

# RETOS COAFINA 2024

## Búsqueda de patrones periódicos en la respuesta de un detector de astropartículas

Retador.a	Dennis Cazar Ramírez
Correo	dcazar@usfq.edu.ec
Institución	Colaboración LAGO
<b>Resumen del reto</b>	
<p>La Colaboración LAGO está implementando una red de detectores de astropartículas en toda Latinoamérica, estos detectores están adquiriendo los impulsos eléctricos que genera un sensor (tubo fotomultiplicador o PMT) las 24 horas del día. A régimen este flujo de datos puede alcanzar algunos Giga por hora por lo que el análisis de estos datos en largos períodos de tiempo (días) requiere el uso de gran poder de computo, es decir acceso a HPCs, recursos que pueden ser costosos y requieren entrenamiento previo.</p> <p>Cualquier sistema de adquisición de datos no está exento de ruido e interferencias debido a los circuitos de alimentación y o convertidores DC-DC que pueden manifestarse como patrones con periodicidad (frecuencia) fija en los datos y los llamados "impulsos oscuros" que son señales generadas dentro del PMT que no corresponden con la detección de astropartículas.</p> <p>La propuesta de este reto es analizar los archivos de datos de los detectores, usando técnicas de ML y DL, para poder discriminar los impulsos por energía y por frecuencia y así poder identificar patrones de ruido.</p> <p>Hecho esto se puede aplicar el método desarrollado a los archivos individuales durante la operación del detector (por lo que no se necesitaría alto poder de computo) y generar un archivo "limpio" que contendría sólo impulsos provenientes de astropartículas menos pesado que puede ser luego analizado off-line</p> <p>Este análisis permitiría también realizar una calibración mas sofisticada del detector y especializada para el estudio de fenómenos de interés específico (Clima espacial, GRBs, entre otros)</p>	
<b>Objetivos y motivación del reto</b>	
<p>Objetivo general : Análisis de los datos de un detector de astropartículas para la discriminación de impulsos por energía y por frecuencia</p> <p>Objetivo 1: aplicación de técnicas de ML o DL para la identificación de patrones periódicos en los datos.</p> <p>Objetivo 2: proponer un algoritmo de preprocesamiento de datos que separen la señal útil del detector de las diferentes fuentes de ruido</p>	

Motivación: facilitar el análisis off-line de los datos de los detectores de la colaboración LAGO
<b>Datasets del reto, ubicación y descripción</b>
Los datos se encuentran en un servidor de la Colaboración Se almacenan en archivos de texto con la evolución del impulso, un contador de eventos y un sello temporal absoluto Los datos son de acceso abierto Cada archivo que contiene una hora de datos puede pesar 1Gb aproximadamente
<b>Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]</b>
Información detallada se encuentra en la página de LAGO <a href="http://lagoproject.net">lagoproject.net</a>
<b>Criterios para evaluar la solución del reto [opcional]</b>
Histogramas por alto de impulso PHD (Pulse Height Distribution) Clasificación de impulsos por frecuencia de suceso Algoritmo de limpieza de datos para aplicar en cada archivo y que no requiera de alto poder de computo, de preferencia que pueda ser ejecutado por un uP de modestas prestaciones y memoria limitada (una Raspberry por ejemplo)
<b>Algún comentario adicional</b>
LAGO Ecuador proporcionará toda la información necesaria para trabajar sobre este reto

## Democratizemos los datos meteorológicos de la historia de Latinoamérica

Retador.a	Elisa Sevilla, Pablo Mendieta y Adrián Vásconez
Correo	<a href="mailto:esevillap@usfq.edu.ec">esevillap@usfq.edu.ec</a> , <a href="mailto:pmendieta@estud.usfq.edu.ec">pmendieta@estud.usfq.edu.ec</a> , <a href="mailto:avasconez@estud.usfq.edu.ec">avasconez@estud.usfq.edu.ec</a>
Institución	Universidad San Francisco de Quito
<b>Resumen del reto</b>	
<p>Varios periódicos latinoamericanos del siglo XIX se encuentran digitalizados en los repositorios digitales de acceso abierto pertenecientes a bibliotecas públicas como la Biblioteca de la Casa de las Culturas Ecuatorianas y la Biblioteca Nacional de Colombia. Uno de estos periódicos es “El Nacional” en Ecuador que, en la década de 1871-1880, publicó los reportes meteorológicos del Observatorio Astronómico de Quito; se registran datos de posición del barómetro, temperatura, presión del vapor y humedad relativa tres veces al día, y cantidad de lluvia diaria. Otro ejemplo es el periódico “Gaceta Oficial” de Colombia, donde Tomás Cipriano Mosquera publicó mediciones de presión, temperatura, lluvia y humedad. Todas estas tablas de datos están en un formato fotográfico</p> <p>Se propone el reto de crear una herramienta que permita identificar y ubicar las tablas de datos de entre las fotografías de los repositorios. Además, la herramienta deberá poder extraer los valores numéricos utilizando diferentes técnicas de procesamiento de imágenes, reconocimiento óptico de caracteres e inteligencia artificial.</p>	
<b>Objetivos y motivación del reto</b>	
<p>“El objetivo de este proyecto es crear herramientas digitales que permitan la extracción de información meteorológica —y otro tipo de información tabulada— de documentos históricos. Esta información es crucial para un análisis efectivo de tendencias climáticas y no ha sido accesible debido a su formato fotográfico que impide su análisis efectivo. La poca información que se tiene fue digitalizada a mano, lo cual es un proceso lento y difícil.</p> <p>Estos datos contribuirán al análisis y comprensión de la historia climática de Latinoamérica, y ayudarán a tomar decisiones informadas en temas de política pública, resiliencia a desastres naturales y desarrollo. Además, la herramienta desarrollada podrá ser aplicada a otros repositorios latinoamericanos, permitiendo democratizar el acceso a la información con un gran impacto en la investigación científica. En una región propensa a fenómenos climáticos destructivos como el fenómeno del Niño, es crucial un entendimiento de los patrones climáticos específicos que influyen en estos procesos naturales. Se ha trabajado con más argumentos para justificar la reconstrucción de patrones climáticos en Mora, 2018; Mora, 2019; y Pabón, 2006.</p> <p>Esta propuesta extendería el trabajo de digitalización realizado por Domínguez-Castro</p>	

et al. en 2017 que resultó en la recopilación de datos meteorológicos de Argentina, Bahamas, Belice, Brasil, Guayana Británica, Chile, Colombia, Costa Rica, Cuba, Ecuador, Francia (Martinica y Guadalupe), Guatemala, Jamaica, México, Nicaragua, Panamá, Perú, Puerto Rico, El Salvador y Surinam durante los siglos XVIII y XIX. Además, contribuiría dentro del marco de preservación y liberación de datos DARE de la Organización Meteorológica Mundial y la GFCS (Global Framework for Climate Services).

Pabón Caicedo, J. D. (2006). El clima de Colombia durante los siglos XVI-XIX a partir de material histórico. Parte I: inventario de fuentes de información. Cuadernos de Geografía: Revista Colombiana de Geografía, (15), 75-92.

