

ANALÍTICOS VISUALES CON INTELIGENCIA ARTIFICIAL APLICADOS EN CIENCIA MARINA

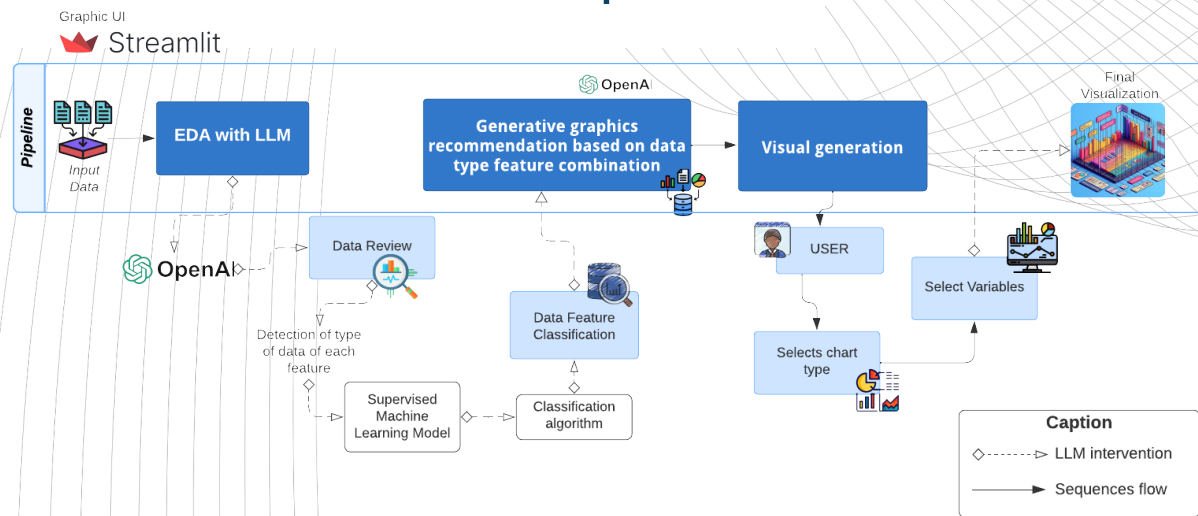
Autor: Dr. Luis Dario Ceballos - Director de Tesis: Dr. Marcos Daniel Zárate

MOTIVACIÓN

La visualización de datos presenta desafíos en la integración de datos heterogéneos y su comunicación efectiva por parte de usuarios sin formación técnica. Además existe una brecha en sistemas que unan el razonamiento semántico con la interacción visual accesible. Esta tesis busca democratizar el análisis de datos y su presentación visual de forma adecuada, reduciendo la carga cognitiva necesaria por parte del usuario.

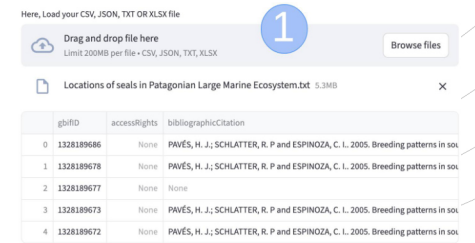


ARQUITECTURA - VisBot Pipeline



VisBot

Visualization Recommender powered with Ai



2 Ai Recommended Visualizations

Given the extensive nature and multi-dimensionality of your dataset, visualization will play a crucial role in unveiling patterns, summarizing data, and discovering insights. Here are some types of visualizations and tools you might consider to effectively represent various aspects of your dataset:

La tesis propone VisBot, un sistema basado en una interfaz interactiva con una arquitectura modular y escalable para la generación de recomendaciones inteligente de visualizaciones adecuadas según la estructura de datos analizados. Sus componentes son:

- 1- Ingesta de Datos:** Soporte para formatos CSV, JSON, Excel y datos enlazados (RDF) mediante técnicas de aplanamiento de grafos.
- 2 - EDA con LLM:** Un módulo de análisis exploratorio que utiliza GPT-4 para inferir tipos semánticos (fechas, coordenadas, taxonomías) y calcular estadísticas descriptivas.
- 3 - Motor de Recomendación:** Genera sugerencias de gráficos basadas en las características del dataset y principios de percepción visual de Cleveland y McGill.
- 4 - Generador Visual:** Presenta las recomendaciones propuestas por ia en gráficos interactivos mediante la librería Python, Plotly Express.

APORTES

Eficiencia Analítica: El sistema redujo el tiempo de completitud de tareas de un promedio de 8.5 minutos con herramientas convencionales a 3.2 minutos.

