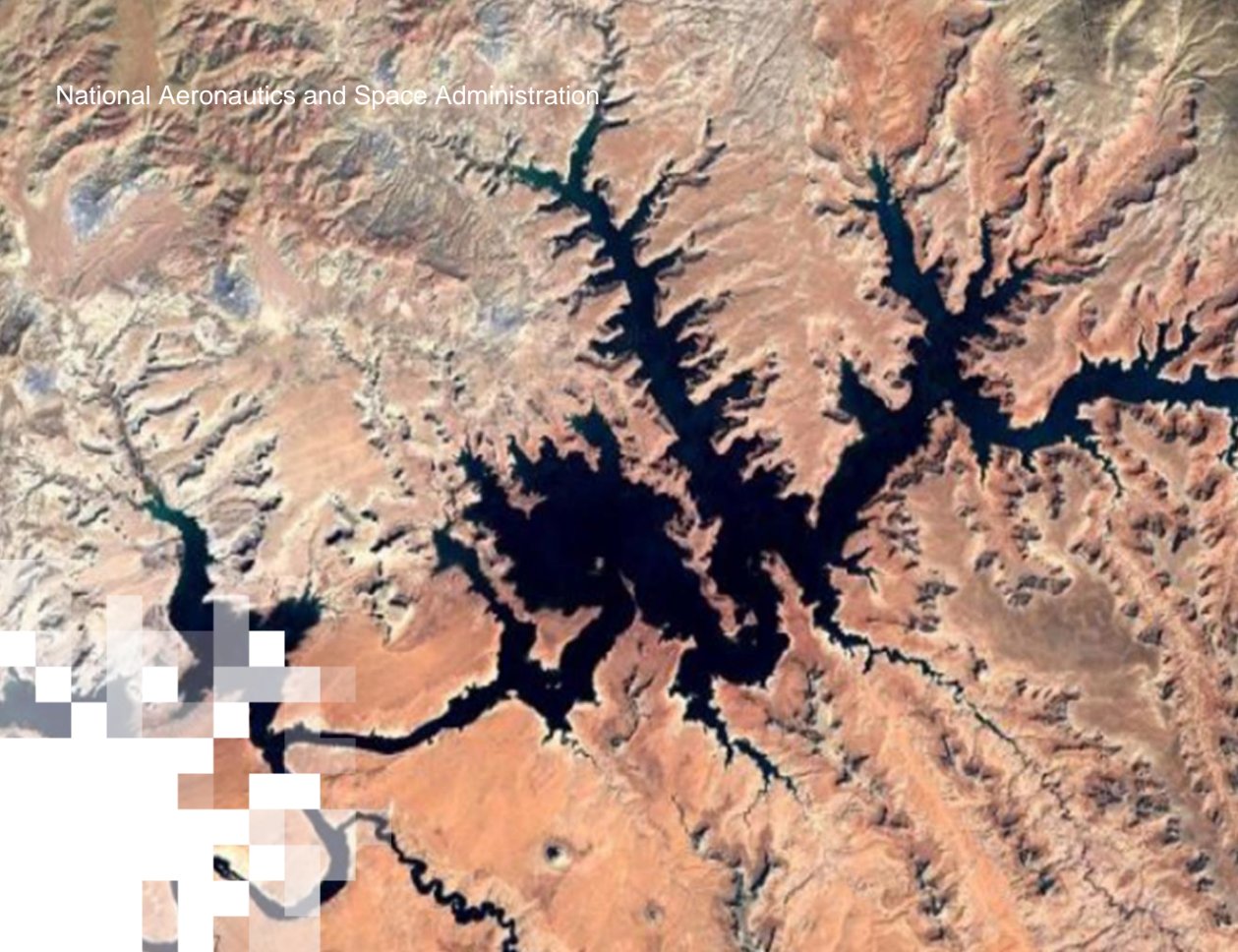
Parte 3:
Ajustes de Modelos,
Optimización de
Parámetros y
Algoritmos de
Aprendizaje
Automático
Adicionales

4 de mayo de 2023

Tarea
Práctica y
aplicación
independientes

Fecha límite: 19 de mayo
Disponible: 4 de mayo





Fundamentos del Aprendizaje Automático para las Ciencias de la Tierra

Parte 3: Ajustes de Modelos, Optimización de Parámetros y Algoritmos de Aprendizaje Automático Adicionales

Instructores: Jordan A. Caraballo-Vega, Mark L. Carroll, Jules Kouatchou, Jian Li, Caleb S. Spradlin
4 de mayo de 2023



Esquema de la Sesión 3

- Ajuste de modelos
- Introducción a la optimización de parámetros
- Ejercicio para optimizar un modelo existente
- Visión general de la Explicabilidad e interpretabilidad de los modelos
- Visión general de los algoritmos de aprendizaje automático
- Ejercicio práctico en Jupyter Notebook: Mejoras al Modelo de la Clasificación del Agua de MODIS
- Tarea para después de la clase
- Sesión de preguntas y respuestas

Recursos para esta capacitación

https://github.com/NASAARSET/ARSET_ML_Fundamentals

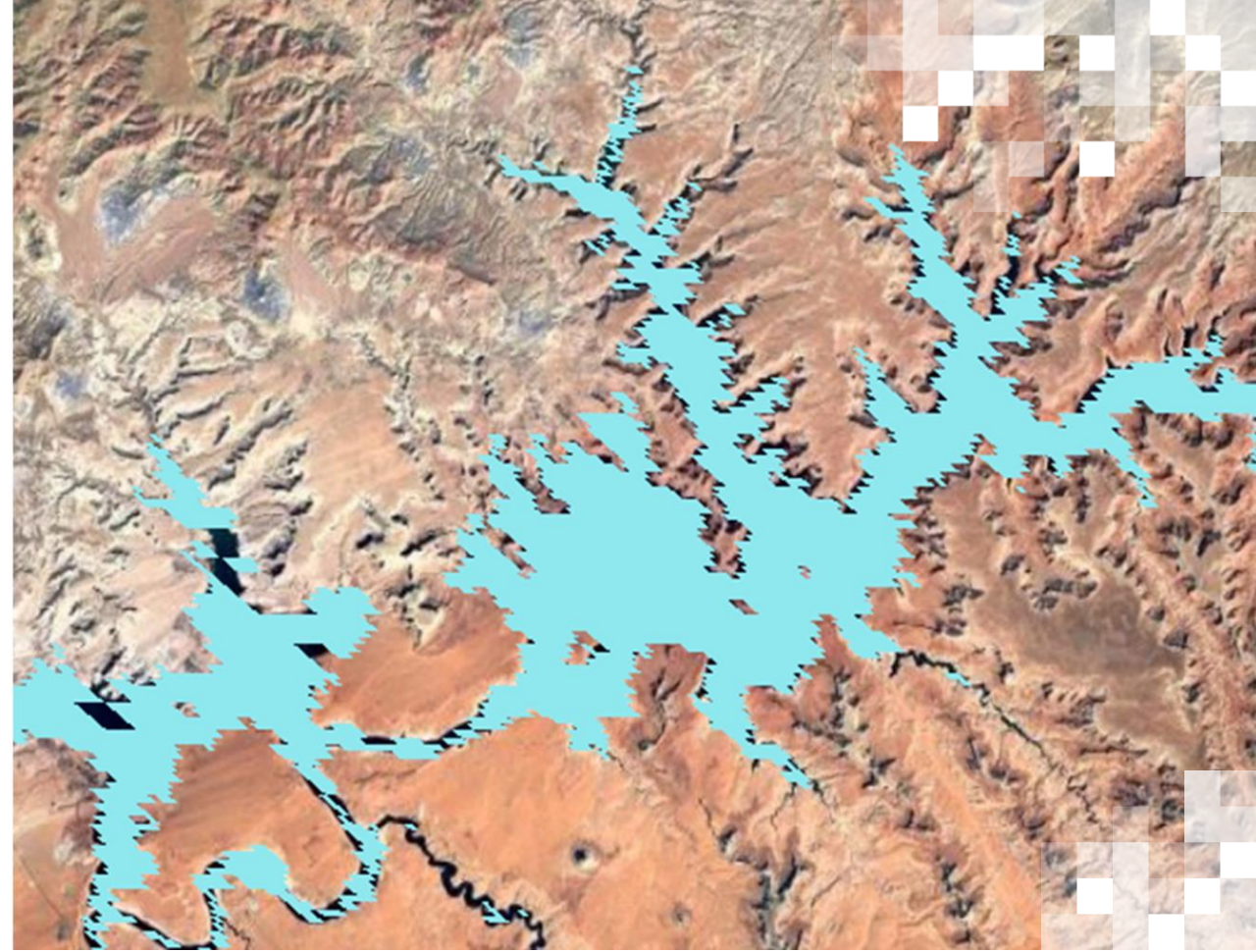
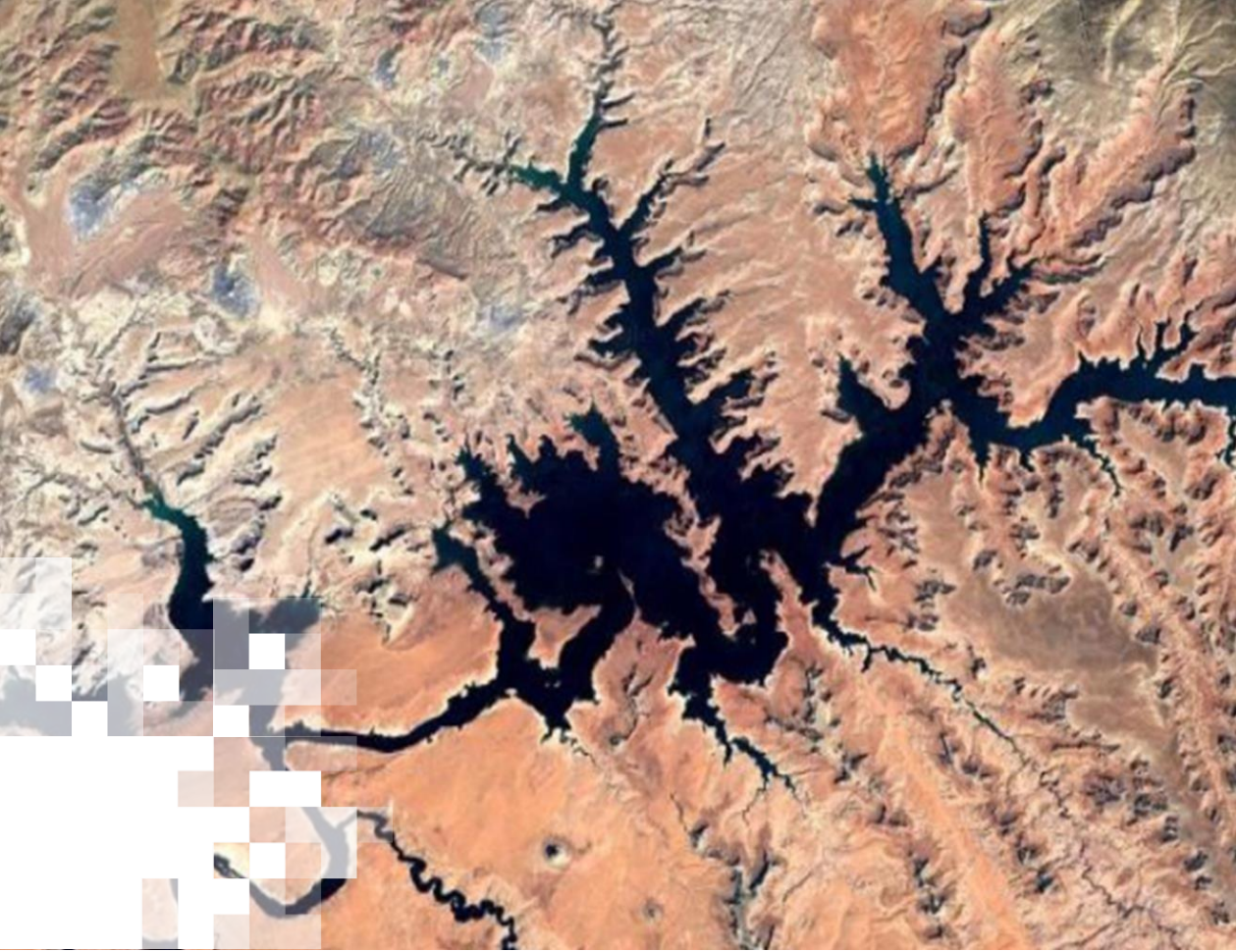


