

Identificación de patrones submarinos en especies de macroinvertebrados de monitoreo ecológico



Minería de Datos
Máster en Ciencia de Datos
Curso 2022-2023

Introducción

Introducción

- Fondo marino dividido en transectos
- Generación de datos a partir de las observaciones marinas
- Fenómenos meteorológicos, cambios en el ambiente...
modifican el fondo marino
- Análisis de los datos de entre 2010 y 2020

Recolección y pre-procesamiento de datos

Recolección y pre-procesamiento de datos

- Proceso realizado columna por columna (generalmente)
- Tareas realizadas:
 - Búsqueda y análisis de valores nulos
 - Búsqueda y análisis de valores anómalos
 - Completar datos nulos/anómalos

Recolección y pre-procesamiento de datos

- id: Todo correcto, sin nulos
- dive_data, dive_month, year: Se elimina dive_month, year
- island: Se completan nulos usando Transect_code
- Transect_code: Todo correcto
- Bioregion: Hay nulos, pero no se pueden deducir
- MPA_Status: Todo correcto, se comprueba la consistencia

Recolección y pre-procesamiento de datos

- Sum_ind: No hay valores anómales
- Site: Todo correcto
- Latitude y Longitude: Todo correcto, se comprueba rango
- Subzone.name: Se eliminan los nulos
- Refuge_Level: Se eliminan los nulos
- epoca: Todo correcto

Recolección y pre-procesamiento de datos

- TaxonID, Domain, Kingdom, PhylumOrDivision, Class, Order, Family, ScientificName, CommonNameEnglish, CommonNameSpanish: Se estudian juntas, todas ellas son la taxonomía.

Se pueden extraer todos los valores a partir del ScientificName

Recolección y pre-procesamiento de datos

- TaxonID, Domain, Kingdom, PhylumOrDivision, Class, Order, Family, ScientificName, CommonNameEnglish, CommonNameSpanish

Lista de Especies de Galápagos

 / [dataZone](#) / [Lista de Especies de Galápagos](#)

En nuestra Lista de Especies de Galápagos compartimos abiertamente la información ingresada a la base de datos de las **Colecciones de Historia Natural** de la Estación Científica Charles Darwin por investigadores de todo el mundo durante varias décadas.

Para buscar información sobre una especie, puedes ingresar su nombre común o su nombre científico, filtrar por categorías de riesgo de UICN, o generar un listado de todas las especies incluidas en una de nuestras Listas de Especies. Un click sobre el nombre de la especie te llevará a su página, donde podrás encontrar contenido sobre su ecología, distribución, y taxonomía, entre otros.

Busqueda por nombre de especie o familia

Filtrar por Lista de Especies

- Seleccionar Lista de Especies - ▾

Filtrar por categoría UICN

- Seleccionar categoría UICN - ▾

Filtrar por origen

- Seleccionar origen del taxon - ▾



Recolección y pre-procesamiento de datos

- TaxonID, Domain, Kingdom, PhylumOrDivision, Class, Order, Family, ScientificName, CommonNameEnglish, CommonNameSpanish

Caretta caretta (Linnaeus, 1758)

CAGUAMA, TORTUGA BOBA, LOGGERHEAD SEA TURTLE

[Taxonomía](#) [Estado](#) [Distribución](#) [Referencias](#)

Taxonomía



Dominio	Reino	Filo	Clase	Orden	Familia	Género	Especie
Eukaryota	> Animalia	> Chordata	> Reptilia	> Testudines	> Chelonidae	> Caretta	> caretta