<https://revistas.unal.edu.co/index.php/rcg/article/view/1288>

Mora, Katherinne. (2018). Tras la pista de “terribles veranos” y “copiosas lluvias”. Elementos para una historia climática del territorio colombiano. Historia Crítica, (74). <http://journals.openedition.org/histcrit/1571>

Mora, Katherinne. (2018). Pensar el pasado para adaptarse al cambio climático. El aporte necesario de la historia ambiental latinoamericana. Letras Verdes, Revista Latinoamericana de Estudios Socioambientales, (24), 8-26. <https://doi.org/10.17141/letrasverdes.24.2018.3317>

#### **Datasets del reto, ubicación y descripción**

"Ubicación: Biblioteca Digital de la Casa de las Culturas Ecuatorianas

Formato: PDF con imágenes escaneadas

Licencia:

Tamaño aproximado: ~5 MB por PDF

Descripción de metadatos: Título, Otros títulos, Palabras clave, Fecha de publicación, Editorial, Resumen, URI, Aparece en colecciones

Esta base de datos contiene archivos PDF para cada número aproximadamente (trisemanal). Conocemos que el periódico EL NACIONAL contiene datos en 1871 y 1872 (hint). Ejemplo para enero de 1871:

<http://repositorio.casadelacultura.gob.ec/handle/34000/9417>

Repositorio: <http://repositorio.casadelacultura.gob.ec/handle/34000/1534>

Ubicación: Biblioteca Nacional de Colombia

Formato: PDF con imágenes escaneadas

Licencia: Obras en dominio público digitalizado por la Biblioteca Nacional de Colombia

Tamaño aproximado: ~5 MB por PDF

Descripción de metadatos: Título, Autor, Editor, Idioma, Materias, Derechos.

Esta base de datos contiene archivos PDF para cada mes aproximadamente.

Conocemos que existen datos meteorológicos al menos en los meses de febrero y abril (hint).

Repositorio: <https://bibliotecanacional.gov.co/content/conservacion?idFichero=89268>

#### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

<b>Criterios para evaluar la solución del reto [opcional]</b>
Se podría tomar una muestra aleatoria de los datos extraídos y verificar el porcentaje de error en la transcripción. Además, si se implementaron herramientas de mejoramiento de imágenes, se podría realizar una comparación cualitativa de la legibilidad de los documentos.
<b>Algún comentario adicional</b>
"Por favor, me gustaría que se copie en las comunicaciones a: Pablo Mendieta, pmendieta@estud.usfq.edu.ec Adrian Vásconez, avasconez@estud.usfq.edu.ec Gracias!"

## Creación de una Herramienta de Visualización Interactiva para Análisis de Desempeños en Pruebas Saber 11 y Saber Pro

Retador.a	David Sierra Porta
Correo	sierraporta@gmail.com
Institución	Universidad Tecnológica de Bolívar - UTB
<b>Resumen del reto</b>	
<p>"En Colombia, las pruebas Saber 11 y Saber Pro son evaluaciones clave que miden el rendimiento académico de los estudiantes en niveles básicos y universitarios. Estas pruebas proporcionan datos valiosos sobre el desempeño educativo, pero actualmente existe una carencia significativa de herramientas accesibles y autocontenidas que permitan a directivos de colegios, autoridades educativas, gerentes de educación y tomadores de decisiones analizar estos datos de manera efectiva y comprensible. La falta de estas herramientas limita la capacidad para identificar y abordar las brechas en la formación y el desempeño de los estudiantes. Los directivos y las autoridades educativas a menudo se enfrentan a dificultades para interpretar los resultados de estas pruebas y utilizar esa información para diseñar estrategias educativas efectivas. Esta carencia de análisis accesible y detallado impide la implementación de mejoras informadas y basadas en datos, afectando negativamente la calidad educativa y la equidad en la educación.</p> <p>El reto propuesto busca desarrollar una herramienta en línea de visualización interactiva que utilice datos abiertos del ICFES sobre las pruebas Saber 11 y Saber Pro. La herramienta deberá incorporar técnicas avanzadas de minería de datos y visualización científica para proporcionar un análisis detallado y accesible de los datos. El objetivo es crear un dashboard interactivo que permita a los usuarios explorar los datos de manera intuitiva, identificando patrones, tendencias y brechas educativas de forma clara y comprensible.</p> <p>Esta herramienta será capaz de presentar visualizaciones dinámicas, incluir filtros personalizables y ofrecer análisis detallados que faciliten la toma de decisiones basada en datos. Con esta solución, se espera empoderar a los directivos de colegios, autoridades educativas y tomadores de decisiones, proporcionándoles una herramienta eficaz para evaluar el rendimiento académico, identificar áreas de mejora y desarrollar estrategias educativas informadas que promuevan un mayor rendimiento académico y equidad en la educación en Colombia.</p> <p>Utilizando datos abiertos del ICFES y técnicas avanzadas de minería de datos y visualización científica, la herramienta facilitará la identificación de estas brechas y apoyará la toma de decisiones informadas. Se diseñará para ser accesible, intuitiva y autocontenida, promoviendo un análisis profundo y comprensible de los datos educativos. Esta herramienta proporcionará visualizaciones dinámicas y filtros personalizables que permitirán a los usuarios explorar los datos de manera intuitiva, identificando patrones, tendencias y brechas educativas de forma clara y comprensible, contribuyendo así a la mejora de la calidad educativa en Colombia."</p>	

## Objetivos y motivación del reto

### "Objetivos del Reto:

Desarrollar una herramienta en línea de visualización interactiva para el análisis de datos de las pruebas Saber 11 y Saber Pro.

Facilitar la identificación de brechas educativas y apoyar la toma de decisiones informadas por parte de directivos y autoridades educativas.

Proporcionar una herramienta accesible y autocontenida que permita un análisis profundo y comprensible de los datos educativos.

### Empoderamiento de la Comunidad Educativa:

Identificar Áreas de Mejora: Al analizar los datos de desempeño, los usuarios pueden identificar rápidamente las áreas donde los estudiantes tienen dificultades y diseñar intervenciones específicas para abordar estas brechas.

Desarrollar Estrategias Basadas en Datos: Las decisiones informadas y basadas en datos son cruciales para implementar estrategias efectivas que mejoren el rendimiento académico y la equidad educativa.

Monitorear el Progreso: Una herramienta interactiva permitirá a las instituciones monitorear el progreso a lo largo del tiempo, evaluando la efectividad de las intervenciones implementadas y ajustándolas según sea necesario.

Fomentar la Transparencia y la Rendición de Cuentas: El acceso a datos claros y comprensibles promueve la transparencia en el sistema educativo y permite a los diferentes actores rendir cuentas sobre el desempeño y las acciones tomadas para mejorar la educación.

### Beneficios para Diversos Actores:

Directivos de Colegios: Podrán utilizar la herramienta para evaluar el rendimiento de sus estudiantes, identificar áreas de mejora, y diseñar planes de acción concretos.

Autoridades Educativas: Podrán analizar los datos a nivel regional o nacional para diseñar políticas educativas más efectivas y equitativas.

Docentes: Los maestros pueden utilizar la información para adaptar sus métodos de enseñanza y brindar apoyo adicional a los estudiantes que lo necesiten.

Estudiantes y Familias: Una mayor transparencia y acceso a la información puede ayudar a los estudiantes y sus familias a comprender mejor su desempeño y buscar apoyo donde sea necesario.

Investigadores y Académicos: Podrán utilizar la herramienta para realizar estudios y análisis que contribuyan al conocimiento y la mejora del sistema educativo.