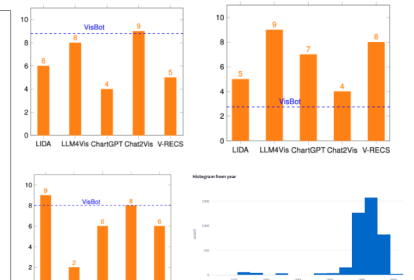
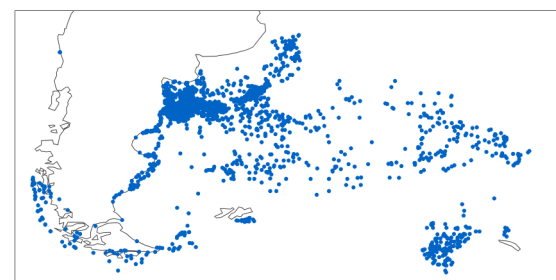
Validación Real: Aplicación exitosa en un caso de estudio sobre la biodiversidad de focas en el ecosistema patagónico (CENPAT-CONICET).

Principios FAIR: Facilita que los datos científicos sean Localizables, Accesibles, Interoperables y Reutilizables.

Explicabilidad: El sistema proporciona justificaciones textuales claras sobre sus elecciones con un 75% de precisión



Geospatial scatter map from decimalLongitude and decimalLatitude



Métrica	Valor	Dev. Est.
CALIDAD DE RECOMENDACIONES		
Relevancia (escala 1-5)	4.7	0.3
Precisión Semántica	4.8	0.2
NDCG@5	0.87	0.05
Correctitud de Mapeo	91%	4%
Diversidad de Gráficos	4.2/5	-
DETECCIÓN AUTOMÁTICA		
Variables Geoespaciales	100%	-
Campos Darwin Core	100%	-
Taxonomía Biológica	95%	-
RENDIMIENTO		
Tiempo por Tarea (min)	3.2	0.8
Tasa de Éxito	95%	-
Tasa de Error	0.8/sesión	0.4
SATISFACCIÓN		
CSAT (1-5)	4.7	0.2
SUS Score	82.3/100	3.8
Explicabilidad	81%	8%
DOMINIO MARINO		
Visual. Geoespaciales (1-5)	4.8	0.3
Patrones Estacionales (1-5)	4.5	0.4
Mapas Distribución (1-5)	4.9	0.2

Métrica	VisBot	Herramientas IA (promedio)	Herramientas Tradicionales (promedio)
DETECCIÓN AUTOMÁTICA			
Campos Darwin Core	95%	59%	N/A
Variables geoespaciales	100%	72%	Manual
Taxonomía biológica	95%	55%	N/A
EFICIENCIA TEMPORAL			
Tiempo total análisis (min)	3.7	6.8	27.8
Mejora temporal	+84%	+59%	Base
RECOMENDACIONES			
Visualizaciones automáticas	4.2 tipos	3.7 tipos	Manual
Explicaciones generadas	81%	58%	N/A
SATISFACCIÓN USUARIOS			
Satisfacción general (1-5)	4.7	4.2	4.0
Adecuación dominio marino (1-5)	4.8	3.7	4.3
Tasa éxito en tareas	95%	86%	91%
ACCESIBILIDAD			
Experiencia técnico requerido	Básico	Básico-Inter.	Avanzado
Tiempo de aprendizaje	30 min	1-2 horas	8-16 horas
Costo anual (USD/usuario)	0-50	0-100	520-1040
CASOS DE USO (minutos)			
Mapas distribución espacial	2.5	4.2	13.5
Análisis estacionalidad	3.8	4.7	19.0
Hotspots biodiversidad	2.1	4.2	11.0

Dimension	Score (avg)	Std. Dev.	Scale
Relevance	4.3	0.4	1-5
Diversity	4/5	-	count
NDCG@5	0.82	0.07	0-1
Task Success Rate	92%	5%	0-100%
Semantic Accuracy	4.6	0.3	1-5
Mapping Correctness	87%	6%	0-100%
Error Rate	1.2	0.5	per session
SUS Score	78.5	4.2	0-100
Task Time	42 sec	6 sec	seconds
Interactions	6.3	1.1	count
CSAT	4.5	0.3	1-5
NPS	+64	-	-100+100
Ambiguity Handling	3.9	0.5	1-5
Explainability	75%	10%	0-100%



LINEAS FUTURAS

Soporte 3D y Batimetría: Incorporación de la dimensión de profundidad para perfiles oceanográficos.

Modelos Multimodales: Integración de capacidades para procesar imágenes y datos no estructurados.

Entornos Colaborativos: Funciones para el análisis compartido y persistencia de sesiones en tiempo real.