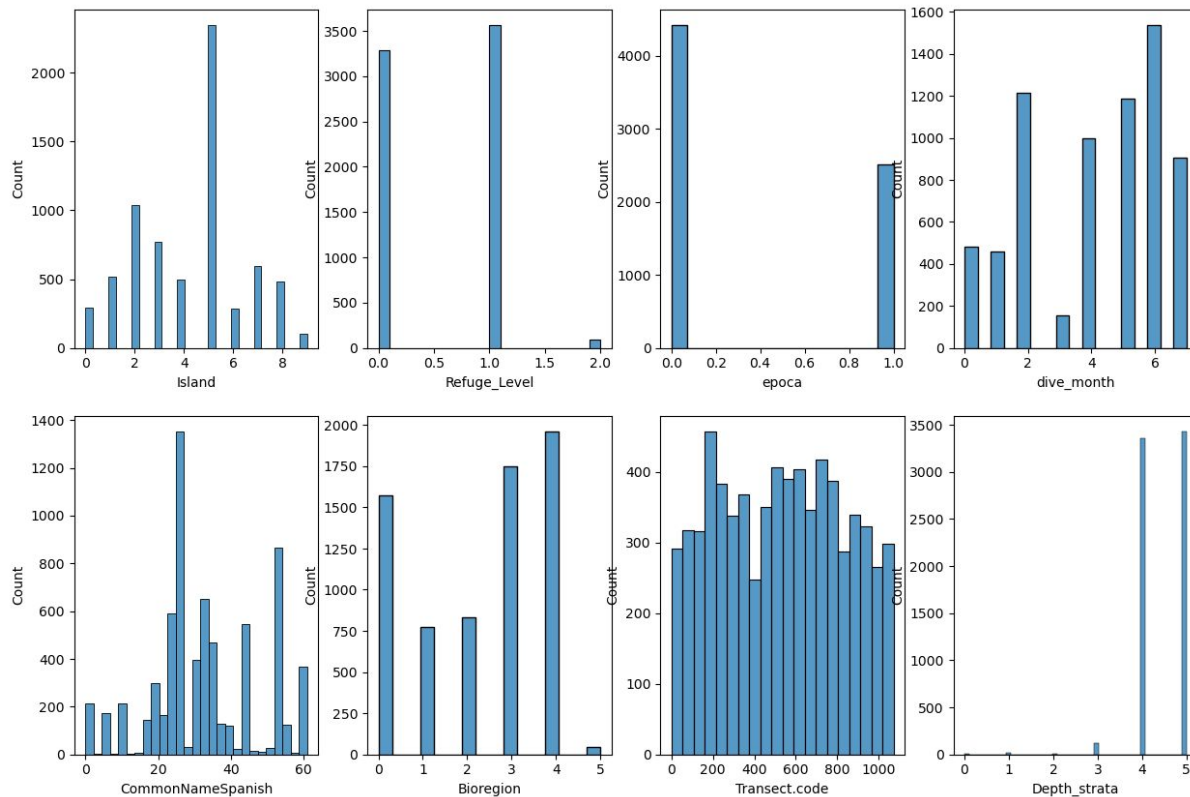
Recolección y pre-procesamiento de datos

- TaxonID, Domain, Kingdom, PhylumOrDivision, Class, Order, Family, ScientificName, CommonNameEnglish, CommonNameSpanish

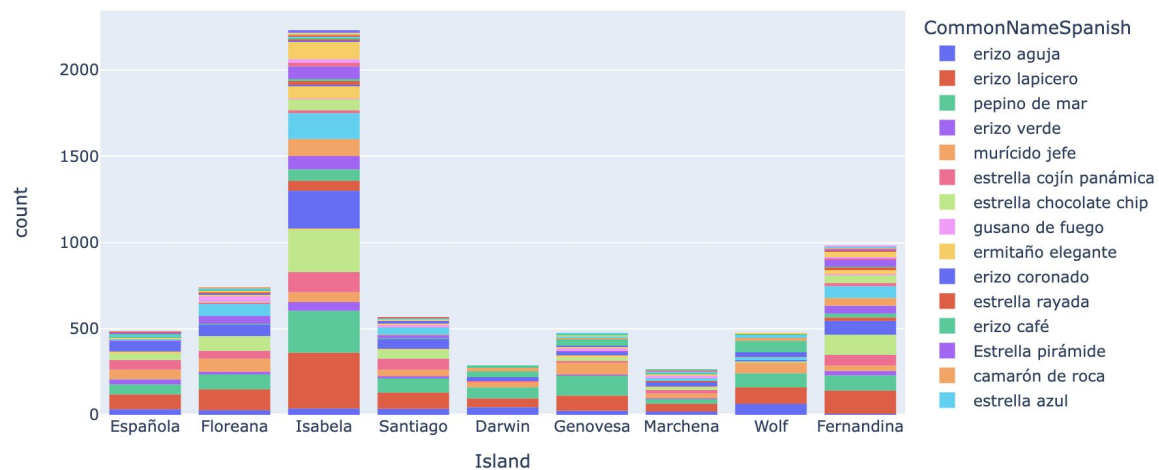
Se completan los valores nulos realizando llamadas a esa web, y 'scrapeando' los datos que se obtienen.