### Contribución a la Sociedad:

Mejorar la Calidad Educativa: Al proporcionar una forma accesible y efectiva de analizar y comprender los datos educativos, se pueden implementar estrategias más acertadas y eficaces para mejorar la calidad de la educación.

Promover la Equidad: Al identificar y abordar las brechas en el rendimiento académico, se promueve una educación más equitativa, donde todos los estudiantes tienen la oportunidad de alcanzar su máximo potencial.

Fomentar la Innovación: El uso de técnicas avanzadas de minería de datos y visualización científica fomentará la innovación en el análisis de datos educativos y el

desarrollo de soluciones tecnológicas en el sector educativo."

### **Datasets del reto, ubicación y descripción**

"Para el desarrollo del reto, se utilizarán conjuntos de datos abiertos proporcionados por el Instituto Colombiano para la Evaluación de la Educación (ICFES) y disponibles en el portal de datos abiertos del Gobierno de Colombia. Estos conjuntos de datos incluyen información detallada sobre los resultados de las pruebas Saber 11 y Saber Pro, así como variables sociodemográficas y contextuales que permiten un análisis integral del desempeño académico de los estudiantes colombianos.

#### Ubicación:

Portal de Datos Abiertos del ICFES: ICFES - Datos Abiertos:  
<https://www.icfes.gov.co/>

Portal de Datos Abiertos del Gobierno de Colombia: Datos Abiertos Colombia - Resultados Únicos Saber 11:  
[https://www.datos.gov.co/Educaci-n/Resultados-nicos-Saber-11/kgxf-xxbe/about\\_data](https://www.datos.gov.co/Educaci-n/Resultados-nicos-Saber-11/kgxf-xxbe/about_data)

#### Formato:

Los datos están disponibles en formato CSV, lo que facilita su manejo y análisis mediante diversas herramientas de minería de datos y visualización.

#### Licencia:

Los conjuntos de datos están disponibles bajo una licencia de acceso abierto que permite su uso, reutilización y redistribución. Se deben seguir los términos especificados en la licencia para garantizar el cumplimiento de las normativas de acceso abierto.

#### Resultados Únicos Saber 11 y Saber Pro:

Descripción: Este conjunto de datos contiene los resultados individuales de los estudiantes que presentaron la prueba Saber 11, incluyendo puntajes por áreas evaluadas y variables sociodemográficas.

Variables Incluidas: Identificador del estudiante, puntajes en áreas específicas (lectura crítica, matemáticas, ciencias naturales, sociales y ciudadanas, inglés), información sociodemográfica (género, edad, estrato socioeconómico, tipo de institución educativa, ubicación geográfica).

Rango Temporal: El conjunto de datos abarca varios años de evaluación, proporcionando una perspectiva longitudinal del desempeño académico.

Metadatos: Incluyen descripciones detalladas de cada variable, su significado, y la forma en que se recolectaron los datos, asegurando que los usuarios comprendan plenamente la estructura y el contenido del conjunto de datos.

Para garantizar que los conjuntos de datos utilizados en el desarrollo del reto sigan los principios internacionales de Acceso Abierto, es necesario que todos los participantes se adhieran a las licencias y condiciones de uso especificadas. Además, se deben tomar las medidas necesarias para proteger la privacidad y la integridad de los datos, especialmente aquellos que contienen información sensible sobre los estudiantes."

<b>Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]</b>
<b>Criterios para evaluar la solución del reto [opcional]</b>
"Pertinencia del Reto: Relación con temas educativos y científicos. Uso de datos abiertos para resolver la problemática planteada. Impacto y Relevancia: Potencial de generar un impacto positivo en la educación en Colombia. Relevancia para la comunidad educativa y tomadores de decisiones. Originalidad y Creatividad: Enfoque innovador y creativo en la solución del problema. Diseño intuitivo y accesible de la herramienta. Viabilidad Técnica: Factibilidad técnica de desarrollar la solución dentro del marco del hackathon como un prototipo que pueda ser fácilmente escalable y generalizado a más años y más datos. Uso de tecnologías web, bibliotecas de visualización de datos y herramientas de minería de datos. Claridad y Detalles: Descripción clara y precisa de la problemática y objetivos. Provisión de detalles suficientes para que los participantes puedan desarrollar soluciones concretas."
<b>Algún comentario adicional</b>

## Desarrollo de Modelos de Inteligencia Artificial para Caracterización y Predicción del Ciclo Solar 25 usando Imágenes Espectrales del SOHO

Retador.a	David Sierra Porta
Correo	sierraporta@gmail.com
Institución	Universidad Tecnológica de Bolívar - UTB
<b>Resumen del reto</b>	
<p>El reto consiste en desarrollar modelos de machine learning e inteligencia artificial que hagan uso de imágenes espectrales del Observatorio Solar y Heliosférico (SOHO) para caracterizar y determinar la completitud del ciclo solar 25, actualmente en curso. Las imágenes a utilizar incluyen EIT171, EIT195, EIT284, EIT304, HMIIGR y HMIMAG, las cuales proporcionan información detallada de la dinámica en varias capas del Sol. Tradicionalmente, la predicción del número de manchas solares se ha basado en métodos como ARIMA, LSTM, regresiones y otros modelos de machine learning que utilizan series de tiempo de sunspot del SILSO y condiciones del medio interplanetario. Sin embargo, en este reto, se propone utilizar las imágenes digitales del Sol para desarrollar modelos más robustos y comprensivos. Las imágenes espectrales capturan características únicas que pueden mejorar significativamente la precisión y profundidad de los modelos predictivos, superando las limitaciones de los enfoques basados únicamente en series de tiempo.</p>	
<b>Objetivos y motivación del reto</b>	
<p>El objetivo principal del reto es desarrollar modelos de inteligencia artificial que utilicen imágenes espectrales del SOHO para caracterizar y predecir la completitud del ciclo solar 25. Se espera que estos modelos proporcionen una visión más detallada y precisa de la dinámica solar, aprovechando la información única contenida en las imágenes espectrales en diferentes filtros solares. La motivación para este reto radica en la necesidad de mejorar los modelos predictivos actuales, que a menudo se basan en datos limitados a series de tiempo y no capturan completamente la complejidad de los fenómenos solares. Utilizando imágenes espectrales, se puede obtener una representación más rica y detallada del comportamiento solar, lo cual es crucial para diversas aplicaciones científicas y tecnológicas, incluyendo la predicción de tormentas solares y su impacto en sistemas aeroespaciales y terrestres. Este reto está alineado con los temas del hackathon Co-Afina, ya que aborda una problemática científica utilizando datos abiertos y tecnologías avanzadas para desarrollar soluciones innovadoras. La herramienta propuesta no solo contribuirá al campo de la heliofísica, sino que también promoverá el uso de inteligencia artificial en la investigación científica, fomentando la innovación y el avance tecnológico.</p>	
<b>Datasets del reto, ubicación y descripción</b>	
"Para el desarrollo del reto, se utilizarán conjuntos de datos abiertos del Observatorio	

Solar y Heliosférico (SOHO) y otras fuentes relevantes. Las imágenes espectrales del SOHO están disponibles en el archivo de datos del SOHO, específicamente en la sección de reprocesamiento completado. Estas imágenes incluyen EIT171, EIT195, EIT284, EIT304, HMIIGR y HMIMAG, que capturan diferentes aspectos de la dinámica solar en varias capas del Sol. La ubicación de estos datos es <https://soho.nascom.nasa.gov/data/REPROCESSING/Completed/>. Los datos están en formato JPEG, un formato estándar en la astronomía que facilita el manejo y análisis mediante diversas herramientas de procesamiento de imágenes. La licencia de estos datos permite su uso abierto, reutilización y redistribución, cumpliendo con los principios internacionales de acceso abierto. El tamaño aproximado de los conjuntos de datos puede variar, pero generalmente cada imagen puede tener decenas de megabytes, acumulando varios gigabytes para períodos prolongados de observación. Los metadatos incluyen descripciones detalladas de cada imagen, su filtro específico, fecha y hora de captura, y otros parámetros relevantes que aseguran una comprensión completa de la información contenida.