Objetivos de Capacitación

Después de participar en esta capacitación, quienes asistieron podrán:

- Reconocer los métodos de optimización más comunes para los algoritmos de aprendizaje automático que se utilizan para el procesamiento de datos de Ciencias de la Tierra
- Describir los beneficios y las limitaciones de la optimización de los algoritmos de aprendizaje automático para el análisis de Ciencias de la Tierra
- Explicar cómo aplicar técnicas de inteligencia artificial para algoritmos de aprendizaje automático de manera significativa a datos de teledetección



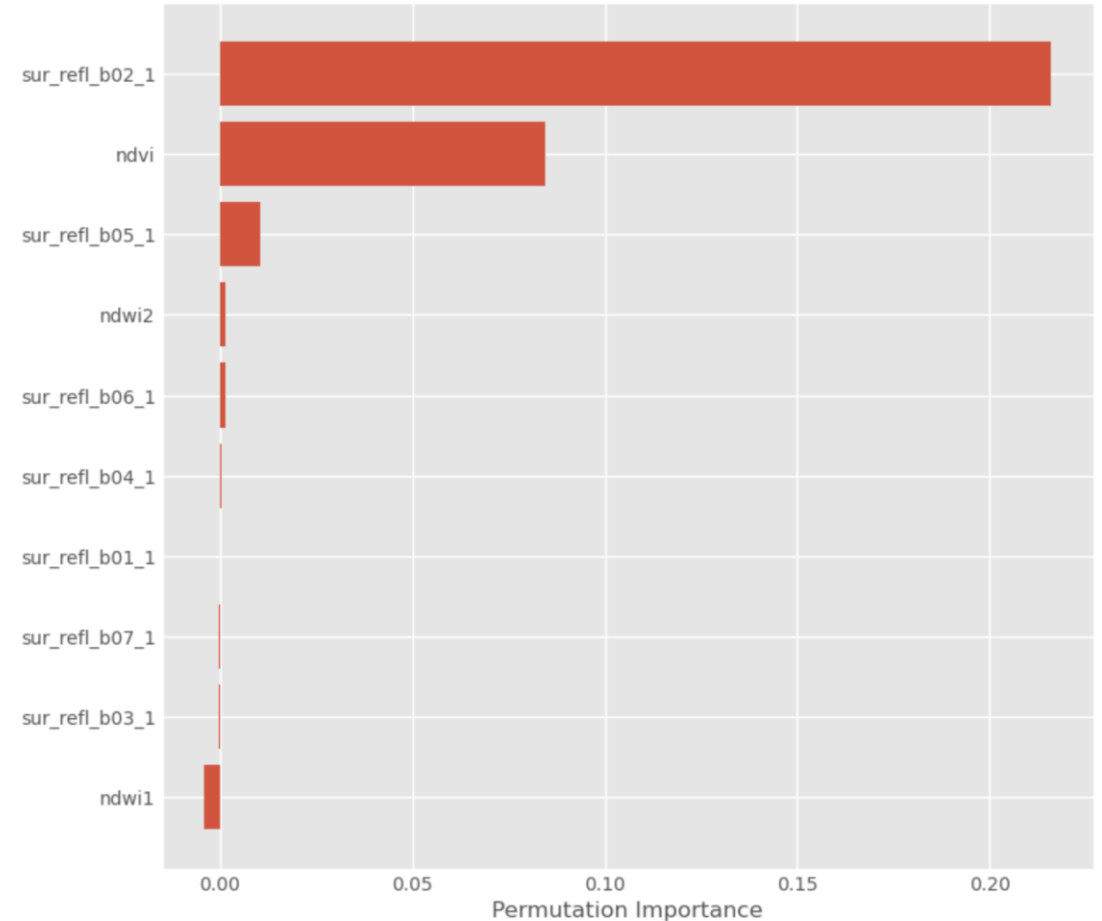


Ajuste y Optimización de Modelos

Formador: Jordan A. Caraballo-Vega

Resumen de la Sesión 2

- Entrenamos un modelo Random Forest que realiza una clasificación binaria de píxeles con agua y sin agua utilizando datos de reflectancia superficial de MODIS.
- Evaluamos nuestro modelo utilizando nuestro conjunto de datos de prueba y realizamos inferencias utilizando mosaicos ráster.
- Nuestro modelo pudo identificar con éxito píxeles de agua, pero también identificamos ubicaciones donde hace falta mejorar el modelo.

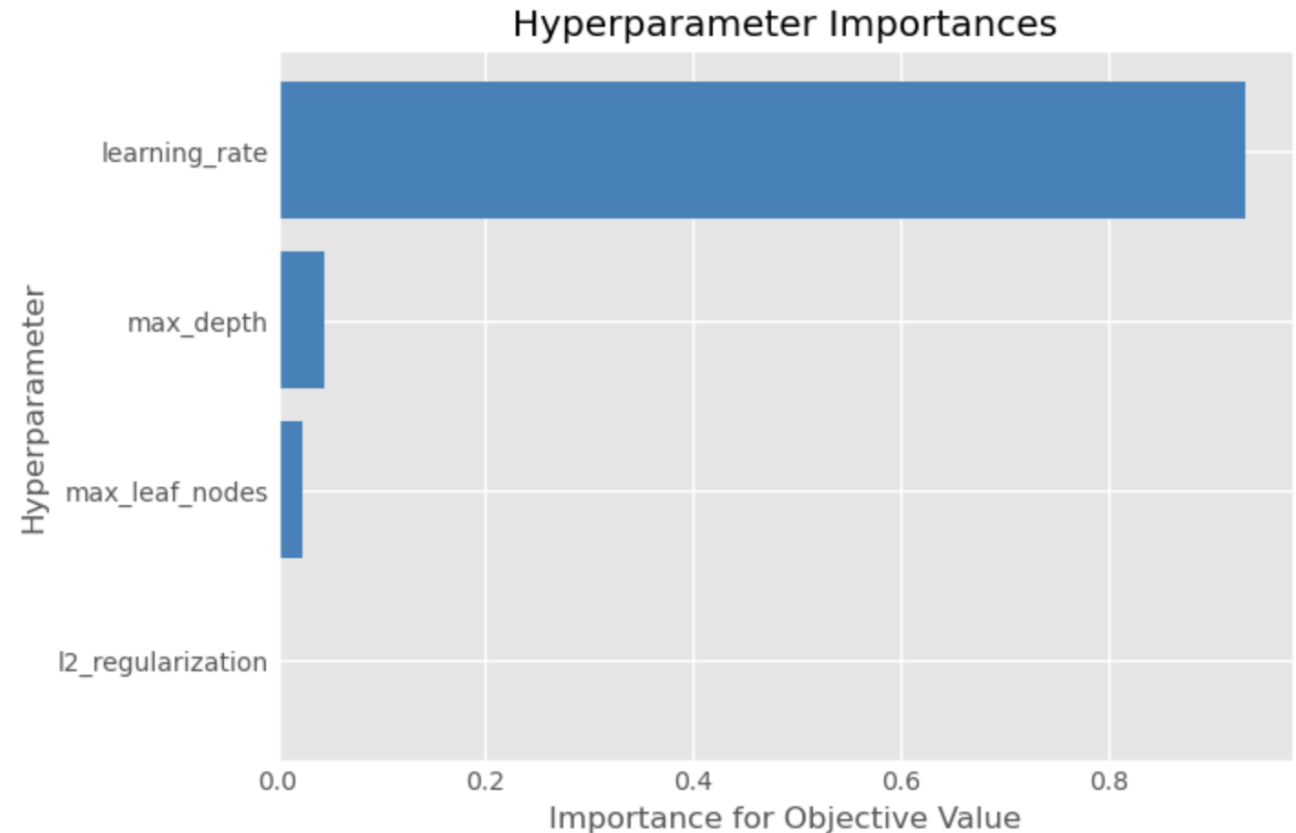


Bandas de Entrada Utilizadas en el Entrenamiento del Modelo



Ajuste y Optimización de Modelos

- Su primer modelo no siempre será el más preciso.
- La mayoría de las interfaces de programación de aplicaciones (Application Programming Interface o API) de los modelos de ML incluirán hiperparámetros predeterminados para inicializar su modelo, lo que puede servir de punto de partida.
- Estos hiperparámetros se establecen antes de que comience el proceso de aprendizaje e influirán en la convergencia del modelo.
- Tenemos varias opciones para mejorar nuestro modelo antes de agregar o generar más datos de entrenamiento.