Distribución y visualización de los datos

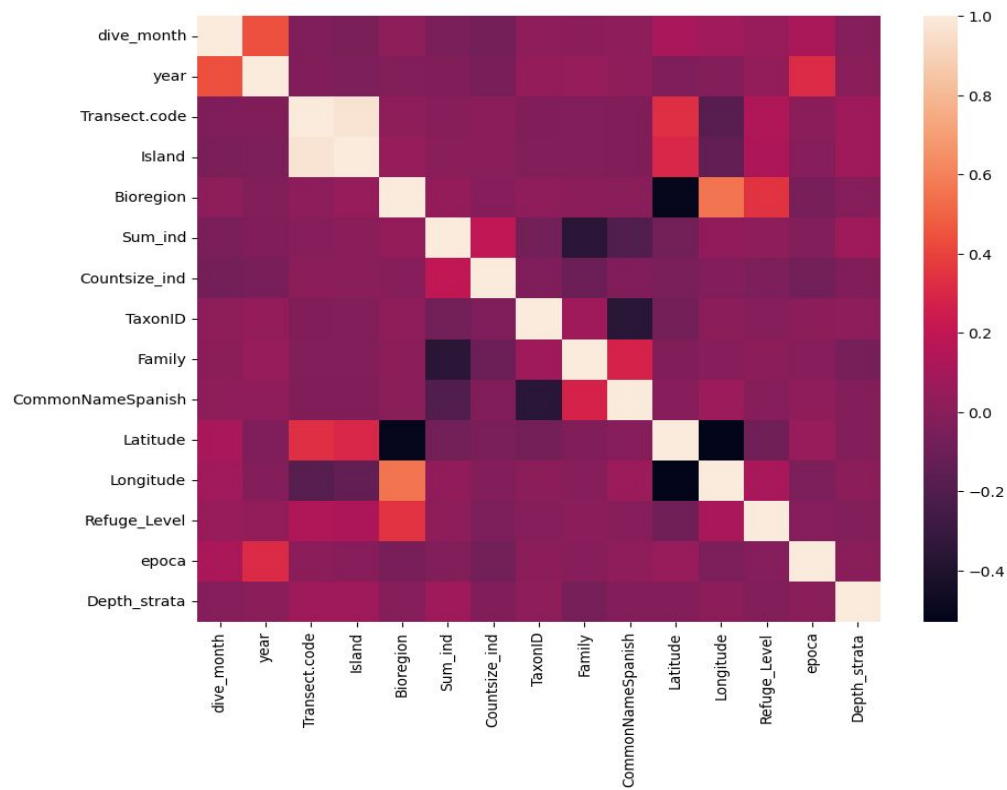
Distribución



Distribución



Correlaciones



Visualización

Data Visualization

Here you can see all the animals

Choose the epoch

Caliente

Choose the island

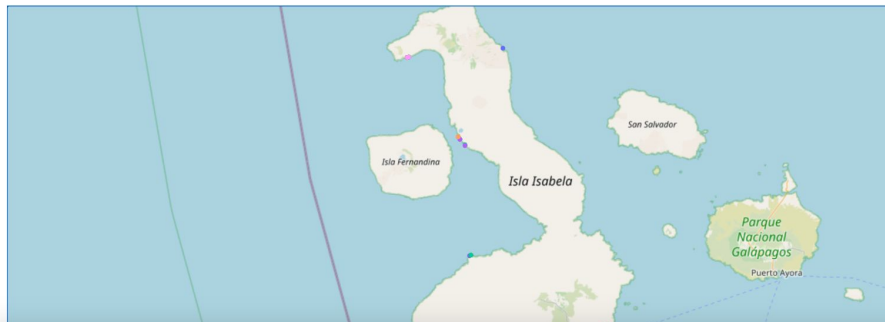
Isabela

Choose the refuge lvl

Extractive use

Choose the dive month

all



CommonNameSpanish

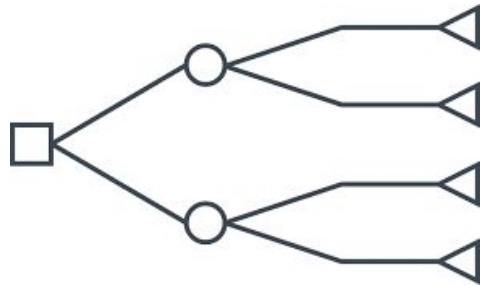
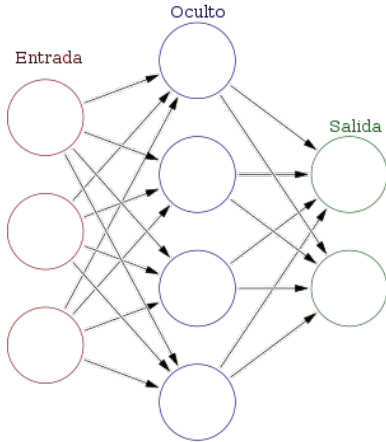
- erizo café
- erizo lapicero
- pepino de mar
- estrella chocolate chip
- cangrejo espuma
- estrella azul
- camarón de roca
- estrella cojín panámica
- estrella variable
- erizo coronado
- Estrella pirámide
- erizo aguja
- muricido jefe
- sol de mar
- erizo verde
- cono
- langosta china
- doridáceo de Mullineri
- erizo morado

Modelos

Tipos de modelos empleados

- Tecnologías

- Árboles de decisiones
- Redes neuronales



- Variables explicativas

- TaxonID (especie)
- dive_month (mes)

- Variables predictoras

- Transect.code (transecto)
- Island (Isla)
- Bioregion (Bioregión)

Árboles de decisiones

Islas

	precision	recall	f1-score	support
0	0.11	0.04	0.06	47
1	0.23	0.20	0.21	107
2	0.36	0.06	0.10	196
3	0.26	0.26	0.26	148
4	0.22	0.30	0.26	79
5	0.46	0.80	0.59	477
6	0.25	0.04	0.07	51
7	0.00	0.00	0.00	9
8	0.33	0.07	0.12	14
9	0.30	0.08	0.12	119
10	0.35	0.28	0.31	98
accuracy			0.39	1345