Para el número de sunspot se puede encontrar datos muy buenos y de buena calidad en: <https://www.sidc.be/SILSO/datafiles>"

#### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

Además se pueden encontrar datos de contexto en <https://omniweb.gsfc.nasa.gov/ow.html>

#### **Criterios para evaluar la solución del reto [opcional]**

Para evaluar las soluciones propuestas en este reto, se considerarán varios parámetros clave. En primer lugar, la pertinencia y precisión de los modelos desarrollados para predecir la dinámica del ciclo solar 25 utilizando imágenes espectrales serán fundamentales. Se evaluará la capacidad de los modelos para capturar y predecir patrones y tendencias en las imágenes solares. La originalidad y creatividad en el uso de técnicas de machine learning e inteligencia artificial también serán cruciales, así como la viabilidad técnica de las soluciones propuestas, asegurando que sean implementables y efectivas. La claridad y detalle en la presentación de las soluciones serán importantes para entender completamente la metodología y resultados obtenidos. Además, el cumplimiento de normativas y principios de acceso abierto, junto con la participación activa del equipo proponente durante el hackathon, serán factores determinantes en la evaluación. Se espera que las soluciones no solo proporcionen predicciones precisas, sino que también ofrezcan una plataforma accesible y útil para la comunidad científica y educativa interesada en la dinámica solar y sus impactos.

#### **Algún comentario adicional**

## Microplásticos: Un Desafío Ciudadano

Retador.a	Marga L Rivas e Iskya Garcia
Correo	<a href="mailto:giskya@gmail.com">giskya@gmail.com</a> , <a href="mailto:mlrivas@uca.es">mlrivas@uca.es</a>
Institución	Universidad de Cadiz y Creative Commons Venezuela
<b>Resumen del reto</b>	
Desarrollar una aplicación móvil o web que permita a los ciudadanos identificar microplásticos utilizando un banco de imágenes de licencia abierta. Esta herramienta facilitará el muestreo de la cantidad de microplásticos en diferentes lugares, permitiendo a científicos y organizaciones estimar el impacto ambiental de este problema.	
<b>Objetivos y motivación del reto</b>	
<p>"Motivación: Los microplásticos son fragmentos de plástico menores a 5 mm que se acumulan en el medio ambiente, representando una grave amenaza para la vida marina y la salud humana. La contaminación por microplásticos es un problema global que requiere soluciones urgentes.</p> <p>Objetivos: *Desarrollar una herramienta accesible para la identificación de microplásticos en la arena. *Facilitar el muestreo de microplásticos por parte de ciudadanos. *Generar datos valiosos para estimar el impacto ambiental de la contaminación por microplásticos. *Promover la participación ciudadana en la ciencia y la conservación del medio ambiente."</p>	
<b>Datasets del reto, ubicación y descripción</b>	
<p><a href="https://universe.roboflow.com/panats-mp-project/microplastic-dataset">https://universe.roboflow.com/panats-mp-project/microplastic-dataset</a> , formato: imágenes, licencia: CC0 1.0 UNIVERSAL, tamaño aprox.: 30 Mb, Versión 19, fecha: 2022-01-02 , autor: Panats MP Project</p>	
<b>Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]</b>	
<p>Habilidades deseables: Machine Learning y desarrollo de aplicaciones. Ejemplos: <a href="https://www.kaggle.com/code/mathieuduverne/microplastic-detection-detectron2-map-50-70">https://www.kaggle.com/code/mathieuduverne/microplastic-detection-detectron2-map-50-70</a> , <a href="https://colab.research.google.com/drive/16jcaJoc6bCFAQ96jDe2HwtXj7BMD_-m5">https://colab.research.google.com/drive/16jcaJoc6bCFAQ96jDe2HwtXj7BMD_-m5</a></p>	
<b>Criterios para evaluar la solución del reto [opcional]</b>	
Facilidad de uso y accesibilidad: Evaluar la simplicidad de la interfaz, la claridad de las	

instrucciones y la disponibilidad de la herramienta en diferentes plataformas (móvil, web, multiplataforma). Precisión de la identificación: La aplicación debe ser capaz de identificar correctamente los microplásticos con un alto grado de precisión, minimizando los errores tanto de falsos positivos como de falsos negativos. Generación de datos: Debe permitir a los usuarios recopilar y compartir datos sobre la ubicación y la cantidad de microplásticos encontrados, de manera organizada y estructurada.

**Algún comentario adicional**

## No es cuento: Análisis de narrativas y sentimientos sobre la migración venezolana en periódicos colombianos.

Retador.a	Mairene Tobón Ospino
Correo	mairenetobon@gmail.com
Institución	Fundación Entre Dos Tierras
<b>Resumen del reto</b>	
<p>"Los medios de comunicación tienen la capacidad de moldear percepciones y actitudes, por esta razón es importante promover una cobertura mediática informativa equilibrada y responsable, particularmente en cuestiones sensibles. La migración venezolana ha sido un tema recurrente en los medios de comunicación colombianos, generando diversas reacciones y opiniones. Este reto busca desarrollar una metodología para analizar las narrativas y sentimientos sobre migración venezolana en Colombia, a partir de las noticias publicadas, en los principales periódicos de las ciudades colombianas que acogen a más del 55% de la población migrante venezolana en Colombia, con el fin de detectar posibles patrones narrativos que estigmatizan a la población migrante o que por el contrario, contribuyen a la cohesión social y confianza.</p> <p>Hasta ahora, no existe una aplicación que permita compilar un rastreo de noticias de esta naturaleza pese a la evidencia de modelos eficientes para tal fin. Se espera que los participantes puedan emplear técnicas de web scraping para extraer noticias y comentarios de usuarios de fuentes periodísticas, y así: a) realizar un análisis de sentimientos; b) identificar sesgos y narrativas predominantes, y d) detectar temas subyacentes relacionados con migración. Este análisis permitirá comprender los temas en los que se relaciona a la migración venezolana, así como los sentimientos que expresan los usuarios a través de los comentarios de las noticias web. Se espera que la solución al reto sea capaz de proporcionar visualizaciones claras y comprensibles de los resultados, facilitando la identificación de patrones; tendencias en la percepción pública por ciudad, alertas tempranas sobre aumento de conversaciones potencialmente xenófobas y finalmente, un mapa de calor con la carga emocional de las noticias.</p>	
<b>Objetivos y motivación del reto</b>	
<p>El objetivo principal del reto es desarrollar una metodología que permita analizar las noticias y comentarios de usuarios sobre la migración venezolana en Colombia en los principales periódicos de las ciudades piloto, utilizando técnicas avanzadas de web scraping y análisis de sentimientos con inteligencia artificial. Se espera que los participantes puedan identificar y caracterizar los sentimientos generados, las narrativas predominantes, y los temas subyacentes relacionados con xenofobia, racismo, odio, y discriminación. La motivación detrás de este reto radica en la necesidad de comprender cómo se percibe y discute la migración venezolana en los medios colombianos, debido a la influencia que tienen los medios de comunicación en la construcción de imaginarios sociales y así, promover un discurso más inclusivo y</p>	

libre de discriminación. Al proporcionar una herramienta que analice estos aspectos de manera detallada y precisa, se espera contribuir a la creación de estrategias de comunicación que favorezcan una convivencia más armoniosa y respetuosa en la sociedad colombiana. Este reto está alineado con los temas del hackathon Co-Afina, ya que aborda una problemática social relevante utilizando datos abiertos y tecnologías avanzadas para desarrollar soluciones innovadoras que promuevan el bienestar social y la inclusión.