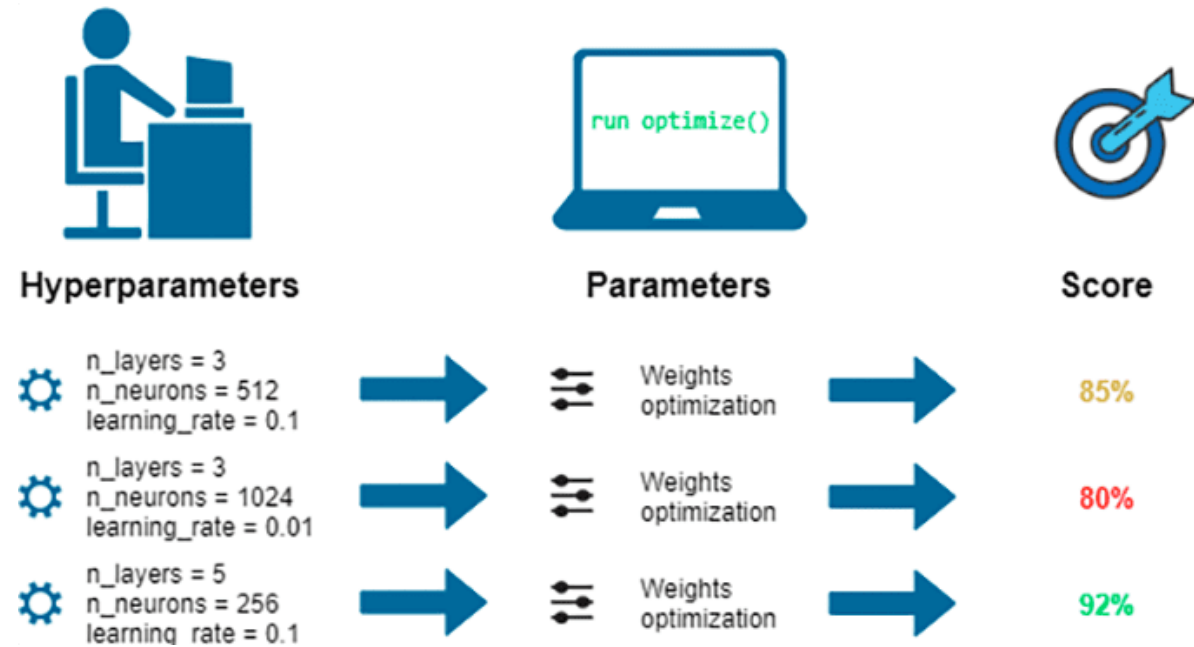


Importancias de Hiperparámetros



Ajuste y Optimización de Modelos, Continuación

- La mayor parte del tiempo en que no estamos optimizando nuestros modelos para las características más comunes, estamos intentando mejorar nuestro modelo para que pueda generalizarse a través de características poco comunes.
- Al ajustar el modelo, podemos maximizar su rendimiento y obtener la mayor tasa de rendimiento posible.
- El ajuste de modelos en general es el proceso experimental de encontrar los valores óptimos de los hiperparámetros para maximizar el rendimiento del modelo.



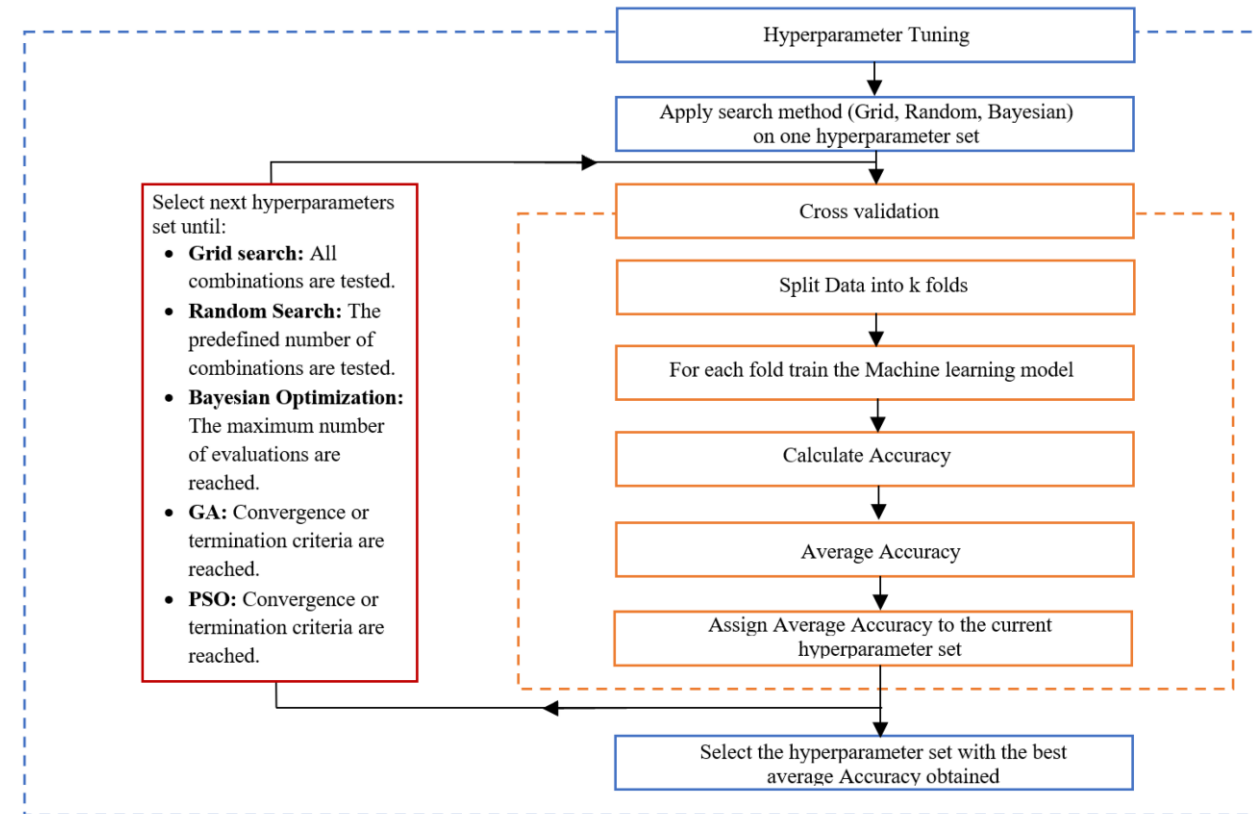
Esta es una métrica para la cual vamos a querer optimizar.

Fuente de la Imagen: medium.com



Técnicas para el Ajuste de Modelos

- Se debe identificar y establecer un criterio de evaluación sólido antes del ajuste del modelo para optimizar los parámetros de ajuste hacia el objetivo específico.
- **Ajuste manual del modelo:** los valores de los hiperparámetros se establecen en función de la intuición o la experiencia pasada. Luego, el modelo se entrena y evalúa para determinar el rendimiento utilizando el conjunto respectivo de hiperparámetros.
- **Ajuste de modelos automatizado:** los valores óptimos de los hiperparámetros se encuentran a través de algoritmos. Aquí, definimos un espacio de búsqueda de hiperparámetros desde el cual se selecciona el conjunto óptimo de valores de hiperparámetros.

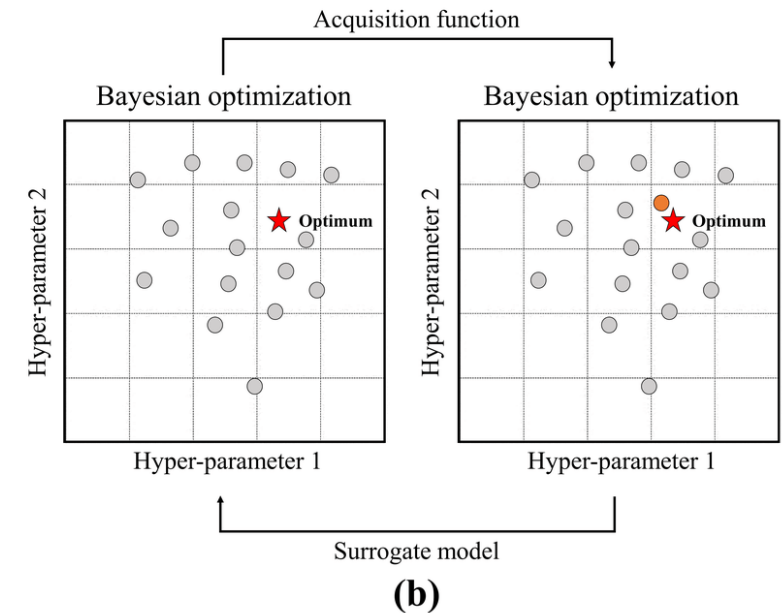
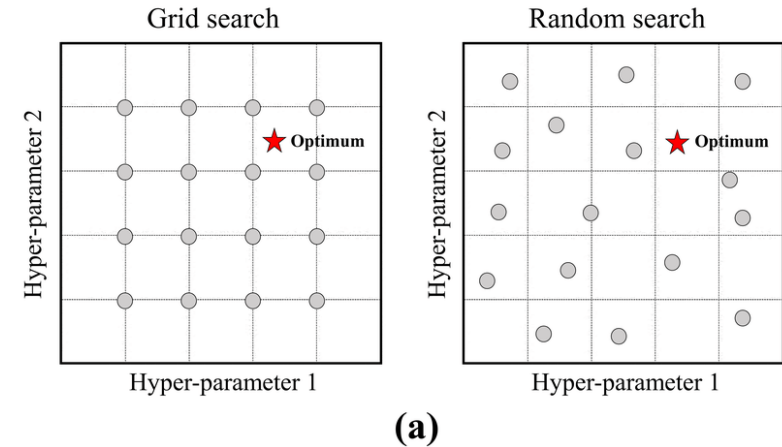


Flujo de trabajo del ajuste de hiperparámetros. Hyperopt es una herramienta excelente para el ajuste de modelos automatizado.