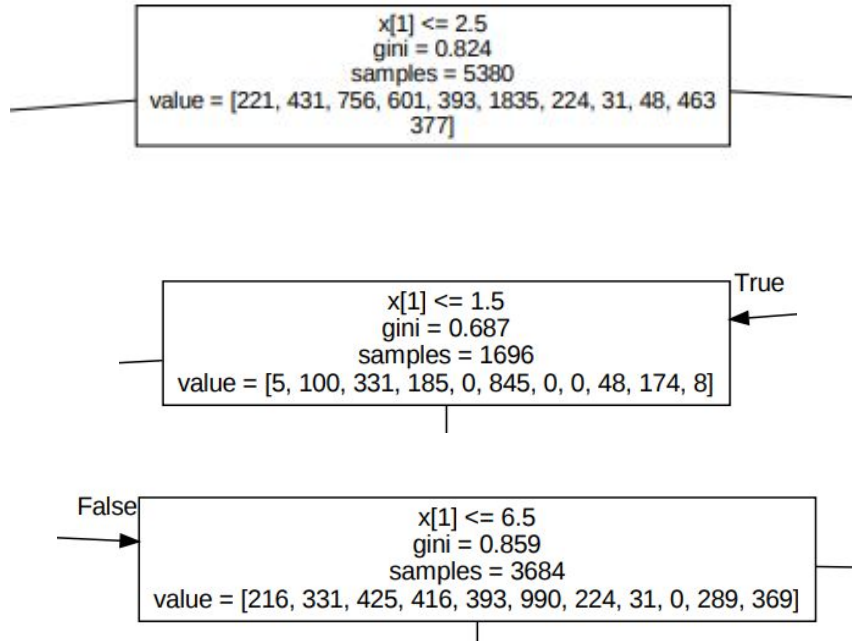
Transecto: 0% accuracy

Bioregión

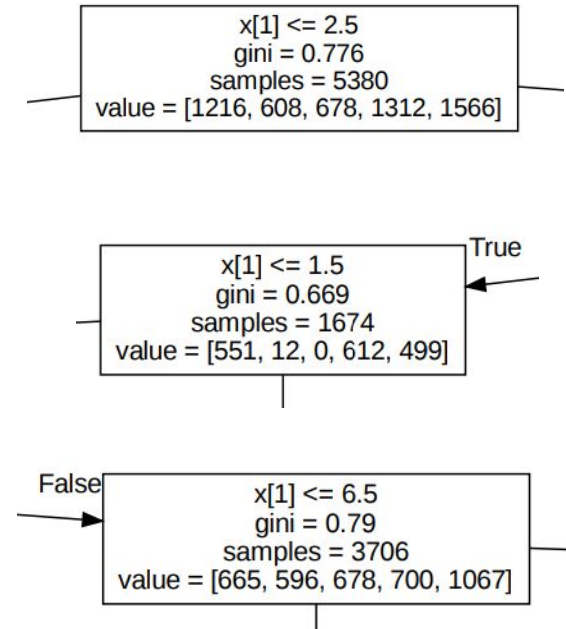
	precision	recall	f1-score	support
0	0.43	0.30	0.36	324
1	0.47	0.53	0.50	159
2	0.29	0.20	0.23	147
3	0.47	0.36	0.41	355
4	0.38	0.60	0.46	360
accuracy			0.41	1345

Árboles de decisiones

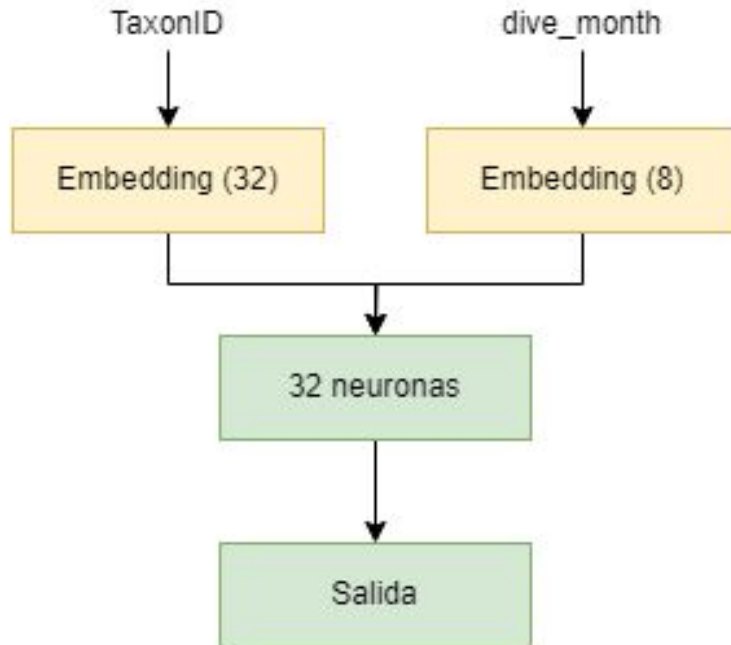
Islas



Bioregión



Redes neuronales



- Función activación: ReLU
- Última capa: SoftMax
- Loss: Categorical cross entropy
- Learning rate: 0.0003
- Optimizador: Adam
- Early Stopping

Redes neuronales

Transecto: 0% accuracy

Islas

	precision	recall	f1-score	support
0	0.10	0.02	0.04	45
1	0.28	0.18	0.22	110
2	0.33	0.03	0.05	195
3	0.33	0.24	0.28	162
4	0.27	0.38	0.31	93
5	0.45	0.86	0.59	447
6	0.12	0.01	0.03	71
7	0.00	0.00	0.00	7
8	0.00	0.00	0.00	18
9	0.39	0.09	0.14	104
10	0.35	0.46	0.40	93
accuracy			0.40	1345

Bioregión

	precision	recall	f1-score	support
0	0.46	0.38	0.42	313
1	0.49	0.60	0.54	138
2	0.36	0.27	0.31	171
3	0.49	0.39	0.43	326
4	0.46	0.62	0.53	397
accuracy			0.46	1345

Conclusiones

Objetivos cumplidos

- Limpieza de datos
- Análisis de los datos
- Creación de los modelos y análisis de los mismos