#### **Datasets del reto, ubicación y descripción**

"Para el desarrollo del reto, los participantes deberán utilizar técnicas de web scraping para extraer noticias y comentarios de usuarios de periódicos colombianos de Bogotá, Barranquilla, Medellín, Cartagena, Cúcuta, Bucaramanga, Santa Marta, Cali y Riohacha,

EL TIEMPO [www.eltiempo.com](http://www.eltiempo.com);

EL HERALDO [www.elheraldo.co](http://www.elheraldo.co);

EL UNIVERSAL [www.eluniversal.com.co](http://www.eluniversal.com.co);

EL COLOMBIANO [www.elcolombiano.com](http://www.elcolombiano.com);

DIARIO DEL NORTE <https://diariodelnorte.net/>;

HOY DIARIO DEL MAGDALENA <https://hoydiariodelmagdalena.com.co/>;

LA OPINIÓN [www.laopinion.com.co](http://www.laopinion.com.co);

VANGUARDIA [www.vanguardia.com](http://www.vanguardia.com);

EL PAÍS [www.elpais.com.co](http://www.elpais.com.co)

Estos datos deberán ser recolectados de las secciones de noticias y comentarios relacionadas con la migración y movilidad humana de venezolanos, generadas en el segundo trimestre de 2024.

Los datos extraídos estarán en formato de texto. La licencia de estos datos dependerá de las políticas de uso de cada sitio web. El tamaño aproximado de los conjuntos de datos puede variar según el volumen de noticias y comentarios recolectados, pero se espera que sean manejables con técnicas estándar de procesamiento de texto. Los metadatos deberán incluir información como la fuente de la noticia, la fecha de publicación, el contenido textual, y cualquier otra información relevante que pueda ayudar en el análisis de sentimientos y discurso."

#### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

#### **Criterios para evaluar la solución del reto [opcional]**

La precisión y efectividad de las técnicas de web scraping. Del mismo modo, la capacidad de los modelos de llevar a cabo un análisis de sentimientos detallado, identificar los patrones, las narrativas predominantes y detectar temas subyacentes serán un parámetro importante para evaluar. La originalidad y creatividad de la aplicación de técnicas de análisis de sentimiento permitirán evaluar también las soluciones propuestas. Además, la viabilidad técnica de la implementación de las soluciones para asegurar que sean factibles y que sean bastantes serán determinantes. La claridad y el detalle en la forma en que se presentan las soluciones también serán

críticos para una comprensión adecuada de la metodología y los resultados. Las soluciones también deben diseñarse en una forma que facilite que los resultados sean fáciles de interpretar y comprender los resultados.

### **Algún comentario adicional**

Para establecer la carga emocional de noticias o comentarios sobre la migración en periódicos digitales, se pueden identificar palabras clave que reflejen diferentes tonalidades emocionales. Las palabras con carga negativa incluyen términos como: crisis, invasión, amenaza, delincuencia, problema, ilegal, conflicto, veneco(a), sobrecarga, desempleo y peligro. Estas palabras suelen asociarse con percepciones adversas y estigmatizantes. Por otro lado, las palabras con carga positiva abarcan conceptos como: contribución, integración, diversidad, enriquecimiento, solidaridad, apoyo, oportunidad, talento, emprendimiento, cohesión, riqueza, desarrollo e inclusión, que destacan aspectos beneficiosos y constructivos de la migración. Finalmente, las palabras neutras, como: migración, desplazamiento, movimiento, tránsito, cambio, población migrante, extranjero, refugiados, retornados, entre otras. Las estadísticas, datos e informes citados en noticias presentan una visión más objetiva y descriptiva del fenómeno migratorio en las ciudades propuestas para abordar en el reto. Estas categorías de palabras clave ayudan a identificar y clasificar el tono emocional de los contenidos relacionados con la migración en los medios digitales.

## Grasp classification based on electromyographic (EMG) signals

Retador.a	Iris Kyranou
Correo	iris.kyranou@cern.ch
Institución	CERN, Open Data Group
<b>Resumen del reto</b>	
Use publicly available muscle activity signals recorded from the surface of the skin of healthy people performing a set of specific grasps that correspond to holding different objects, and classify the signals to the intended grasps.	
<b>Objetivos y motivación del reto</b>	
<p>"Grasp classification based on EMG signals has been explored for a while by research to control of robotic prosthetic hands. Even though it performs well in controlled experimental environments, this technique is not used in commercially available prostheses due to the performance of the algorithm being very sensitive to changes over time of the signal and the difficulty to generalize between different people. In this challenge you are going to explore the following tasks:</p> <ol style="list-style-type: none"><li>1. Familiarize yourself with the machine learning pipeline for grasp recognition based on the EMG signals:<ul style="list-style-type: none"><li>- Raw signal pre-processing</li><li>- Design of classification algorithm</li><li>- Test the classification algorithm</li><li>- Explain the results (visualization, quantification)</li></ul></li><li>2. Explore the between different sessions' (a or b) generalization of your algorithm<ul style="list-style-type: none"><li>- Can you use the algorithm trained on the 'a' set to classify the grasps from the same subjects' 'b' set?</li><li>- Can you adapt the algorithm to improve the performance of the classifier?</li></ul></li><li>3. Explore the between different subjects' generalization of your algorithm<ul style="list-style-type: none"><li>- Can you adapt the trained algorithm on one (or more) subject(s) to recognize the grasps performed by a subject whose data you have not used in the training?</li></ul></li></ol> <p>The recommendation is to pick only one of the 2. and 3. tasks and focus on this. If you think you have found a good solution for the chosen task you can try your solution to the other task and report on the results."</p>	
<b>Datasets del reto, ubicación y descripción</b>	
<p>"The datasets can be found on the Ninapro website: <a href="https://ninapro.hevs.ch/instructions/DB6.html">https://ninapro.hevs.ch/instructions/DB6.html</a>. Each folder of this dataset contains files of the format 'Sx_D1_Ty.mat', where x is the subject number and y is the time of the day</p>	

that the recording corresponds to. On the website you can find pictures showing the acquisition protocol, the grasps performed during the exercise and the description of the information stored in the matlab file.

The dataset includes more information than just the electromyography (EMG) signals. For this challenge focus first on the EMG signal. After you have a working solution for EMG you can try to add information from other sensory modalities (like accelerometers) that are also included in the dataset. "

#### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

"Notice that part of the challenge is the preparation of the raw signal to be of the correct format to be used for classification. This process includes:

- Signal filtering
- Signal normalization
- Windowing
- Extract features (might not be needed, depending on the algorithm you are using)

A python library that can be found useful that implements the above, specifically for the ninapro dataset, can be found here: <https://pypi.org/project/nina-funcs/>

The papers that accompany the datasets are a good source of information on the process of recording the data and building algorithms for the task on hand."