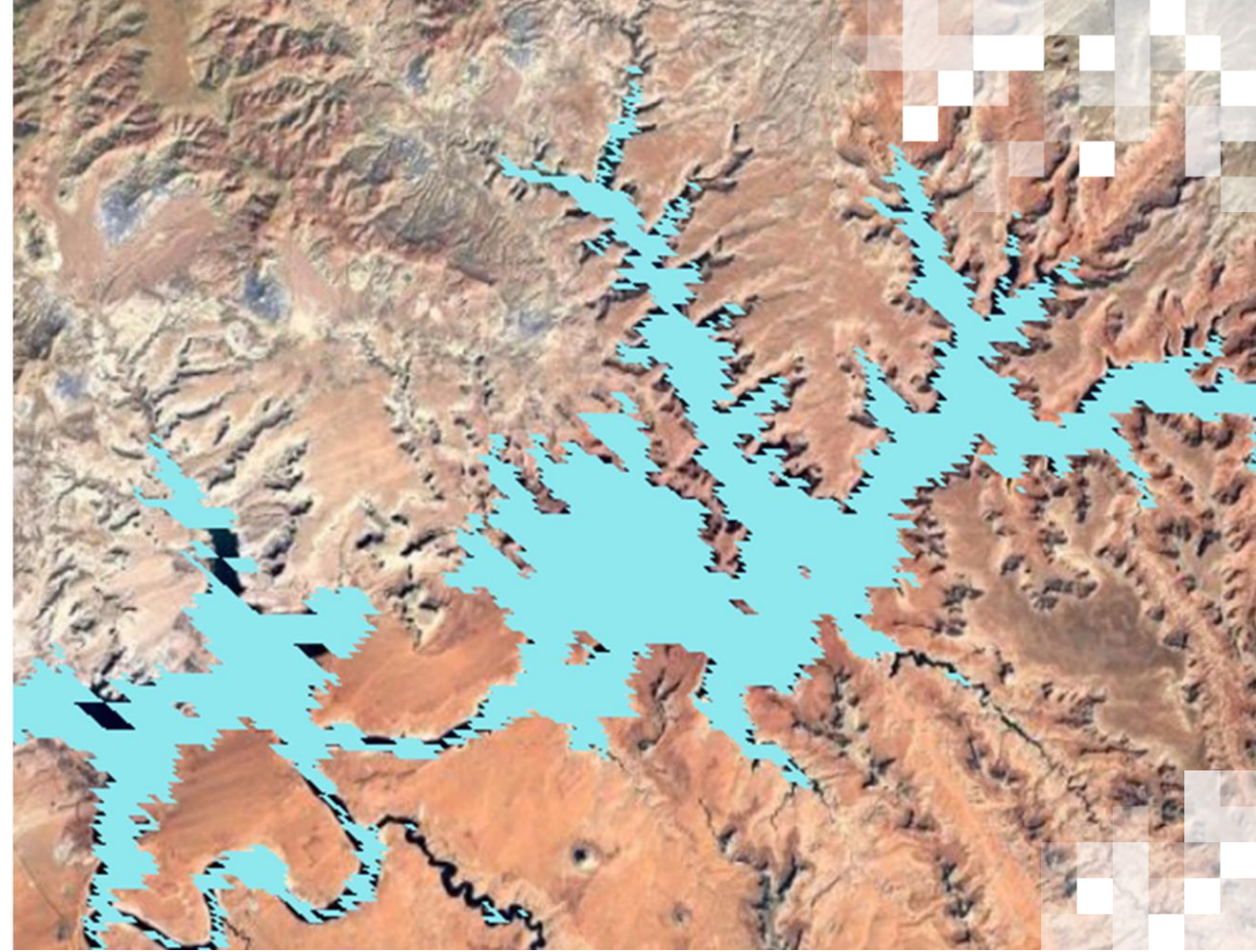
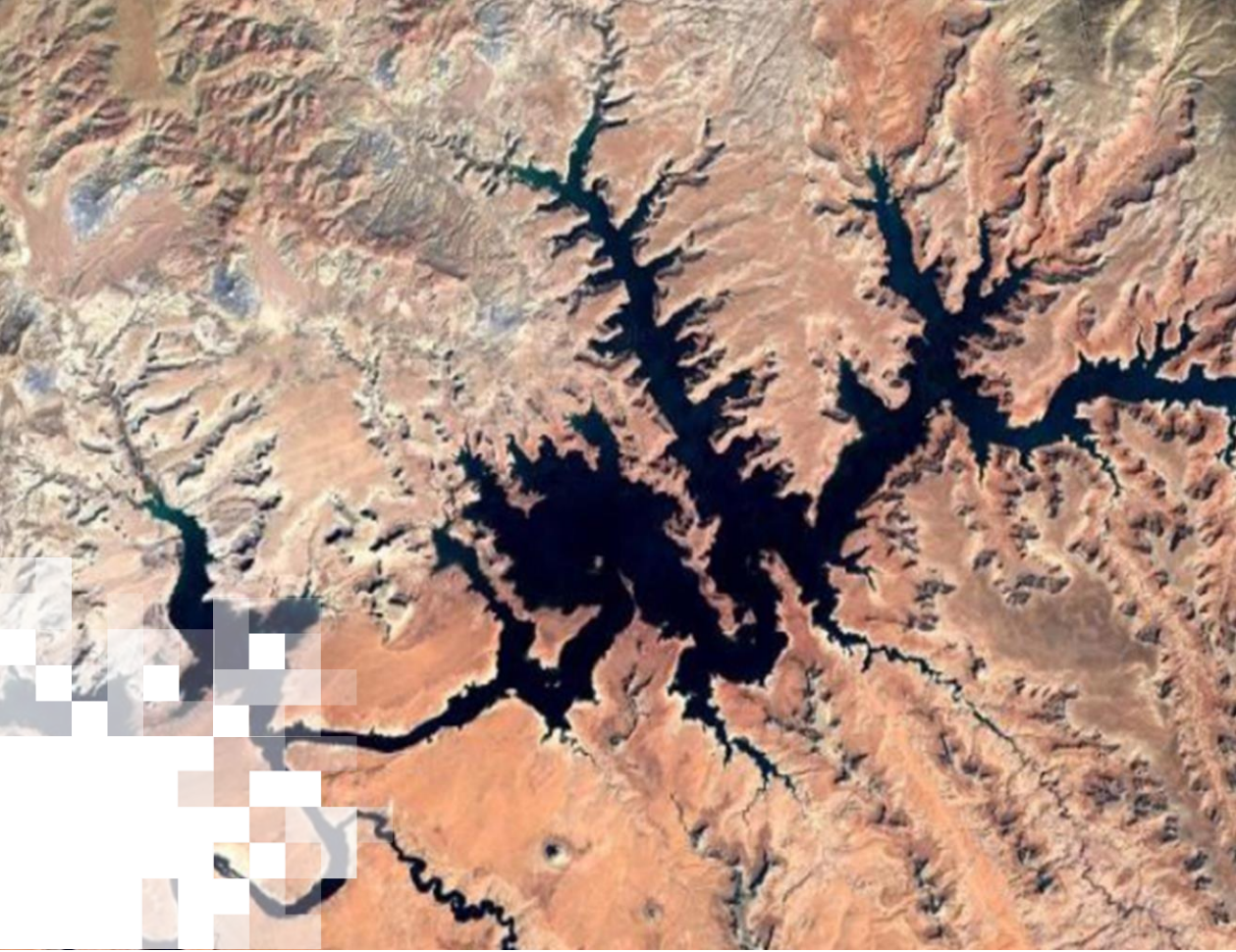
Técnicas para el Ajuste de Modelos

- **Búsqueda en cuadrícula:** el usuario define un conjunto de valores para cada hiperparámetro para formar una cuadrícula. Se prueban diferentes combinaciones de estos valores de hiperparámetros y la combinación que produce el mejor resultado se selecciona como el conjunto final de hiperparámetros óptimos.
- **Búsqueda aleatoria:** el algoritmo solo probará combinaciones aleatorias de valores de hiperparámetros en vez de todas las combinaciones posibles.
- **Búsqueda bayesiana:** realiza un seguimiento de los resultados de evaluaciones pasadas para formar la información utilizada para tomar decisiones futuras al seleccionar valores futuros de hiperparámetros.
- Al final, la combinación que produzca el mejor resultado a partir de esto es la que se selecciona como el conjunto óptimo de hiperparámetros.



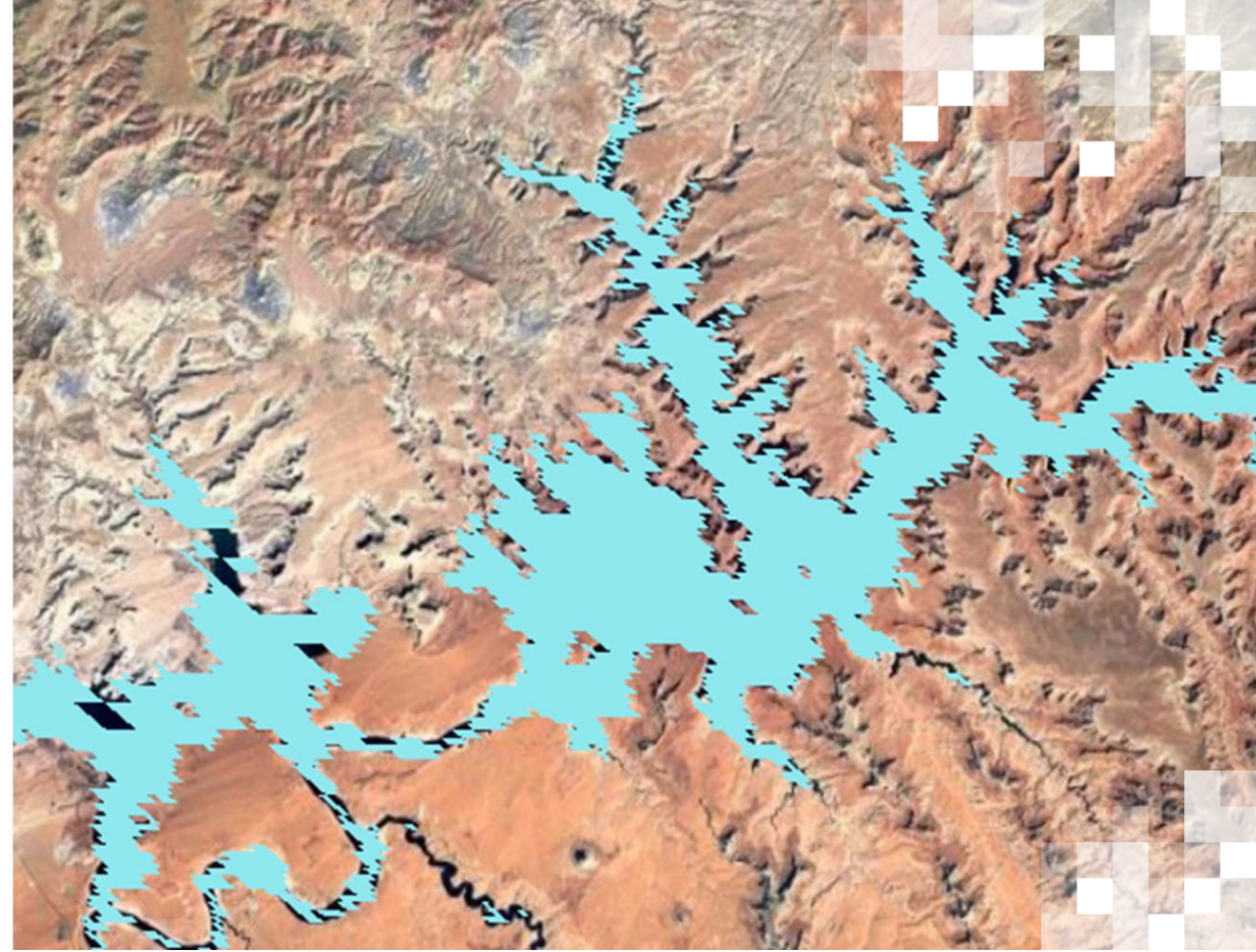
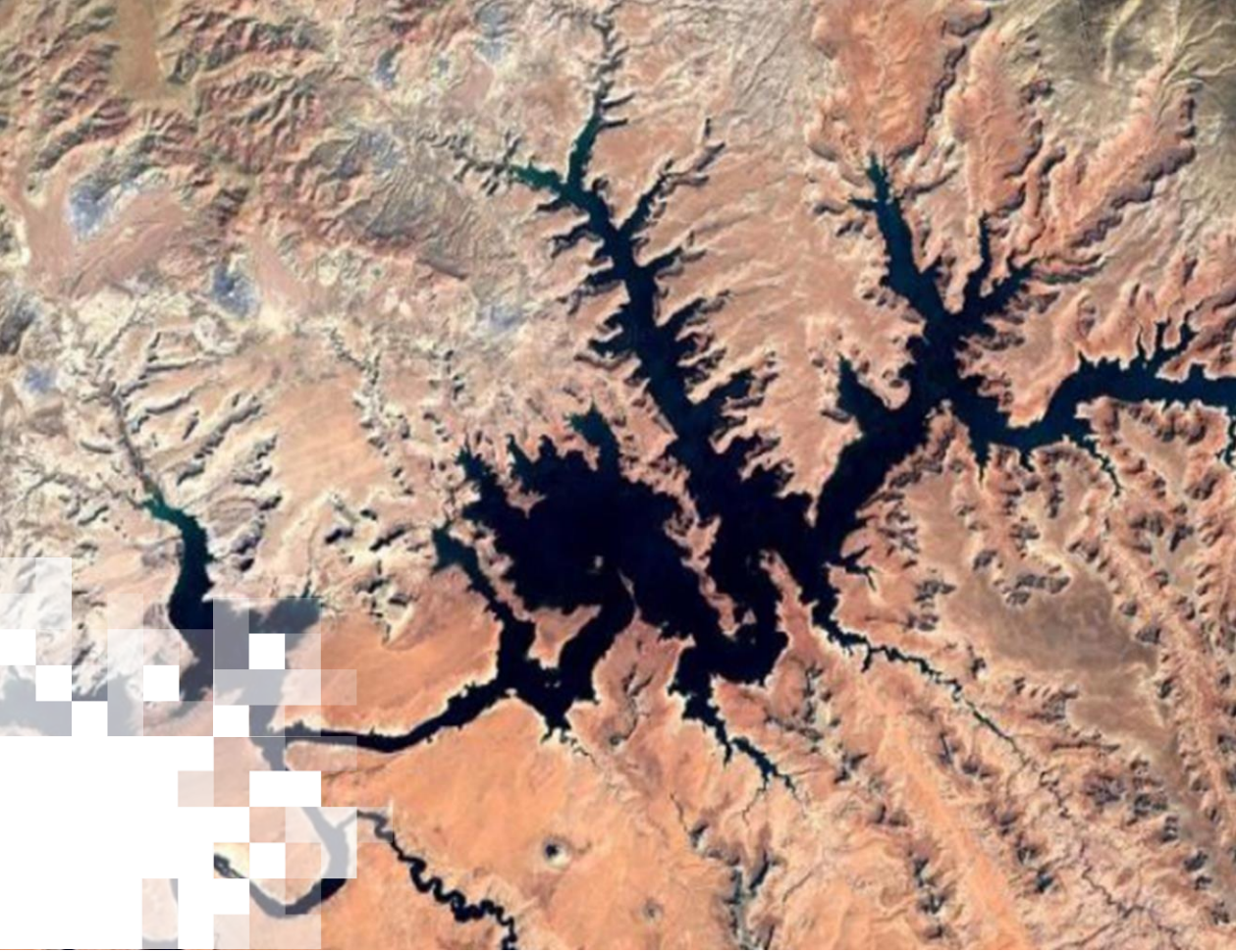
Optimización en cuadrícula, aleatoria y bayesiana ilustradas. Kim et al. (2021), <https://doi.org/10.1109/ACCESS.2021.3051619>





Ejercicio: Ajuste y Optimización del Modelo Random Forest

Formador: Jordan A. Caraballo-Vega

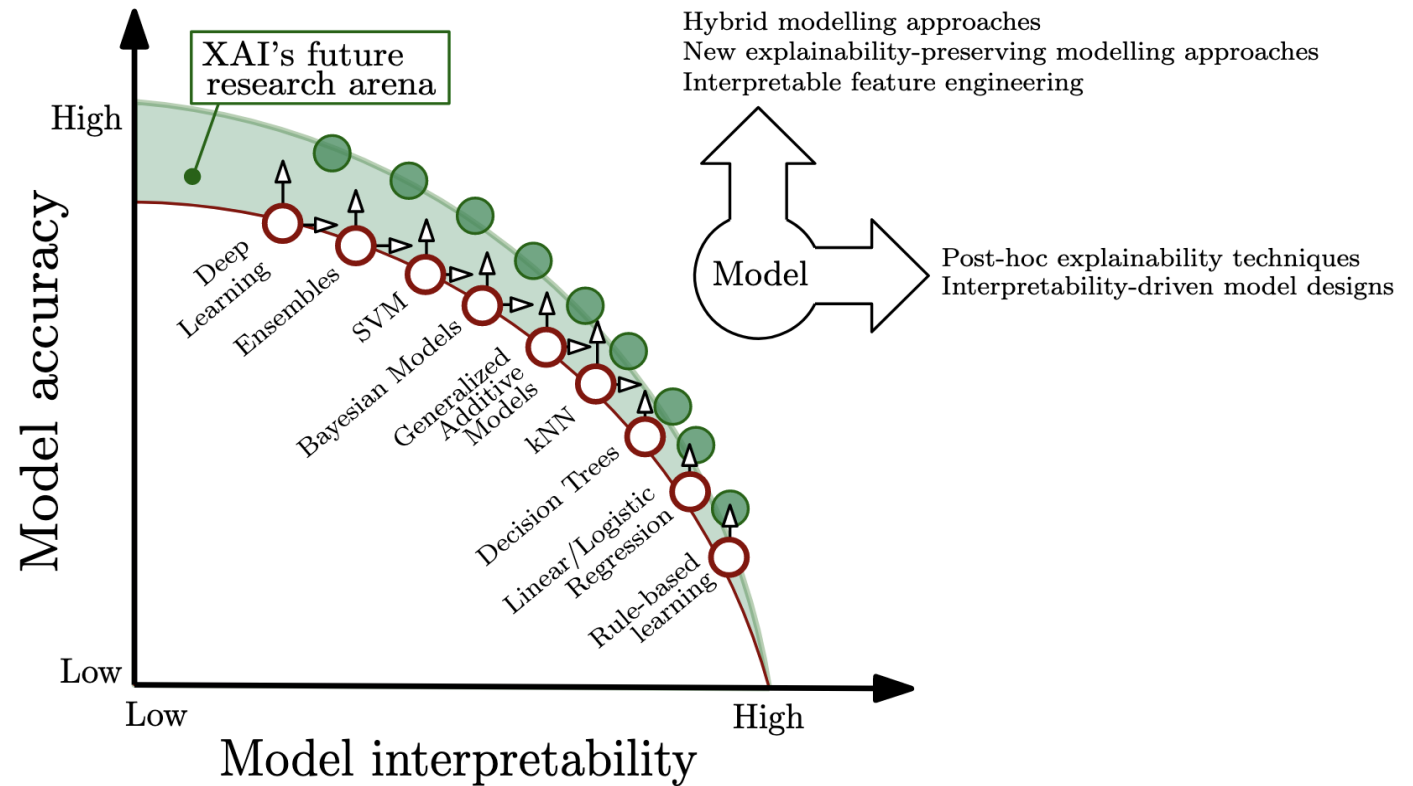


Explicabilidad e Interpretabilidad de los Modelos - XAI

Formador: Caleb S. Spradlin

Explicabilidad e Interpretabilidad de los Modelos – XAI

A medida que llegamos a depender de las inferencias proporcionadas por los modelos de aprendizaje automático, es importante que estos modelos sean precisos e interpretables.