#### **Criterios para evaluar la solución del reto [opcional]**

The visualization of the results is up to you. Report any experiment you are performing, for example if you explore more than one algorithm or if you implement the between-subjects objective, and try to justify your observations.

#### **Algún comentario adicional**

## Buscador de obras en dominio público

Retador.a	Jose Luis Mendoza
Correo	jluismendoza@clisi.org
Institución	Centro Latinoamericano de Investigaciones Sobre Internet
<b>Resumen del reto</b>	
<p>"Las obras en dominio publico son aquellas que, de acuerdo a la legislación nacional, salen del monopolio de la propiedad intelectual y conforman un acervo cultural común a la población mundial, por lo que cualquiera puede hacer uso de estas obras sin permiso expreso, licencia o cualquier acuerdo previo. Esto permite el desarrollo cultural y científico de los pueblos ya que representan acceso gratuito y abierto a obras y materiales creados que cumplan los requisitos legales. En el caso de Venezuela, se consideran obras en dominio publico todas aquellas pertenecientes a un autor que haya muerto 60 años atrás, es decir, a partir de la muerte del autor, o del ultimo de los autores en el caso en que sean varios autores de la misma obra, se comienza un computo de 60 años continuos a partir del primero de enero del año inmediatamente siguiente, cumplido este plazo, la obra entra en dominio publico (por ello a nivel mundial se celebra el 1 de enero de cada año el día del dominio publico, ya que es el día en que entra todo un nuevo bloque de obras a dicha área.</p> <p>A partir de los datos de Wikipedia, y en una alianza entre Creative Commons Venezuela y Wikidata, en 2020 se creo una herramienta de consulta que se actualiza automáticamente y genera un listado de los venezolanos que cumplen con el requisito anterior y que por consiguiente sus sobras están en dominio publico, y podemos usarla de manera gratuita y abierta.</p> <p>Ahora, este listado es un poco crudo, creo que a partir de los mismos datos abiertos (wikimedia) se puede lograr algo mas amigable con el usuario final que quiera saber si puede o no usar determinada obra de un autor, incluso que se pueda escoger la nacionalidad del autor y así ajuste el plazo para que no se restrinja solo a los venezolanos"</p>	
<b>Objetivos y motivación del reto</b>	
<p>Las obras en dominio publico son de gran importancia para la cultura abierta ya que permiten su uso gratuito y sin condiciones, por ello es importante determinar que obras están o no en dominio publico para poderlas utilizar. Existe un nivel añadido de dificultad si incluimos que algunos autores utilizar la llamada Licencia Zero de Creative Commons, la cual les permite que su obra, bajo esa licencia, entre inmediatamente al dominio publico, sin esperar su muerte ni el computo de ningún plazo, pero esta característica no se encuentra recogida en ninguna base de datos como para poderla incorporar. La importancia educativa de poder determinar si una obra esta en dominio publico, y en consecuencia se puede usar, es gigantesca, ya que facilita su uso en aula o dentro de las practicas de los estudiantes</p>	

**Datasets del reto, ubicación y descripción**

"Los datos en Wikidata están publicados bajo la licencia Creative Commons Dedicación al Dominio Público 1.0, lo que permite la reutilización de los datos en muchos escenarios diferentes. Puede copiar, modificar, distribuir y presentar los datos, incluso con fines comerciales, sin necesidad de pedir permiso.

<https://www.wikidata.org/wiki/Wikidata:Introduction/es#:~:text=Los%20datos%20en%20Wikidata%20est%C3%A1n,sin%20necesidad%20de%20pedir%20permiso.>

En esta base de datos reposan todos los datos que se utilizan en los diversos proyectos de la fundación wikimedia, con la gran ventaja que pueden ser editados, o creados nuevos, por cualquiera usuario, lo que permite incluir mas autores y creadores y así hacer mas precisa la herramienta que se desarrolle"

**Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

"una guía pudiera ser lo que hemos desarrollado hasta ahora

<https://w.wiki/EkU>

que se encuentra reseñada en:

<https://ve.creativecommons.net/2021/02/09/editathon-2020-dominio-publico/>

"

**Criterios para evaluar la solución del reto [opcional]**

Una herramienta capaz de dar una respuesta afirmativa o negativa al introducir el nombre completo y nacionalidad de un autor, artista, creador o académico. Basándose en la normativa nacional de propiedad intelectual, según la nacionalidad del autor, y los tratados internacionales aplicables

**Algún comentario adicional**

## ATLAS Open data

Retador.a	Carlos Sandoval, Dilia
Correo	<a href="mailto:cesandovalu@unal.edu.co">cesandovalu@unal.edu.co</a>
Institución	ATLAS CERN
<b>Resumen del reto</b>	
<p>"Choose one (or more) of the 13 TeV Open data python examples listed here: <a href="https://nbviewer.org/github/atlas-outreach-data-tools/notebooks-collection-opendata/tree/master/13-TeV-examples/python/">https://nbviewer.org/github/atlas-outreach-data-tools/notebooks-collection-opendata/tree/master/13-TeV-examples/python/</a> depending on the physics interest of ht eparticipants. For example, the full Hyy analysis using the 13 TeV dataset (<a href="https://nbviewer.org/github/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/python/ATLAS_OpenData_13-TeV_python_full_HyyAnalysis_5min.ipynb">https://nbviewer.org/github/atlas-outreach-data-tools/notebooks-collection-opendata/blob/master/13-TeV-examples/python/ATLAS_OpenData_13-TeV_python_full_HyyAnalysis_5min.ipynb</a>) is a nice candidate. Also, this notebook is very useful: <a href="https://www.kaggle.com/code/meirinevans/how-to-rediscover-the-higgs">https://www.kaggle.com/code/meirinevans/how-to-rediscover-the-higgs</a> .</p>	
<b>Objetivos y motivación del reto</b>	
<p>"1) Follow and ensure the student can run the example as it is and identify the parts that technically and scientifically are more complex  2) Improve the physics explanation to make it more pedagogically accessible. Consider also add equations, references and diagrams to improve the documentation.  4) Propose intermediate plots or other final plots to help understand the final state or the result and extend the investigation further  5) Also provide the documentation in Spanish to extend the reach of these resources. We think this is feasible in 3 days and very useful to the ATLAS outreach group. But if the students want to go further, and they have a good background in python and have notions of ATLAS analysis, they can also propose another final state and create their own example.</p>	
<b>Datasets del reto, ubicación y descripción</b>	
<p><a href="https://nbviewer.org/github/atlas-outreach-data-tools/notebooks-collection-opendata/tree/master/13-TeV-examples/python/">https://nbviewer.org/github/atlas-outreach-data-tools/notebooks-collection-opendata/tree/master/13-TeV-examples/python/</a></p>	
<b>Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]</b>	
<p></p>	
<b>Criterios para evaluar la solución del reto [opcional]</b>	
<p>"¿Qué limitaciones encontró el equipo con respecto a la versión actual?  ¿Fue la explicación física del análisis mejorada basados en dichas limitaciones?  ¿Adaptada al público: estudiantes de bachillerato/universidad?  Implementación</p>	

¿Se hizo un update de la documentación? La documentación tiene dos enfoques:

Documentación del código

Se documentó el código en un repositorio, página web, notebook?