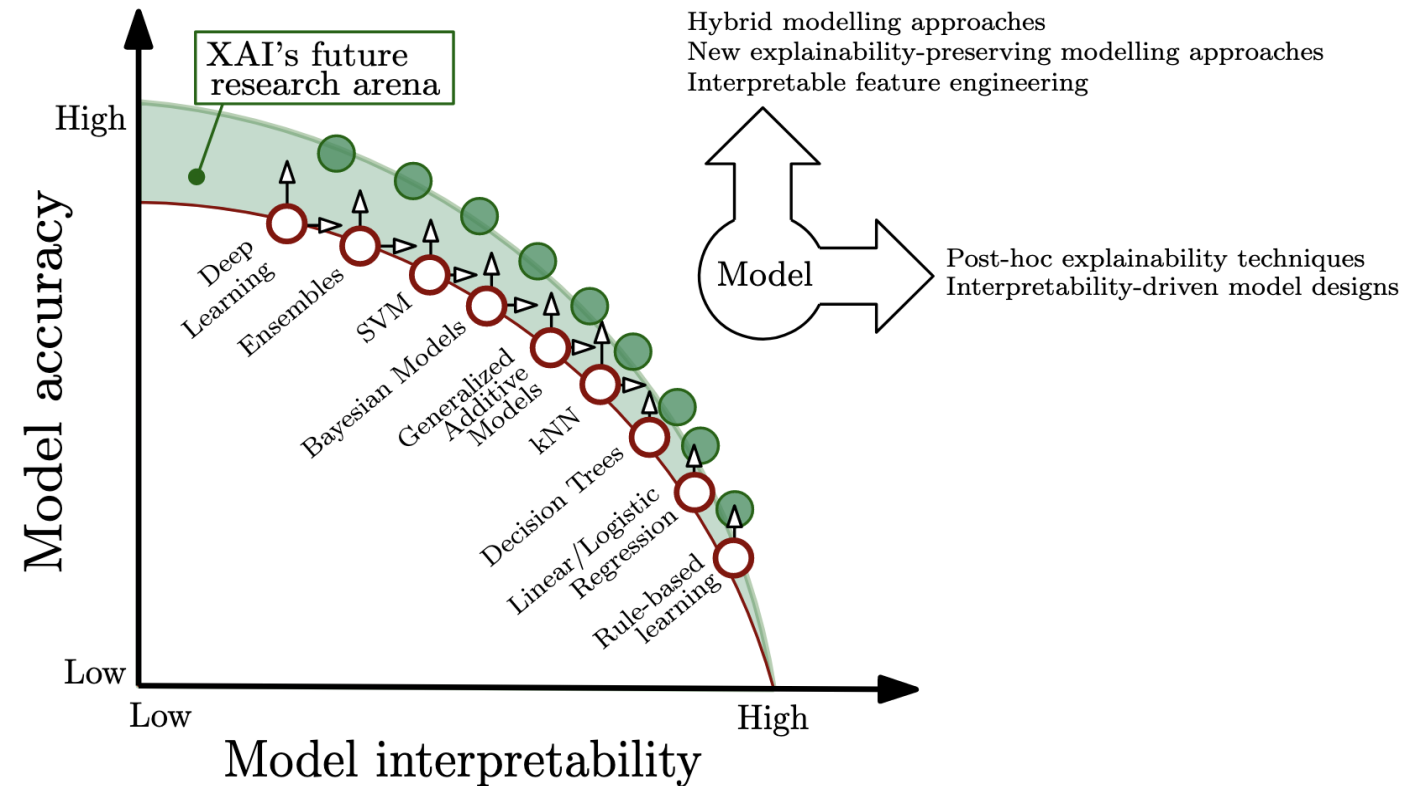


Arrieta et al. (2019), <https://doi.org/10.3389/fnsys.2021.766980>



¿Por Qué Necesitamos Modelos Confiables? – XAI

- La precisión puede que no sea suficiente.
- Los modelos de aprendizaje automático deben ser confiables.
- La confiabilidad está determinada por la interpretabilidad y la robustez.
- Interpretabilidad: podemos explicar por qué se predijo un resultado determinado.
- Robustez: la entrada puede ser ruidosa; aún así logramos predicciones precisas.

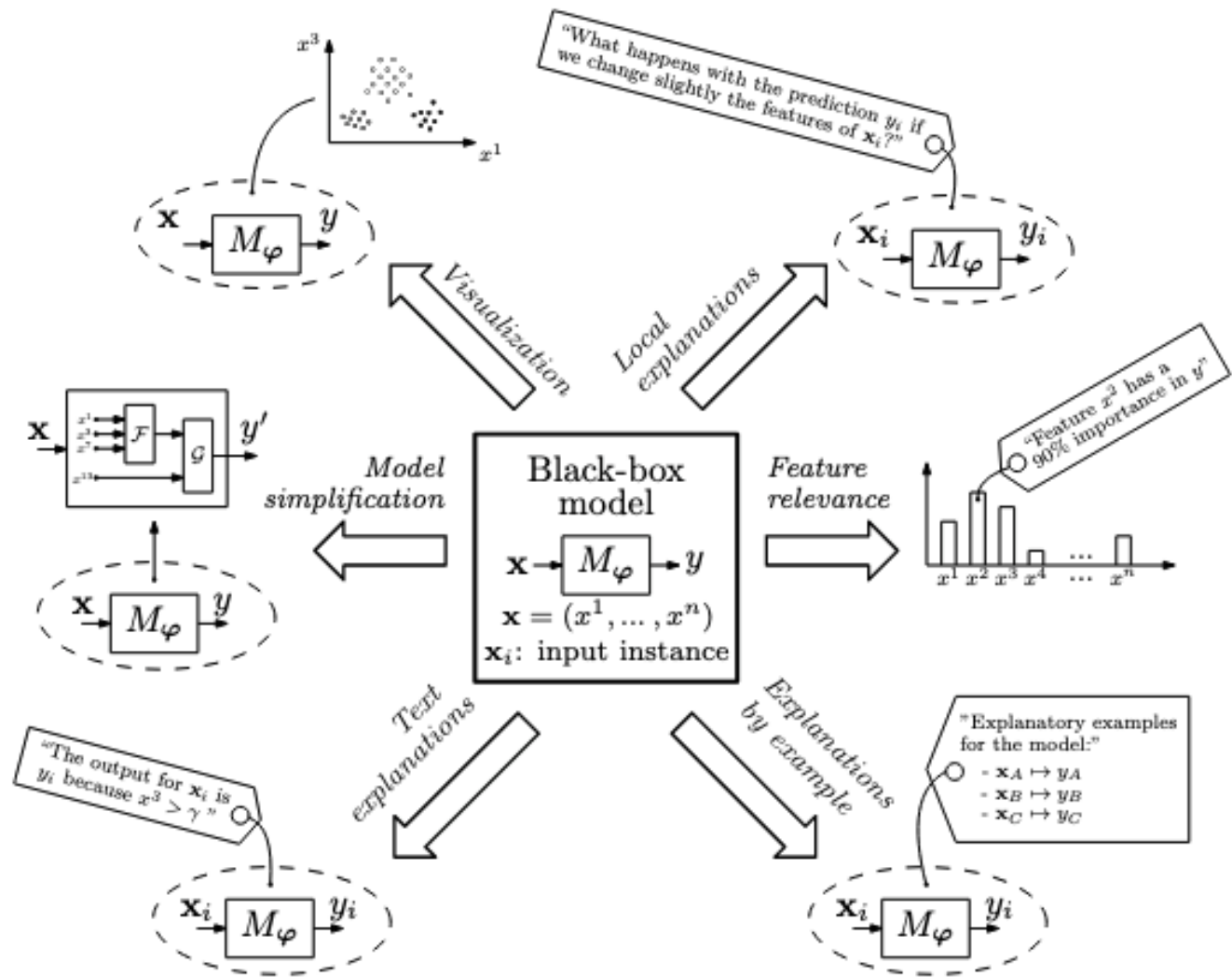


Arrieta et al. (2019), <https://doi.org/10.3389/fnsys.2021.766980>



Enfoques de Explicabilidad Post-hoc– XAI

- Uno de los métodos más comunes para lograr un modelo de ML interpretable es a través de métodos de explicación post-hoc (es decir, después de entrenar el modelo).
- Estos métodos utilizan la salida del modelo junto con las entradas para extraer información sobre las decisiones del modelo.

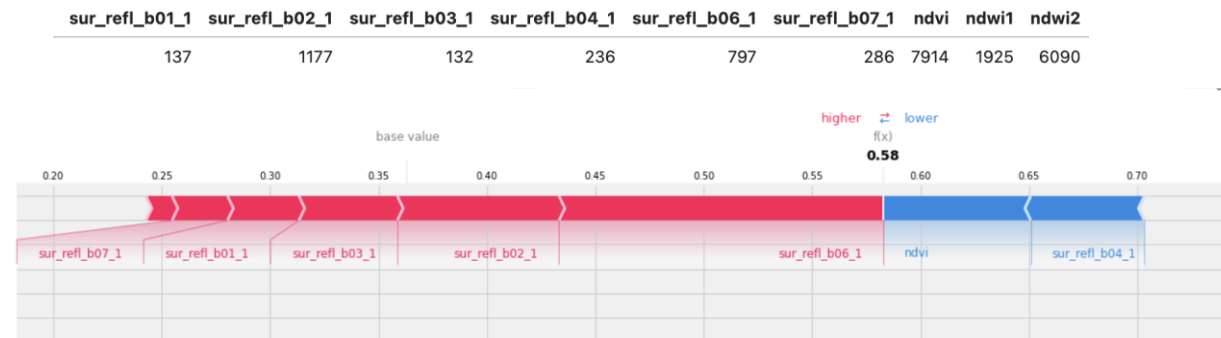
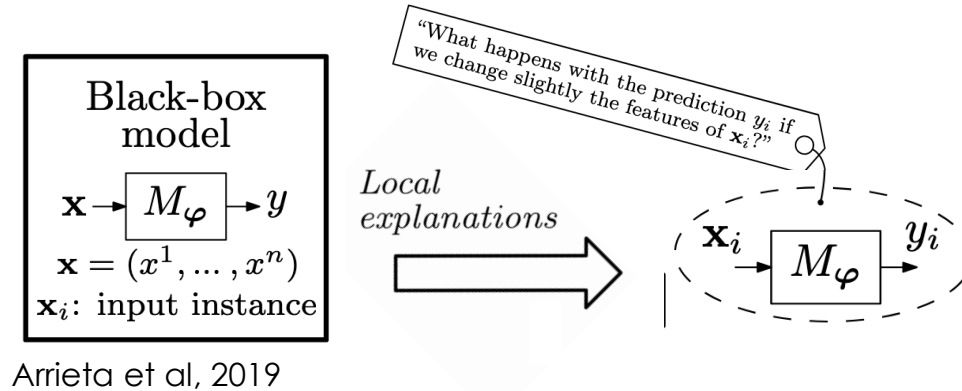


Explicabilidad e Interpretabilidad de los Modelos – XAI

- Una herramienta comúnmente utilizada es SHAP (SHapley Additive exPlanations).
- SHAP es un enfoque agnóstico en cuanto al modelo que puede calcular una puntuación aditiva de la importancia de las características para cada predicción.

Usando Valores SHAP para Explicaciones Locales

- Usando valores de Shapely para generar explicaciones de decisiones singulares para modelos de caja negra (black box)