Es asequible al público general, ordenada y clara?

¿se hizo un full implementation o solo una prueba de concepto?

Documentación conceptual de la física

¿Cómo se hizo? ¿nuevos plots/figuras? ¿videos?

Otro productos de documentacion: manual para docentes, introducción conceptual

Documentación en español

Calidad de la traducción

Implementación de la misma

Bonus: El producto final está en ambos idiomas: español e inglés?

MEGA bonus: Se implementó un NUEVO análisis, i.e. nuevo estado final, nuevo canal de decaimiento del Higgs o nuevos procesos?

"

**Algún comentario adicional**

## Maestro del Arte: Descubre el Pintor y su Estilo

Retador.a	Mariana Vivas
Correo	marianaivivas@gmail.com
Institución	
<b>Resumen del reto</b>	
<p>En el desafío "Maestro del Arte", los participantes deberán desarrollar una aplicación basada en inteligencia artificial, específicamente en tecnologías de transferencia de aprendizaje. Esta debe ser capaz de analizar imágenes de pinturas y predecir tanto el autor como el estilo artístico de cada una. Utilizando la base de datos de WikiArt, los equipos crearán modelos de aprendizaje automático que identificarán patrones y características distintivas en las obras de arte.</p>	
<b>Objetivos y motivación del reto</b>	
<p>"El reto ""Maestro del Arte"" tiene como objetivo desarrollar una aplicación de aprendizaje automático que pueda analizar imágenes de pinturas y predecir tanto el autor como el estilo artístico utilizando datos abiertos de WikiArt. Los participantes deberán crear modelos de aprendizaje automático que identifiquen patrones y características distintivas en las obras de arte para realizar estas predicciones.</p> <p>La motivación detrás de este reto es combinar el arte y la tecnología para ofrecer una solución que pueda ser utilizada en contextos educativos. Al desarrollar una herramienta que permite identificar el autor y el estilo de una pintura, se busca descubrir si existen características intrínsecas en los pintores y los estilos. La clasificación facilita el estudio y la apreciación del arte, proporcionando a estudiantes y profesores una herramienta para el análisis artístico. Esto no solo fomenta la creatividad y el pensamiento crítico entre los participantes, sino que también proporciona un recurso valioso para la educación artística.</p> <p>El reto se relaciona con los temas del hackathon Co-Afina en varios aspectos. En primer lugar, promueve el uso de datos abiertos para resolver problemas educativos, alineándose con la misión de Co-Afina de fomentar soluciones colaborativas en ciencia y educación. Además, tiene el potencial de generar una mejora en la educación artística, ya que la aplicación creada permitirá a los estudiantes identificar fácilmente el autor y el estilo de una pintura, simplificando el estudio de diferentes artistas y movimientos. La originalidad de este reto radican en la idea de utilizar tecnologías de inteligencia artificial en un campo tradicionalmente analógico como el arte. Además existe suficiente documentación, sin caer en un problema demasiado popular, o sobre analizado. "</p>	
<b>Datasets del reto, ubicación y descripción</b>	

"Ubicación: dataset WikiArt en kaggle (<https://www.kaggle.com/datasets/steubk/wikiart>).

Formato: Carpetas con imágenes

Licencia: WikiArt es de dominio público.

Tamaño aproximado: 34Gb

Descripción de metadatos: La descripción completa del dataset se encuentra en kaggle. En el repositorio original del dataset (<https://github.com/cs-chan/ArtGAN/tree/master/WikiArt%20Dataset>) hay información para la organización de los datos para entrenamiento y validación (que los participantes podrían decidir utilizar o no, dependiendo de que tan familiarizados estén con trabajar directorios de imágenes, les podría ser útil)"

### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

"Librerías/habilidades sugerida: Conocimiento sobre redes neuronales. TensorFlow o PyTorch para hacer transfer learning. Alguna herramienta para hacer una app sencilla, como streamlit.

Cómputo mínimo: Se propondría el uso de kaggle directamente, o google colab (<https://www.freecodecamp.org/news/how-to-download-kaggle-dataset-to-google-colab/>) para desarrollar el modelo, ya que habilita el uso de GPU, lo que ahorra tiempo al momento de entrenar los modelos, y elimina algunas limitaciones de cómputo local. También se evita el problema de descargar los archivos localmente.

Referencias bibliográficas:

- [https://www.tensorflow.org/tutorials/images/transfer\\_learning](https://www.tensorflow.org/tutorials/images/transfer_learning)
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9173940/>
- <https://www.kaggle.com/code/steubk/wikiart-top25-artists-classifier-baseline>
- <https://arxiv.org/pdf/1812.08052>"

### **Criterios para evaluar la solución del reto [opcional]**

"El accuracy, que mide la frecuencia con la que un modelo de aprendizaje automático predice correctamente el resultado, es una buena métrica para este reto. A través del accuracy se pueden comparar los resultados con resultados previos de distintos modelos:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0248414>

Si rehacen alguno de los modelos ya probados, se puede intentar llegar al mismo accuracy o mejorarlo"

### **Algún comentario adicional**



## Reconocimiento ciudadano de las transformaciones de humedales en Colombia

Retador.a	Alejandra Melfo, Oscar Altuve y Yelitza León
Correo	alejandramelfo@gmail.com, altuveoscar95@gmail.com, yleon64@hotmail.com
Institución	Universidad de Los Andes (ULA), Universidad Simón Bolívar (USB) y Universidad de Los Andes (ULA)
<b>Resumen del reto</b>	
<p>Entre los ecosistemas de necesaria protección en el mundo, los humedales están identificados como zonas a defender, dada su importancia en la mitigación y adaptación al cambio climático. Estas extensiones, naturales o artificiales, se encuentran distribuidas en el mundo cubiertas de agua, de manera permanente o temporal, con una profundidad no mayor a seis metros, según la convención internacional Ramsar. Dada su importancia ecológica, se han generado acuerdos entre los países para generar políticas públicas de protección ambiental que designe y reconozca a los humedales como zonas de preservación necesaria, pues, además de aportar agua, mantienen biodiversidad y son los más eficientes sumideros de carbono, claves para mitigar el cambio climático. Implica un desafío evitar que estos ecosistemas desaparezcan con el tiempo, sobre todo por la incidencia de la actividad humana. En este reto se busca usar diversas bases de datos abiertas relacionadas con la identificación de humedales en Colombia, sus cambios en el tiempo y su ubicación dentro de las áreas naturales protegidas pertenecientes al Sistema de Parques Nacionales Naturales. Dado que muchos de los humedales podrían encontrarse en zonas con distintas condiciones de intervención humana, esta visualización organizada en una aplicación digital podría convertirse en una herramienta útil para su uso en términos ciudadanos, de investigación y decisiones en políticas públicas.</p>	
<b>Objetivos y motivación del reto</b>	
<p>"Objetivo general: Generar conciencia ciudadana sobre los humedales en Colombia, sus posibles cambios y los factores de riesgo en su preservación, desde una herramienta de uso amigable para la identificación de estas zonas, su ubicación respecto a áreas de actividad humana o zonas de protección ambiental, y la visualización de su transformación en el tiempo.</p>	
<b>Datasets del reto, ubicación y descripción</b>	
<p>"-&gt; Web de los sitios Ramsar en el mundo: <a href="https://rsis.ramsar.org/">https://rsis.ramsar.org/</a> -&gt; Ejemplo de timelapses: <a href="https://earthengine.google.com/timelapse/">https://earthengine.google.com/timelapse/</a> (buscar lugar de interés y presionar en ""Share Content"" analizar los enlaces). -&gt; Municipios, departamentos, población,...:</p>	