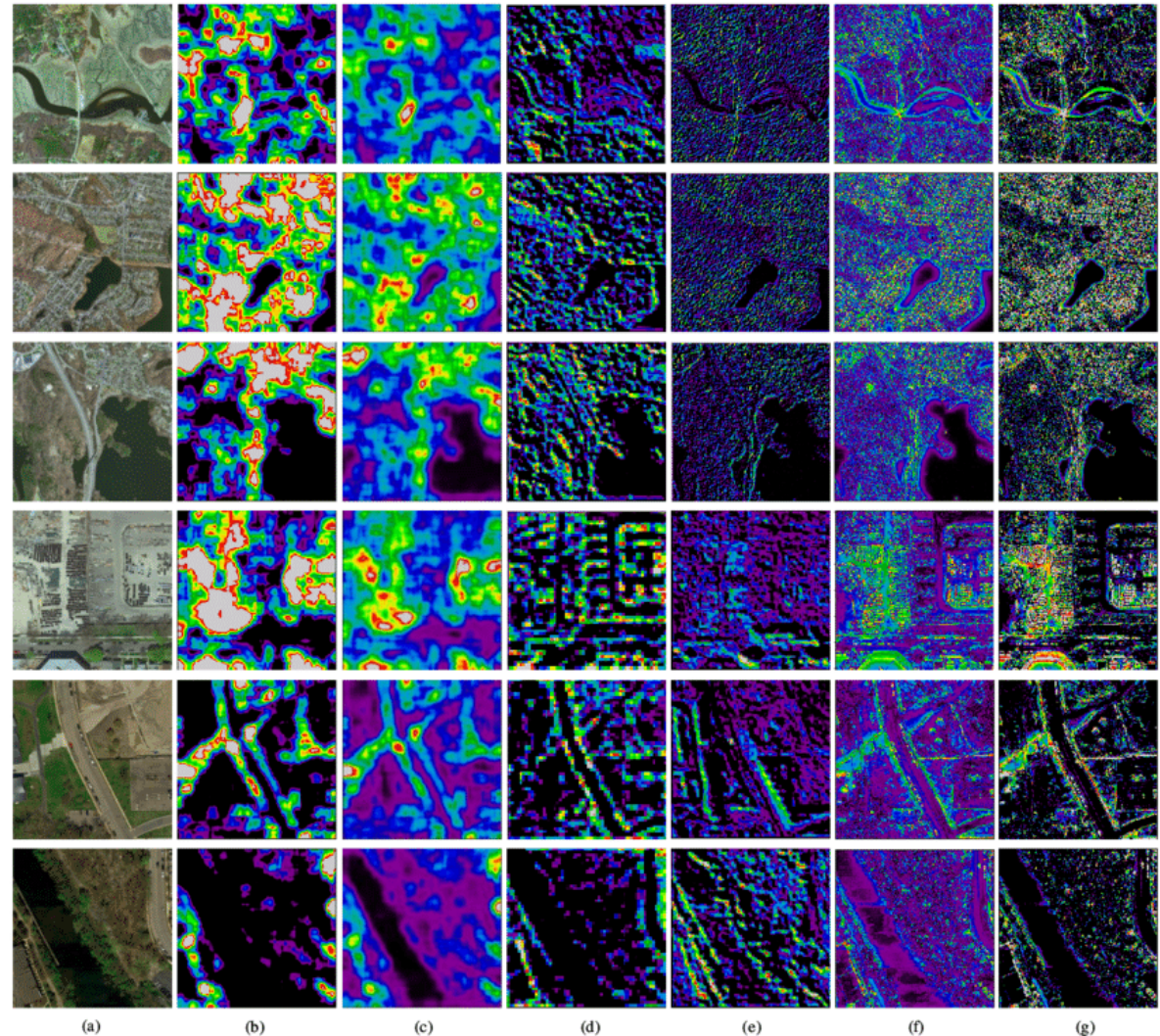


Fuente de la Imagen: <https://shap.readthedocs.io/en/latest/index.html>



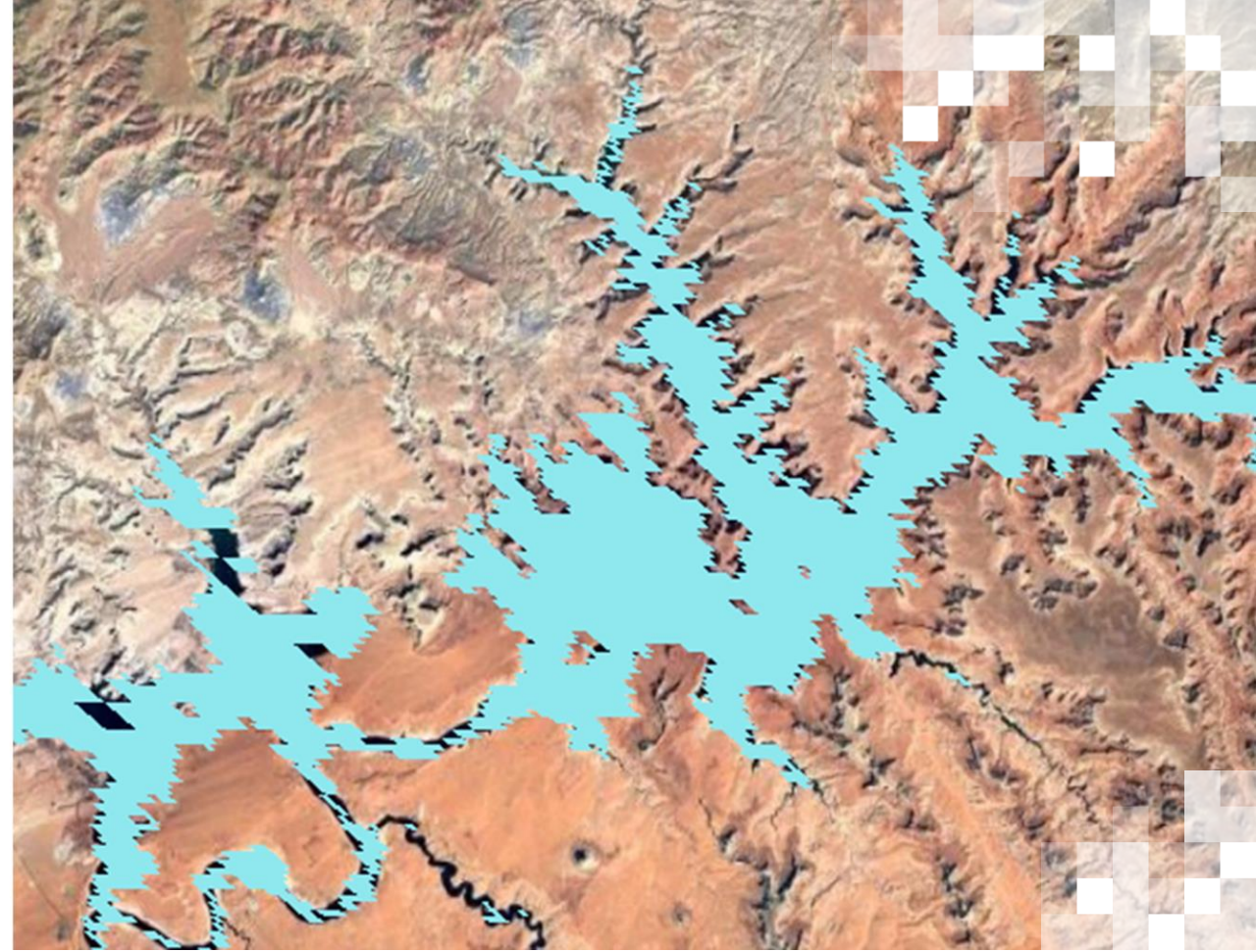
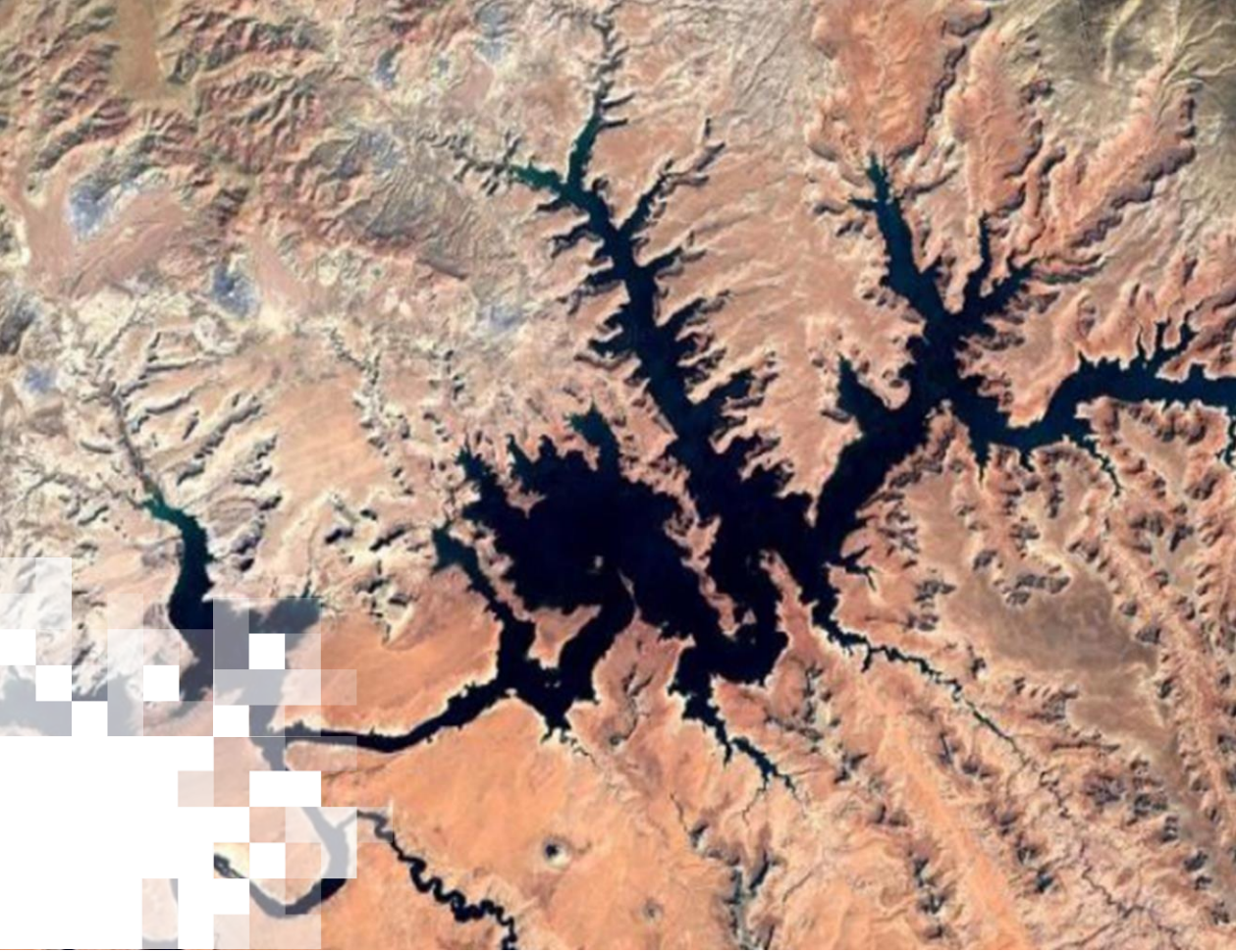
Atención y Explicabilidad – XAI

- Los transformadores visuales (Visual Transformers o ViT) pueden generar mapas de atención como salida.
- Los mapas de atención son la salida intermedia del modelo que resalta la región importante en la imagen para la clase objetivo.
- La visualización de mapas de atención puede llevar a una mejor comprensión de cómo el modelo procesa las entradas y cuáles son las características más importantes para la predicción.