<https://geoportal.dane.gov.co/geovisores/sociedad/cnpv-2018/>

--

<https://geoportal.dane.gov.co/servicios/descarga-y-metadatos/datos-geoestadisticos/?cod=4>

Parques nacionales naturales de Colombia:

<https://datosabiertos.esri.co/datasets/esri-colombia::parques-nacionales-naturales-de-colombia/about>

-> (Enlace alternativo) Datos de población de Colombia y ubicación:

<https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/herramientas#:~:text=Es%20una%20herramienta%20que%20permite%20consultar%20y%20descargar,del%20Censo%20Nacional%20de%20Poblaci%C3%B3n%20y%20Vivienda%202018.>"

### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

"Ejemplo de código:

-> Visualizar mapas: <https://www.kaggle.com/code/alfredomaussa/colombia-municipios>

Web para curiosear:

->

<https://github.com/nick-murray/gee-global-intertidal-change-data-descriptor/tree/main>"

### **Criterios para evaluar la solución del reto [opcional]**

"Objetivos específicos:

- Clasificar humedales en Colombia, reconocidos en la web de sitios Ramsar, según su ubicación, tomando como criterio las zonas de protección ambiental.

- [Opcional] Clasificar los sitios Ramsar según la densidad de población en los municipios que lo rodean.

- Construir/vincular timelapses con datos disponibles en la red de los sitios Ramsar de Colombia.

- Crear una aplicación digital, página web, con un mapa interactivo de Colombia que permita observar las regiones Ramsar que están dentro y fuera de las zonas de protección ambiental (y según densidad de población); que además garantice la observación de los timelapses.

Importante: Es fundamental que el prototipo sea genérico y replicable en otros países. Por ejemplo: Brasil, Argentina, entre otros."

### **Algún comentario adicional**

## Asistente Nutricional Global

Retador.a	Arturo Sánchez Pineda y María Isabel García Lorenzo
Correo	<a href="mailto:argaby2017@gmail.com">argaby2017@gmail.com</a> ,
Institución	inait.ai
<b>Resumen del reto</b>	
Descripción del Proyecto: Los participantes deben crear una web app interactiva que permita a los usuarios hacer preguntas sobre nutrición y recibir respuestas basadas en datos extraídos de los documentos oficiales de la WHO. La solución debe integrar un sistema RAG para mejorar la precisión y relevancia de las respuestas generadas por un modelo de lenguaje grande (LLM).	
<b>Objetivos y motivación del reto</b>	
Objetivo del Reto: Desarrollar una web app que ofrezca recomendaciones nutricionales precisas y actualizadas utilizando técnicas de RAG (Retrieval Augmented Generation). La aplicación debe ser capaz de entender y responder preguntas en múltiples idiomas, utilizando documentos oficiales de la Organización Mundial de la Salud (WHO) como fuente principal de información.	
<b>Datasets del reto, ubicación y descripción</b>	
Recursos y Herramientas Sugeridas: Lenguajes de Programación: Python, JavaScript Frameworks: Flask o Django para el backend, React o Vue.js para el frontend Modelos de Lenguaje: Modelos LLM open source de Hugging Face y Meta, como Mistral y Llama Herramientas de Traducción: API de Google Translate, DeepL Herramientas de RAG: Weaviate ( <a href="https://verba.weaviate.io/">https://verba.weaviate.io/</a> )  Datos y Recursos Adicionales: Videos y Artículos: Video sobre RAG: <a href="https://www.youtube.com/watch?v=oTRrfabdk8">https://www.youtube.com/watch?v=oTRrfabdk8</a> WHO ELENA (e-Library of Evidence for Nutrition Actions): <a href="https://www.who.int/tools/elena/overview">https://www.who.int/tools/elena/overview</a> Casos de uso de RAG en Google Cloud: <a href="https://cloud.google.com/use-cases/retrieval-augmented-generation">https://cloud.google.com/use-cases/retrieval-augmented-generation</a>	

Glosario de Databricks sobre RAG:

<https://www.databricks.com/glossary/retrieval-augmented-generation-rag>

Emergency Nutrition Network (ENN): <https://www.ennonline.net/>

Food and Agriculture Organization (FAO): <https://www.fao.org/home/ru/>

Wikipedia Articles, like those from the latest Editathon:

<https://editathon.github.io/page>

<https://www.cpnh.cornell.edu/editathon>

### **Otros recursos disponibles/requeridos para el desarrollo del reto [Opcional]**

Infraestructura y Despliegue:

Los equipos que elijan este reto tendrán la oportunidad de utilizar infraestructura de GitHub.com, Google Cloud Platform (GCP) y/o Azure de Microsoft para desplegar la solución que construyan. Esto les permitirá experimentar con tecnologías en la nube y aprender a implementar soluciones escalables y robustas.

### **Criterios para evaluar la solución del reto [opcional]**

Criterios de Evaluación:

Innovación y Creatividad: Grado de innovación en la solución propuesta y su implementación.

Funcionalidad y Usabilidad: Facilidad de uso de la web app y precisión de las respuestas generadas.

Calidad Técnica: Calidad del código, implementación de la arquitectura RAG y uso de modelos de lenguaje.

Impacto y Escalabilidad: Potencial de la solución para ser utilizada en el mundo real y su capacidad para escalar.

Presentación Final:

Demo en video: Mostrar la web app en funcionamiento, respondiendo preguntas en diferentes idiomas.

Explicación Técnica: Describir brevemente cómo se implementó la arquitectura RAG y cómo se integraron los diferentes componentes.

Esta parte se debería hacer en el README del repositorio en GitHub.com ha entregar.

Impacto Potencial: Discutir cómo la solución podría ayudar a mejorar la nutrición en regiones remotas y su potencial para ser expandida.

### **Algún comentario adicional**

Funcionalidades Clave:

Carga y Procesamiento de Documentos:

Descargar y procesar documentos PDF de la WHO para extraer texto utilizable.

Preprocesar el texto para facilitar su indexación y búsqueda.

Indexación y Recuperación de Información:

Generar embeddings de los textos de los documentos.

Indexar estos embeddings en una base de datos de vectores para realizar búsquedas rápidas y eficientes utilizando herramientas como Weaviate (<https://verba.weaviate.io/>).

Implementar un sistema de recuperación de información para obtener fragmentos relevantes basados en consultas de los usuarios.

Generación de Respuestas:

Utilizar un modelo de lenguaje grande (LLM) para generar respuestas, utilizando los fragmentos recuperados como contexto adicional.

Interfaz de Usuario Multilingüe:

Desarrollar una interfaz web amigable que permita a los usuarios hacer preguntas y recibir respuestas en múltiples idiomas.

Implementar capacidades de traducción automática para soportar la interacción en diferentes idiomas.

Pruebas y Despliegue:

Realizar pruebas para asegurar la funcionalidad y precisión del sistema.

Desplegar la aplicación en una plataforma en la nube para garantizar su disponibilidad global.