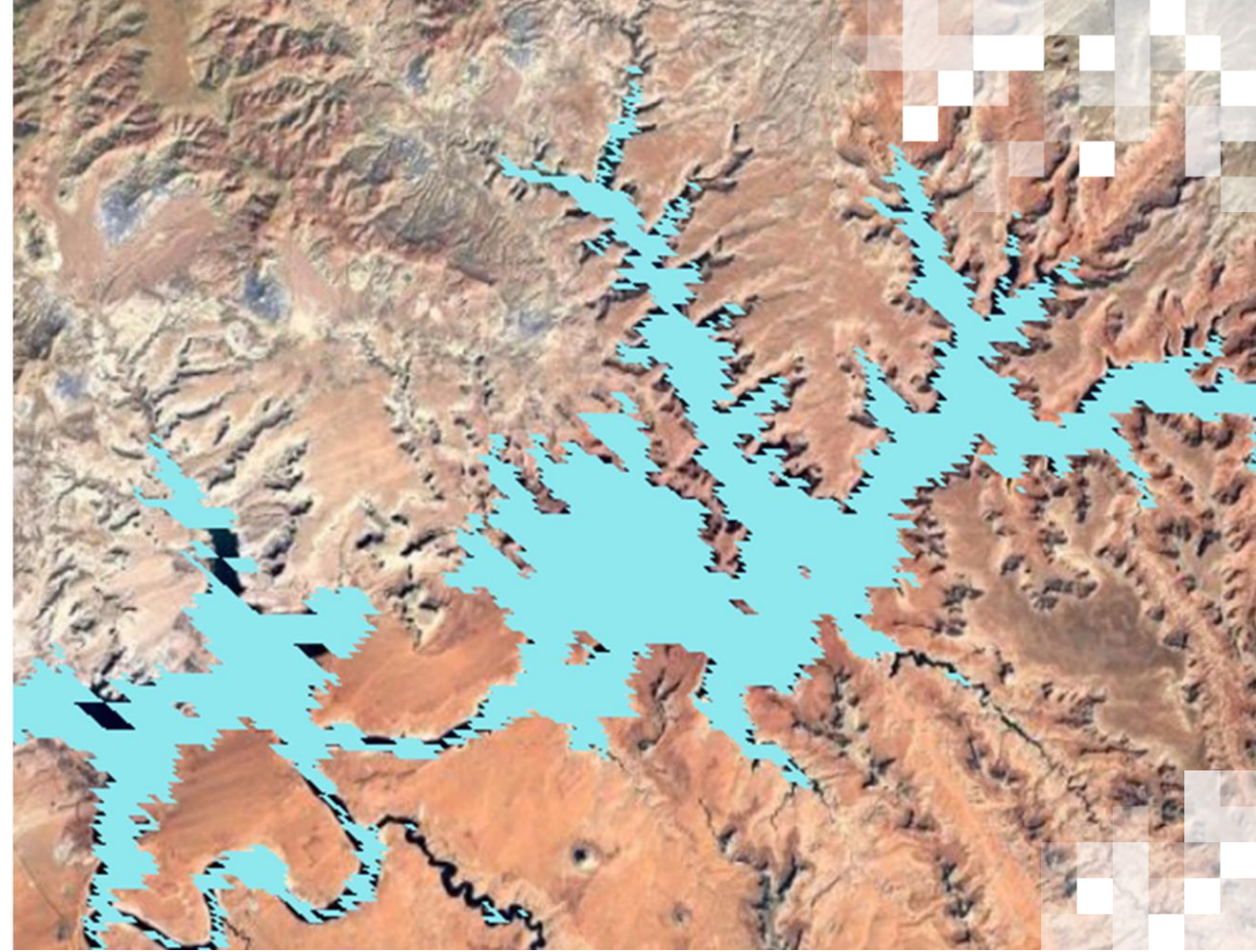
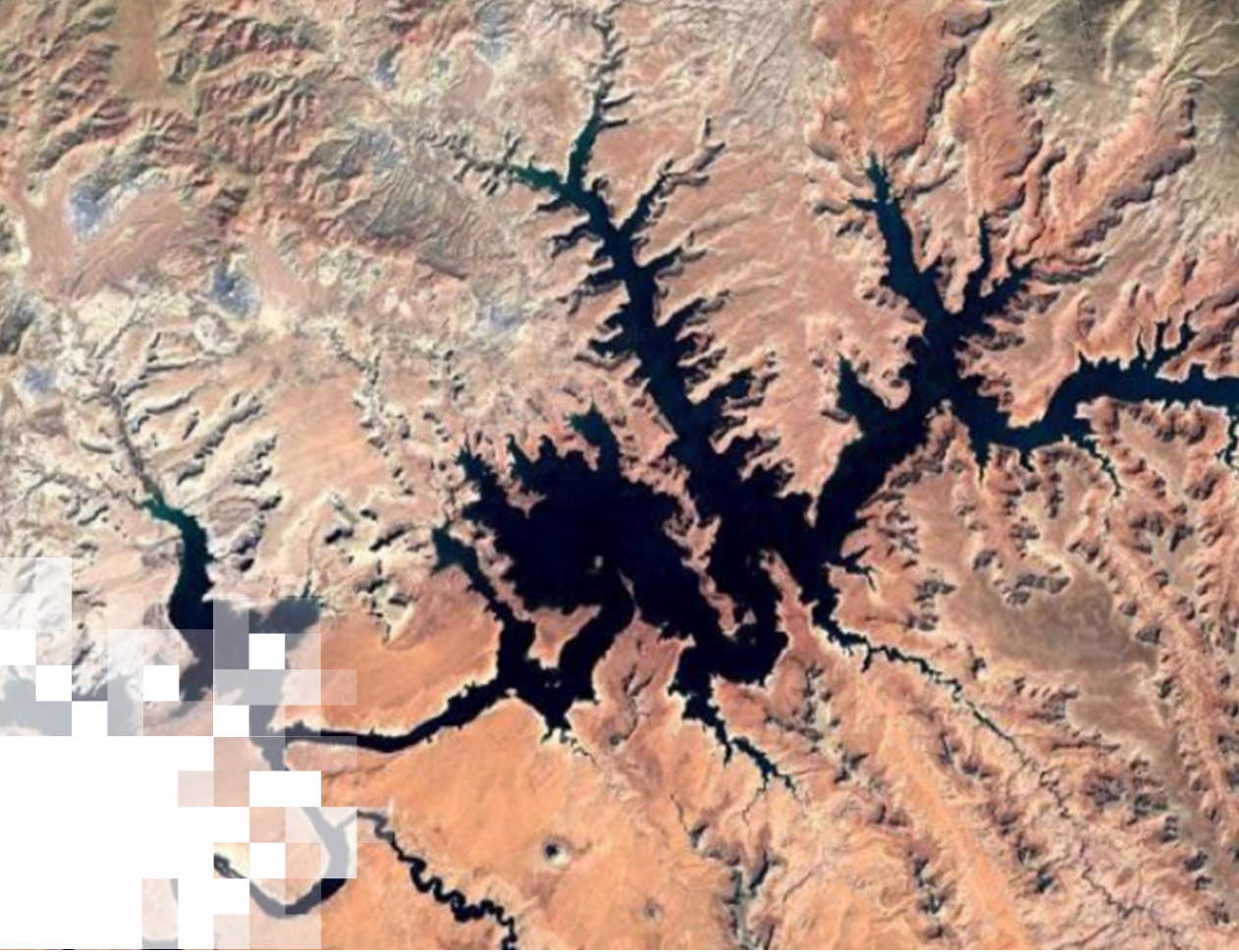
Imágenes de teledetección y visualización de mapas de atención en diferentes módulos. Shamsolmoali et al. (2020), <https://doi.org/10.1109/TGRS.2021.3112481>





Ejercicio: Explicabilidad e Interpretabilidad de los Modelos – XAI

Formador: Caleb S. Spradlin



Para Concluir

Formador: Jordan A. Caraballo-Vega

Para Concluir

- Hemos:
 - Proporcionado una base sobre los fundamentos del aprendizaje automático para las Ciencias de la Tierra usando un problema de clasificación binaria como ejemplo.
 - Presentado el concepto general del aprendizaje automático y los posibles escenarios de sus beneficios en diferentes dominios.
 - Proporcionado la base para producir un conjunto de datos de prueba, validación y entrenamiento eficaz a partir de fuentes de datos ráster y tabulares.
 - Proporcionado las herramientas para entrenar y realizar la inferencia de un modelo Random Forest, incluso su ajuste fino y análisis XAI.

Esta es solo una introducción al campo muy amplio que es el Aprendizaje Automático. Los fundamentos aprendidos en esta capacitación proporcionarán la base para comprender la literatura y saber cuándo un algoritmo específico podría ser el más aplicable.



Contactos

- Instructores:
 - Jordan A. Caraballo-Vega: jordan.a.caraballo-vega@nasa.gov
 - Jules Kouatchou: jules.kouatchou-1@nasa.gov
 - Caleb S. Spradlin: caleb.s.spradlin@nasa.gov
 - Jian Li: jian.li@nasa.gov
- Página Web de la Capacitación:
 - <https://appliedsciences.nasa.gov/join-mission/training/spanish/arset-fundamentos-del-aprendizaje-automatico-para-las-ciencias-de-la>
- Página Web de ARSET:
 - <https://appliedsciences.nasa.gov/arset>

Check out our sister programs:



¿Preguntas?

- Por favor escriban sus preguntas en la ventana de “Questions”. Las responderemos en el orden que las recibimos.
- Publicaremos las preguntas y respuestas a la página web de la capacitación después de la conclusión del webinar.





¡Gracias!



Referencias

- Crankshaw, D. y Gonzalez, J. (2018). Prediction-Serving Systems: What happens when we wish to actually deploy a machine learning model to production?. *Queue*, 16(1), 83-97.
- Elders, A., Carroll, M. L., Neigh, C. S., D'Agostino, A. L., Ksoll, C., Wooten, M. R. y Brown, M. E. (2022). Estimating crop type and yield of small holder fields in Burkina Faso using multi-day Sentinel-2. *Remote Sensing Applications: Society and Environment*, 27, 100820.
- Fleming, S. W., Watson, J. R., Ellenson, A., Cannon, A. J. y Vesselinov, V. C. (2021). Machine learning in Earth and environmental science requires education and research policy reforms. *Nature Geoscience*, 14(12), 878-880.
- Prša, A., Kochoska, A., Conroy, K. E., Eisner, N., Hey, D. R., IJspeert, L., ... y Winn, J. N. (2022). TESS Eclipsing Binary Stars. I. Short-cadence Observations of 4584 Eclipsing Binaries in Sectors 1–26. *The Astrophysical Journal Supplement Series*, 258(1), 16.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., y Carvalhais, N. the National Energy Research Supercomputing Center in Lawrence Berkeley National Laboratory, Berkeley, CA, USA: Deep learning and process understanding for data-driven Earth system science. *Nature*, 566, 195-204.
- Yu, S., y Ma, J. (2021). Deep learning for geophysics: Current and future trends. *Reviews of Geophysics*, 59(3), e2021RG000742.



Contribuidores

- Jordan A. Caraballo-Vega
- Mark L. Carroll
- Jules R. Kouatchou
- Jian Li
- Caleb S. Spradlin
- Brock Blevins
- Melanie Follette-Cook
- Erika Podest
- Brian Powell
- Akiko Elders

